Forum of Mathematics, Sigma (2015), Vol. 3, e13, 93 pages
doi:10.1017/fms.2015.14

THE HYPERBOLIC QUADRATIC EIGENVALUE PROBLEM

XIN LIANG[1] and REN-CANG LI[2]

[1] Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1, 39106 Magdeburg, Germany;
email: liang@mpi-magdeburg.mpg.de

[2] Department of Mathematics, University of Texas at Arlington, PO Box 19408, Arlington, TX 76019, USA;
email: rcli@uta.edu

Received 30 January 2014; accepted 2 June 2015

Abstract

The hyperbolic quadratic eigenvalue problem (HQEP) was shown to admit Courant–Fischer type min–max principles in 1955 by Duffin and Cauchy type interlacing inequalities in 2010 by Veselić. It can be regarded as the closest analog (among all kinds of quadratic eigenvalue problem) to the standard Hermitian eigenvalue problem (among all kinds of standard eigenvalue problem). In this paper, we conduct a systematic study on the HQEP both theoretically and numerically. On the theoretical front, we generalize Wielandt–Lidskii type min–max principles and, as a special case, Fan type trace min/max principles and establish Weyl type and Wielandt–Lidskii–Mirsky type perturbation results when an HQEP is perturbed to another HQEP. On the numerical front, we justify the natural generalization of the Rayleigh–Ritz procedure with existing principles and our new optimization principles, and, as consequences of these principles, we extend various current optimization approaches—steepest descent/ascent and nonlinear conjugate gradient type methods for the Hermitian eigenvalue problem—to calculate a few extreme eigenvalues (of both positive and negative type). A detailed convergence analysis is given for the steepest descent/ascent methods. The analysis reveals the intrinsic quantities that control convergence rates and consequently yields ways of constructing effective preconditioners. Numerical examples are presented to demonstrate the proposed theory and algorithms.

2010 Mathematics Subject Classification: 15A42, 65F15 (primary); 15A18, 65F08, 65F50, 65G99 (secondary)

1. Introduction

It was argued in [27] that the hyperbolic quadratic eigenvalue problem (HQEP) is the closest analog to the standard Hermitian eigenvalue problem among quadratic eigenvalue problems (QEPs)

$$(\lambda^2 A + \lambda B + C)x = 0. \tag{1.1}$$

In many ways, both problems share common properties: the eigenvalues are all real and semisimple, and for the HQEP there is a version of the min–max principles [13, 1955] that are very much like the Courant–Fischer min–max principles.

One source of QEPs (1.1) is dynamical systems with friction, where A and C are associated with the kinetic-energy and potential-energy quadratic forms, respectively, and B is associated with the Rayleigh dissipation function [17, 67]. When A, B, and C are Hermitian, and A and B are positive definite and C positive semidefinite, we say that the dynamical system is *overdamped* if

$$(x^{\mathrm{H}}Bx)^2 - 4(x^{\mathrm{H}}Ax)(x^{\mathrm{H}}Cx) > 0 \quad \text{for any nonzero vector } x.$$

Overdamped dynamical systems are common in elevator and car braking systems (W. Kahan, private communications, November 2013). An HQEP is slightly more general than an overdamped QEP in that B and C are no longer required to be positive definite or positive semidefinite, respectively. However, a suitable shift in λ can turn an HQEP into an overdamped QEP [23].

In this paper, we undertake a systematic study of the HQEP both in theory and in numerical computation that further reinforces the belief that this class of QEP is the closest analog to the standard Hermitian eigenvalue problem. On the theoretical front, we will

- review existing results of Courant–Fischer type min–max principles and Cauchy interlacing inequalities;
- establish Wielandt–Lidskii type min–max principles for the sums of selected eigenvalues and, as corollaries, trace min/max type principles;
- establish perturbation results in the spectral norm, as well as general unitarily invariant norms, on how the eigenvalues change if A, B, C are perturbed.

On the numerical front, we will

- justify a naturally extended Rayleigh–Ritz type procedure, with the existing and newly established min–max principles, and why the procedure will produce the best approximations to eigenvalues/eigenvectors;

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

- propose extended steepest descent/ascent and conjugate gradient type methods for computing extreme eigenpairs;
- establish convergence results, including the rate of convergence for the extended steepest descent/ascent methods, which shed light on preconditioning in what constitutes a good preconditioner and how to construct one.

In a separate paper, we will extend most of the development in this paper to the hyperbolic polynomial eigenvalue problem. The rest of this paper is organized as follows. Section 2 sets up our notational convention for the rest of this paper. In Section 3, we collect some properties for hyperbolic quadratic matrix polynomials, and in Section 4 we establish important eigen-properties of an HQEP through its linearization. Wielandt–Lidskii type min–max principles, among others, are given in Section 5. Eigen-perturbation analysis for an HQEP is done in Section 6. In Section 7, we justify the use of the Rayleigh–Ritz procedure for extracting interesting eigenvalues and their associated eigenvectors within a given subspace. The steepest descent/ascent method and its extended variation are studied in Section 8, where a detailed convergence analysis is performed. Section 9 investigates the preconditioning techniques to speed up the extended steepest descent/ascent method and explain how an effective preconditioner should be constructed from two different perspectives. Section 10 introduces block variations of the methods in the previous two sections. Various conjugate gradient methods—the plain, locally optimal, and extended subspace search versions combined with suitable preconditioners and blocking—are described in detail in Section 11. Two numerical examples are presented in Section 12 to demonstrate the effectiveness of the locally optimal block preconditioned conjugate gradient method in the previous section. Finally, in Section 13, we present our concluding remarks. We use appendices A and C to take care of long and difficult proofs for three of our theorems in Sections 6 and 8. In Appendix B, we review the Jordan canonical form of a positive semidefinite matrix pencil and establish a perturbation theory for a positive definite matrix pencil for use in Section 6.

2. Notation

Throughout this paper, $\mathbb{C}^{n\times m}$ is the set of all $n \times m$ complex matrices, $\mathbb{C}^n = \mathbb{C}^{n\times 1}$, and $\mathbb{C} = \mathbb{C}^1$. $\mathbb{R}$ is the set of all real numbers. I_n (or simply I if its dimension is clear from the context) is the $n \times n$ identity matrix, and e_j is its jth column. X^{H} is the conjugate transpose of a vector or matrix. For $X \in \mathbb{C}^{n\times m}$, $\sigma_{\min}(X)$ is the smallest singular value of X (X has $\min\{m, n\}$ singular values), $\|X\|_2$ and $\|X\|_{\mathrm{F}}$ and $\|X\|_{\mathrm{ui}}$ are the spectral, Frobenius, and a general unitarily invariant norm of X, and $\kappa_2(X) = \|X\|_2\|X^{-1}\|_2$ is the condition number of a square matrix X.

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

We use $A \succ 0$ ($A \succeq 0$) to mean that A is Hermitian positive (semi-)definite, and $A \prec 0$ ($A \preceq 0$) if $-A \succ 0$ ($-A \succeq 0$). For $A \succeq 0$, $A^{1/2}$ is the unique positive semidefinite square root of A.

The integer triplet

$$(i_-(H), i_0(H), i_+(H))$$

denotes the inertia of an Hermitian matrix H, meaning that H has $i_-(H)$ negative, $i_0(H)$ zero, and $i_+(H)$ positive eigenvalues, respectively, and $\lambda_{\min}(H)$ and $\lambda_{\max}(H)$ are its smallest eigenvalue and its largest eigenvalue, respectively.

The generic notation $\mathrm{eig}(\cdot)$ is the set of all eigenvalues, counting algebraic multiplicities, of a matrix or a matrix pencil, depending on its argument(s): $\mathrm{eig}(A)$ is for a square matrix A, and $\mathrm{eig}(A, B)$ is for a square matrix pencil $A - \lambda B$.

3. Hyperbolic quadratic matrix polynomial

Given $A, B, C \in \mathbb{C}^{n\times n}$, define

$$\boldsymbol{Q}(\lambda) := \lambda^2 A + \lambda B + C, \tag{3.1}$$

a quadratic matrix polynomial of order n. The quadratic eigenvalue problem (QEP) for $\boldsymbol{Q}$, and similarly below, is to find $\lambda \in \mathbb{C}$ and $0 \neq x \in \mathbb{C}^n$ such that

$$\boldsymbol{Q}(\lambda)x = 0.$$

When this equation is satisfied, λ is called an *eigenvalue* and x the associated *eigenvector*. Evidently all eigenvalues of $\boldsymbol{Q}(\cdot)$ are the roots of $\det \boldsymbol{Q}(\lambda) = 0$, which has $2n$ (complex) roots (some of them may be infinite if A is singular), counting multiplicities, assuming that $\det \boldsymbol{Q}(\lambda) \not\equiv 0$. In what follows, we will use $\mathrm{spec}(\boldsymbol{Q})$ to denote the set of all $2n$ eigenvalues of $\boldsymbol{Q}(\cdot)$.

DEFINITION 3.1. $\boldsymbol{Q}(\lambda)$ is said to be *Hermitian* if A, B, and C are all Hermitian, *hyperbolic* if it is Hermitian, $A \succ 0$, and

$$(x^{\mathrm{H}}Bx)^2 - 4(x^{\mathrm{H}}Ax)(x^{\mathrm{H}}Cx) > 0 \quad \text{for all } 0 \neq x \in \mathbb{C}^n, \tag{3.2}$$

and *overdamped* if it is hyperbolic and $B \succ 0$, $C \succeq 0$.

The next theorem summarizes some of the relevant theoretical results on hyperbolic quadratic polynomials. They can be found in Guo and Lancaster [23], which is an excellent gateway to references of origins for these results. Item 3(c) can be found in [66, (0.7)].

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

THEOREM 3.1. *Let $\boldsymbol{Q}(\lambda) = \lambda^2 A + \lambda B + C$ be Hermitian with $A \succ 0$.*

(1) *$\boldsymbol{Q}(\lambda)$ is hyperbolic if and only if there exists $\lambda_0 \in \mathbb{R}$ such that $\boldsymbol{Q}(\lambda_0) \prec 0$.*

(2) *If $\boldsymbol{Q}(\lambda)$ is hyperbolic then its eigenvalues are all real and semisimple.*

(3) *Suppose that $\boldsymbol{Q}(\lambda)$ is hyperbolic. Denote its eigenvalues by $\lambda_i^{\pm}$ and arrange them in the order*

$$\lambda_1^- \leqslant \cdots \leqslant \lambda_n^- < \lambda_1^+ \leqslant \cdots \leqslant \lambda_n^+. \tag{3.3}$$

Then

(a) *$\boldsymbol{Q}(\lambda) \prec 0$ for all $\lambda \in (\lambda_n^-, \lambda_1^+)$;*

(b) *$\boldsymbol{Q}(\lambda) \succ 0$ for all $\lambda \in (-\infty, \lambda_1^-) \cup (\lambda_n^+, +\infty)$;*

(c) *the inertia of $\boldsymbol{Q}(\lambda)$ is $(n-k, 0, k)$ for $\lambda \in (\lambda_k^+, \lambda_{k+1}^+)$ or $\lambda \in (\lambda_{n-k}^-, \lambda_{n+1-k}^-)$ for $k = 1, \ldots, n-1$, concluding that $\boldsymbol{Q}(\lambda)$ is indefinite for $\lambda \in (\lambda_1^{\mathrm{typ}}, \lambda_n^{\mathrm{typ}})$, $\mathrm{typ} \in \{+, -\}$;*

(d) *$\boldsymbol{Q}(\lambda)$ is overdamped if and only if $\lambda_n^+ \leqslant 0$.*

An immediate consequence of Theorem 3.1 is a test to determine whether a Hermitian $\boldsymbol{Q}(\lambda)$ with $A \succ 0$ is hyperbolic or not [23]: check if its eigenvalues are all real and, in the case that they are all real, check if $\boldsymbol{Q}(\lambda_0) \prec 0$, where $\lambda_0 = (\lambda_n^- + \lambda_1^+)/2$.

The next theorem seems to be new. It gives a matrix version of the defining property of a hyperbolic quadratic matrix polynomial.

THEOREM 3.2. *Let $\boldsymbol{Q}(\lambda) = \lambda^2 A + \lambda B + C$ be hyperbolic. Then, for any $X \in \mathbb{C}^{n\times m}$ satisfying $X^{\mathrm{H}} A X = I_m$,*

$$(X^{\mathrm{H}} B X)^2 - 4(X^{\mathrm{H}} C X) \succ 0. \tag{3.4}$$

Proof. For any $y \in \mathbb{C}^m$ with $\|y\|_2 = 1$, write $x = Xy$. We have

$$\begin{aligned}
&y^{\mathrm{H}}[(X^{\mathrm{H}} B X)^2 - 4(X^{\mathrm{H}} C X)]y \\
&\quad = (X^{\mathrm{H}} B X y)^{\mathrm{H}} (X^{\mathrm{H}} B X y) - 4(Xy)^{\mathrm{H}} C (Xy) \\
&\quad = \|y\|_2^2 \cdot \|X^{\mathrm{H}} B X y\|_2^2 - 4(Xy)^{\mathrm{H}} C (Xy) \cdot y^{\mathrm{H}} (X^{\mathrm{H}} A X) y && (3.5) \\
&\quad \geqslant [y^{\mathrm{H}} (X^{\mathrm{H}} B X y)]^2 - 4(Xy)^{\mathrm{H}} C (Xy) \cdot (Xy)^{\mathrm{H}} A (Xy) && (3.6) \\
&\quad = (x^{\mathrm{H}} B x)^2 - 4 x^{\mathrm{H}} C x \cdot x^{\mathrm{H}} A x \\
&\quad > 0, && (3.7)
\end{aligned}$$

where we have used $\|y\|_2 = 1$ and $X^{\mathrm{H}} A X = I_m$ for (3.5), and used the Cauchy–Bunyakovsky–Schwarz inequality for (3.6). Therefore $(X^{\mathrm{H}} B X)^2 - 4(X^{\mathrm{H}} C X) \succ 0$ by (3.7). □

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

4. The HQEP and linearization

A common technique for solving QEP (1.1), or more generally the polynomial eigenvalue problem, is *linearization* that converts a polynomial eigenvalue problem to an equivalent generalized (linear) eigenvalue problem of a matrix pencil [17, 25, 44].

Under the condition that A is nonsingular, QEP (1.1) is equivalent to the generalized eigenvalue problem of the following matrix pencil,

$$\mathscr{L}_{\boldsymbol{Q}}(\lambda) := \begin{bmatrix} -C & 0 \\ 0 & A \end{bmatrix} - \lambda \begin{bmatrix} B & A \\ A & 0 \end{bmatrix} = \mathscr{A} - \lambda \mathscr{B}, \tag{4.1}$$

in the sense that $\operatorname{spec}(\boldsymbol{Q}) = \operatorname{eig}(\mathscr{A}, \mathscr{B})$ and associated eigenvectors of one can be recovered from those for the other. Relevant results, including the case that $\boldsymbol{Q}(\lambda)$ is hyperbolic, are summarized in the following theorem. These results can be found in [1, 5, 10, 26], [27, Theorem 3.6], and [65, Theorem 5A].

THEOREM 4.1. *Let $\boldsymbol{Q}(\lambda) = \lambda^2 A + \lambda B + C$ be a quadratic matrix polynomial of order n, and let $\mathscr{L}_{\boldsymbol{Q}}(\lambda)$ be as in (4.1). Suppose that A is nonsingular.*

(1) $\operatorname{spec}(\boldsymbol{Q}) = \operatorname{eig}(\mathscr{A}, \mathscr{B})$*; that is, the set of eigenvalues of $\boldsymbol{Q}(\cdot)$ is the same as that of the matrix pencil $\mathscr{A} - \lambda\mathscr{B}$.*

(2) *If $A \succ 0$ and B is Hermitian, then the inertia of $\mathscr{B}$ is $(n, 0, n)$.*

(3) *If (μ, x) is an eigenpair of $\boldsymbol{Q}(\lambda)$, then $\left(\mu, \left[\begin{smallmatrix} x \\ \mu x \end{smallmatrix}\right]\right)$ is an eigenpair of $\mathscr{L}_{\boldsymbol{Q}}(\lambda)$.*

(4) *If $\left(\mu, \left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]\right)$ is an eigenpair of $\mathscr{L}_{\boldsymbol{Q}}(\lambda)$, then (μ, x) is an eigenpair of $\boldsymbol{Q}(\lambda)$ and $y = \mu x$.*

(5) *Suppose that $\boldsymbol{Q}(\lambda)$ is Hermitian. $\boldsymbol{Q}(\lambda)$ is hyperbolic if and only if $\mathscr{L}_{\boldsymbol{Q}}(\lambda)$ is a positive definite pencil, that is, there exists a $\lambda_0 \in \mathbb{R}$ such that $\mathscr{L}_{\boldsymbol{Q}}(\lambda_0) \succ 0$.*

(6) *Suppose that $\boldsymbol{Q}(\lambda)$ is hyperbolic, and adopt the notation in item 3 of Theorem 3.1. Then $\mathscr{L}_{\boldsymbol{Q}}(\lambda) \succ 0$ for all $\lambda \in (\lambda_n^-, \lambda_1^+)$.*

Proof. Since, for any $\lambda \in \mathbb{C}$,

$$\begin{bmatrix} I & 0 \\ -\lambda I & I \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} -\boldsymbol{Q}(\lambda) & 0 \\ 0 & A \end{bmatrix} \begin{bmatrix} I & 0 \\ -\lambda I & I \end{bmatrix} = \begin{bmatrix} -C - \lambda B & -\lambda A \\ -\lambda A & A \end{bmatrix} = \mathscr{L}_{\boldsymbol{Q}}(\lambda), \tag{4.2}$$

we have $(-1)^n \det \boldsymbol{Q}(\lambda) \cdot \det A \equiv \det \mathscr{L}_{\boldsymbol{Q}}(\lambda)$, and thus item (1) follows. For item (2), $A \succ 0$ guarantees that there is a nonsingular matrix $X \in \mathbb{C}^{n\times n}$ such that

$$X^{\mathrm{H}} A X = I_n, \quad X^{\mathrm{H}} B X = \operatorname{diag}(\omega_1, \ldots, \omega_n) =: \Omega,$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

where $\omega_i \in \mathbb{R}$. We have

$$\begin{bmatrix} X & \\ & X \end{bmatrix}^{\mathrm{H}} \mathscr{B} \begin{bmatrix} X & \\ & X \end{bmatrix} = \begin{bmatrix} \Omega & I_n \\ I_n & 0 \end{bmatrix}, \tag{4.3}$$

whose eigenvalues are the union of all the eigenvalues of

$$\begin{bmatrix} \omega_i & 1 \\ 1 & 0 \end{bmatrix} \quad \text{for } i = 1, 2, \ldots, n.$$

But the two eigenvalues of each one of these 2×2 matrices are

$$\frac{\omega_i - \sqrt{\omega_i^2 + 4}}{2} < 0, \quad \frac{\omega_i + \sqrt{\omega_i^2 + 4}}{2} > 0.$$

Therefore the last matrix in (4.3) has n positive and n negative eigenvalues, as expected. Items (3) and (4) can be verified in a straightforward way by using (4.2). Also, by using (4.2), we see that $\mathrm{diag}(-\boldsymbol{Q}(\lambda), A)$ and $\mathscr{L}_{\boldsymbol{Q}}(\lambda)$ are congruent for all $\lambda \in \mathbb{R}$, and hence items (5) and (6) follow from items (1) and 3(a) of Theorem 3.1, respectively. □

One consequence of Theorem 4.1 is that any hyperbolic $\boldsymbol{Q}(\lambda) = \lambda^2 A + \lambda B + C$ gives rise to a positive definite matrix pencil $\mathscr{L}_{\boldsymbol{Q}}(\lambda)$ as defined by (4.1) with $\mathscr{B}$ having inertia $(n, 0, n)$. There is a converse to the statement, too. The details can be found in [**39**, Theorem 2.3].

In Theorems 4.2–4.5 below, we investigate the eigen-properties of $\boldsymbol{Q}$ through the eigen-decomposition of its linearization $\mathscr{L}_{\boldsymbol{Q}}(\lambda)$ in (4.1). Define, for a hyperbolic $\boldsymbol{Q}(\lambda)$,

$$\varsigma(x) := \left[(x^{\mathrm{H}} B x)^2 - 4(x^{\mathrm{H}} A x)(x^{\mathrm{H}} C x)\right]^{1/2}, \quad \varsigma_0(x) := \frac{\varsigma(x)}{x^{\mathrm{H}} x}. \tag{4.4}$$

THEOREM 4.2. *Let $\boldsymbol{Q}(\lambda) = \lambda^2 A + \lambda B + C$ be a hyperbolic quadratic matrix polynomial of order n, denote by $\lambda_i^{\pm}$ its eigenvalues which are arranged as in (3.3), and set*

$$\Lambda_+ = \mathrm{diag}(\lambda_1^+, \ldots, \lambda_n^+), \quad \Lambda_- = \mathrm{diag}(\lambda_1^-, \ldots, \lambda_n^-). \tag{4.5}$$

Let $\boldsymbol{Q}(\lambda)$ be linearized to $\mathscr{L}_{\boldsymbol{Q}}(\lambda)$ in (4.1). Then there exists a nonsingular $Z \in \mathbb{C}^{2n \times 2n}$ of the form

$$Z = \begin{bmatrix} U_+ & U_- \\ U_+ \Lambda_+ & U_- \Lambda_- \end{bmatrix}, \tag{4.6}$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

where $U_+,\ U_- \in \mathbb{C}^{n\times n}$ *are nonsingular and*

$$\Upsilon := U_+^{-1}U_- \tag{4.7}$$

is unitary, such that

$$Z^{\mathrm{H}}\mathscr{A}Z \equiv Z^{\mathrm{H}}\begin{bmatrix} -C & \\ & A \end{bmatrix}Z = \begin{bmatrix} \Lambda_+ & \\ & -\Lambda_- \end{bmatrix}, \tag{4.8a}$$

$$Z^{\mathrm{H}}\mathscr{B}Z \equiv Z^{\mathrm{H}}\begin{bmatrix} B & A \\ A & \end{bmatrix}Z = \begin{bmatrix} I_n & \\ & -I_n \end{bmatrix}. \tag{4.8b}$$

Moreover, the i*th column* u_i^+ *of* U_+ *and the* j*th column* u_j^- *of* U_- *are the eigenvectors associated with* λ_i^+ *and* λ_j^-*, respectively; that is,*

$$\boldsymbol{Q}(\lambda_i^+)u_i^+ = 0, \quad \boldsymbol{Q}(\lambda_j^-)u_j^- = 0 \quad \textit{for } i,\ j = 1, 2, \ldots, n. \tag{4.9}$$

These eigenvectors are normalized in the sense that

$$\varsigma(u_i^{\pm}) = 1 \quad \textit{for } i = 1, 2, \ldots, n.$$

Proof. Since $\boldsymbol{Q}(\lambda)$ is hyperbolic, $\mathscr{L}_{\boldsymbol{Q}}(\lambda)$ in (4.1) is a positive definite pencil. By Theorem B.1, there exists a nonsingular $Z \in \mathbb{C}^{2n\times 2n}$ to give (4.8). We have to show that Z must take the form (4.6). Since each column of Z is an eigenvector of the pencil $\mathscr{L}_{\boldsymbol{Q}}(\lambda)$, by Theorem 4.1, we conclude that the ith column of Z can be expressed as $\left[\begin{smallmatrix} u_i^+ \\ \lambda_i^+ u_i^+ \end{smallmatrix}\right]$ for $1 \leqslant i \leqslant n$ or $\left[\begin{smallmatrix} u_j^- \\ \lambda_j^- u_j^- \end{smallmatrix}\right]$ for $1 \leqslant j = i - n \leqslant n$, where u_i^+ and u_j^- are the corresponding eigenvectors of $\boldsymbol{Q}(\lambda)$ associated with λ_i^+ and λ_j^-, respectively. Hence Z takes the form (4.6) with $U_\pm$ given by

$$U_+ = [u_1^+, u_2^+, \ldots, u_n^+], \quad U_- = [u_1^-, u_2^-, \ldots, u_n^-]. \tag{4.10}$$

Blockwise, the equations in (4.8) yield

$$U_+^{\mathrm{H}}CU_+ - \Lambda_+U_+^{\mathrm{H}}AU_+\Lambda_+ = -\Lambda_+, \tag{4.11a}$$

$$U_-^{\mathrm{H}}CU_- - \Lambda_-U_-^{\mathrm{H}}AU_-\Lambda_- = \Lambda_-, \tag{4.11b}$$

$$U_+^{\mathrm{H}}CU_- - \Lambda_+U_+^{\mathrm{H}}AU_-\Lambda_- = 0, \tag{4.11c}$$

$$U_+^{\mathrm{H}}BU_+ + U_+^{\mathrm{H}}AU_+\Lambda_+ + \Lambda_+U_+^{\mathrm{H}}AU_+ = I, \tag{4.11d}$$

$$U_-^{\mathrm{H}}BU_- + U_-^{\mathrm{H}}AU_-\Lambda_- + \Lambda_-U_-^{\mathrm{H}}AU_- = -I, \tag{4.11e}$$

$$U_+^{\mathrm{H}}BU_- + U_+^{\mathrm{H}}AU_-\Lambda_- + \Lambda_+U_+^{\mathrm{H}}AU_- = 0. \tag{4.11f}$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

We claim that U_+ is nonsingular. Consider $U_+x = 0$ for some $x \in \mathbb{C}^n$. We will prove that $x = 0$, and thus U_+ is nonsingular. By (4.11d),

$$x^{\mathrm{H}}x = x^{\mathrm{H}}Ix = x^{\mathrm{H}}(U_+^{\mathrm{H}}BU_+ + U_+^{\mathrm{H}}AU_+\Lambda_+ + \Lambda_+U_+^{\mathrm{H}}AU_+)x = 0,$$

which implies that $x = 0$, as was to be shown. Similarly, U_- is nonsingular.

Next, we define

$$\widehat{\Lambda}_+ := U_+\Lambda_+U_+^{-1}, \quad \widehat{\Lambda}_- := U_-\Lambda_-U_-^{-1}. \tag{4.12}$$

We deduce from (4.11c) and (4.11f) the expressions for C and B in (4.13a) below, and then use $C = C^{\mathrm{H}}$ and $B = B^{\mathrm{H}}$ to get (4.13b).

$$C = \widehat{\Lambda}_-^{\mathrm{H}}A\widehat{\Lambda}_+, \quad B = -A\widehat{\Lambda}_+ - \widehat{\Lambda}_-^{\mathrm{H}}A, \tag{4.13a}$$

$$C = \widehat{\Lambda}_+^{\mathrm{H}}A\widehat{\Lambda}_-, \quad B = -A\widehat{\Lambda}_- - \widehat{\Lambda}_+^{\mathrm{H}}A. \tag{4.13b}$$

Using the second equation in (4.13a), we deduce from (4.11d) and (4.11e) that

$$U_+^{-\mathrm{H}}U_+^{-1} = B + A\widehat{\Lambda}_+ + \widehat{\Lambda}_+^{\mathrm{H}}A = (\widehat{\Lambda}_+ - \widehat{\Lambda}_-)^{\mathrm{H}}A,$$
$$U_-^{-\mathrm{H}}U_-^{-1} = -B - A\widehat{\Lambda}_- - \widehat{\Lambda}_-^{\mathrm{H}}A = A(\widehat{\Lambda}_+ - \widehat{\Lambda}_-).$$

So $U_+^{-\mathrm{H}}U_+^{-1} = (U_-^{-\mathrm{H}}U_-^{-1})^{\mathrm{H}} = U_-^{-\mathrm{H}}U_-^{-1}$. Thus,

$$(U_+^{-1}U_-)^{\mathrm{H}}U_+^{-1}U_- = U_-^{\mathrm{H}}U_+^{-\mathrm{H}}U_+^{-1}U_- = I,$$

which leads to $\Upsilon := U_+^{-1}U_-$ being unitary.

It is straightforward to verify that the columns of $U_\pm$ are eigenvectors and that (4.9) holds. We now prove that $\varsigma(u_i^+) = 1$, and the case for u_i^- can be handled in exactly the same way. Write $a_i = (u_i^+)^{\mathrm{H}}Au_i^+$, $b_i = (u_i^+)^{\mathrm{H}}Bu_i^+$, and $c_i = (u_i^+)^{\mathrm{H}}Cu_i^+$. By (4.11a) and (4.11d), we have

$$c_i - (\lambda_i^+)^2a_i = -\lambda_i^+, \quad b_i + 2a_i\lambda_i^+ = 1,$$

which yield $c_i = -\lambda_i^+ + (\lambda_i^+)^2a_i$ and $b_i = 1 - 2a_i\lambda_i^+$. Thus

$$b_i^2 - 4a_ic_i = (1 - 2a_i\lambda_i^+)^2 - 4a_i[-\lambda_i^+ + (\lambda_i^+)^2a_i] = 1;$$

that is, $\varsigma(u_i^+) = 1$. □

Through the eigen-decomposition (4.8) of the linearization $\mathscr{L}_{\boldsymbol{Q}}(\lambda)$ of $\boldsymbol{Q}(\lambda)$, Theorem 4.2 defines $U_\pm$, $\Lambda_\pm$, and Υ (they are not independent because of (4.7)).

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

Mathematically, these matrices are defined by the coefficient matrices A, B, and C of $\boldsymbol{Q}(\lambda)$, assuming that $\boldsymbol{Q}$ is hyperbolic. In return, the next theorem says that A, B, and C can be parameterized in terms of $U_\pm$, $\Lambda_\pm$, and Υ as well.

THEOREM 4.3. *Under the condition of Theorem 4.2 and the notation there, we have the following.*

(1) $\boldsymbol{Q}(\lambda)$ *admits the factorizations*

$$\boldsymbol{Q}(\lambda) = U_-^{-\mathrm{H}}(\lambda I - \Lambda_-)U_-^{\mathrm{H}} A U_+ (\lambda I - \Lambda_+)U_+^{-1}, \tag{4.14a}$$

$$\boldsymbol{Q}(\lambda) = U_+^{-\mathrm{H}}(\lambda I - \Lambda_+)U_+^{\mathrm{H}} A U_- (\lambda I - \Lambda_-)U_-^{-1}. \tag{4.14b}$$

(2) A, B, C, *and* $\boldsymbol{Q}(\lambda)$ *can be expressed in terms of* $\Lambda_\pm$ *and any two of* U_+, U_-, *and* Υ *as follows:*

$$A = U_+^{-\mathrm{H}} \Theta U_+^{-1}, \tag{4.15a}$$

$$B = U_+^{-\mathrm{H}}(I - \Theta\Lambda_+ - \Lambda_+\Theta)U_+^{-1}, \tag{4.15b}$$

$$C = U_+^{-\mathrm{H}}(\Lambda_+\Theta\Lambda_+ - \Lambda_+)U_+^{-1}, \tag{4.15c}$$

$$\boldsymbol{Q}(\lambda) = U_+^{-\mathrm{H}}[(\lambda I - \Lambda_+)\Theta(\lambda I - \Lambda_+) + (\lambda I - \Lambda_+)]U_+^{-1}, \tag{4.15d}$$

where

$$\Theta = (\Lambda_+ - \Upsilon\Lambda_-\Upsilon^{\mathrm{H}})^{-1}. \tag{4.15e}$$

Proof. For item (1), we have, by (4.13),

$$\boldsymbol{Q}(\lambda) = (\lambda I - \widehat{\Lambda}_-^{\mathrm{H}})A(\lambda I - \widehat{\Lambda}_+), \quad \boldsymbol{Q}(\lambda) = (\lambda I - \widehat{\Lambda}_+^{\mathrm{H}})A(\lambda I - \widehat{\Lambda}_-),$$

which, together with (4.12), yield (4.14). For item (2), write $\Lambda_{-;\Upsilon} = \Upsilon\Lambda_-\Upsilon^{\mathrm{H}}$; then $\Lambda_+ - \Lambda_{-;\Upsilon} \succ 0$ because, for $x \neq 0$,

$$x^{\mathrm{H}}(\Lambda_+ - \Lambda_{-;\Upsilon})x \geqslant \lambda_1^+ x^{\mathrm{H}}x - \lambda_n^- x^{\mathrm{H}}\Upsilon^{\mathrm{H}}\Upsilon x = (\lambda_1^+ - \lambda_n^-)x^{\mathrm{H}}x > 0,$$

which also implies that

$$0 \prec (\Lambda_+ - \Lambda_{-;\Upsilon})^{-1} \preceq (\lambda_1^+ - \lambda_n^-)^{-1} I. \tag{4.16}$$

Substitute $U_- = U_+\Upsilon$ into (4.11c) to get $U_+^{\mathrm{H}} C U_+ - \Lambda_+ U_+^{\mathrm{H}} A U_+ \Lambda_{-;\Upsilon} = 0$. Then, by (4.11a), we have

$$\begin{aligned} 0 &= U_+^{\mathrm{H}} C U_+ - \Lambda_+ U_+^{\mathrm{H}} A U_+ \Lambda_+ + \Lambda_+ \\ &= \Lambda_+ U_+^{\mathrm{H}} A U_+ \Lambda_{-;\Upsilon} - \Lambda_+ U_+^{\mathrm{H}} A U_+ \Lambda_+ + \Lambda_+ \\ &= \Lambda_+[I - U_+^{\mathrm{H}} A U_+ (\Lambda_+ - \Lambda_{-;\Upsilon})]. \end{aligned} \tag{4.17}$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

Substitute $U_+ = U_-\Upsilon^{\mathrm{H}}$ into (4.11c) to get $U_-^{\mathrm{H}}CU_- - \Lambda_{+;\Upsilon}U_-^{\mathrm{H}}AU_-\Lambda_- = 0$, where $\Lambda_{+;\Upsilon} = \Upsilon^{\mathrm{H}}\Lambda_+\Upsilon$. Then, by (4.11b), we have

$$\begin{aligned} 0 &= U_-^{\mathrm{H}}CU_- - \Lambda_-U_-^{\mathrm{H}}AU_-\Lambda_- - \Lambda_- \\ &= \Lambda_{+;\Upsilon}U_-^{\mathrm{H}}AU_-\Lambda_- - \Lambda_-U_-^{\mathrm{H}}AU_-\Lambda_- - \Lambda_- \\ &= -[I - (\Lambda_{+;\Upsilon} - \Lambda_-)U_-^{\mathrm{H}}AU_-]\Lambda_-. \end{aligned} \tag{4.18}$$

We note that at least one of Λ_+ and Λ_- is nonsingular. If Λ_+ is nonsingular, then (4.17) implies that

$$U_+^{\mathrm{H}}AU_+(\Lambda_+ - \Lambda_{-;\Upsilon}) = I \quad \Rightarrow \quad U_+^{\mathrm{H}}AU_+ = (\Lambda_+ - \Lambda_{-;\Upsilon})^{-1}. \tag{4.19}$$

If Λ_- is nonsingular, then (4.18) implies that $(\Lambda_{+;\Upsilon} - \Lambda_-)U_-^{\mathrm{H}}AU_- = I$ which, upon using $U_- = U_+\Upsilon$, also implies that the second equation in (4.19) holds. So, $U_+^{\mathrm{H}}AU_+ = \Theta$, $U_+^{\mathrm{H}}BU_+ = -\Theta\Lambda_+ - \Lambda_{-;\Upsilon}\Theta$, and $U_+^{\mathrm{H}}CU_+ = \Lambda_{-;\Upsilon}\Theta\Lambda_+$. Noticing that

$$\Lambda_{-;\Upsilon}\Theta = -(\Lambda_+ - \Lambda_{-;\Upsilon})\Theta + \Lambda_+\Theta = -I + \Lambda_+\Theta,$$

we have (4.15). □

REMARK 4.1. (1) Each of the decompositions in (4.14) does not reflect the symmetry property in $\boldsymbol{Q}(\lambda)$. However, using the fact that $\Upsilon = U_+^{-1}U_-$ is unitary, we can turn them into

$$\boldsymbol{Q}(\lambda) = U_+^{-\mathrm{H}}(\lambda I - \Upsilon\Lambda_-\Upsilon^{\mathrm{H}})(\Lambda_+ - \Upsilon\Lambda_-\Upsilon^{\mathrm{H}})^{-1}(\lambda I - \Lambda_+)U_+^{-1}, \tag{4.20a}$$

$$\boldsymbol{Q}(\lambda) = U_-^{-\mathrm{H}}(\lambda I - \Upsilon^{\mathrm{H}}\Lambda_+\Upsilon)(\Upsilon\Lambda_+\Upsilon^{\mathrm{H}} - \Lambda_-)^{-1}(\lambda I - \Lambda_-)U_-^{-1}. \tag{4.20b}$$

These equations are essentially the decomposition in [45, Theorem 31.24] but with more detail.

(2) Lemma 6.1 in [22] and Problem `gen_hyper2` in [6] provide a different set of formulas for B and C:

$$B = U_+^{-\mathrm{H}}[-\Theta(\Lambda_+^2 - \Upsilon\Lambda_-^2\Upsilon^{\mathrm{H}})\Theta]U_+^{-1}, \tag{4.21a}$$

$$\begin{aligned} C = U_+^{-\mathrm{H}}[&-\Theta(\Lambda_+^3 - \Upsilon\Lambda_-^3\Upsilon^{\mathrm{H}})\Theta \\ &+ \Theta(\Lambda_+^2 - \Upsilon\Lambda_-^2\Upsilon^{\mathrm{H}})\Theta(\Lambda_+^2 - \Upsilon\Lambda_-^2\Upsilon^{\mathrm{H}})\Theta]U_+^{-1}. \end{aligned} \tag{4.21b}$$

Corollary 6 in [31] provides yet another formula for C:

$$C = U_+^{-\mathrm{H}}[-(\Lambda_+^{-1} - \Upsilon\Lambda_-^{-1}\Upsilon^{\mathrm{H}})^{-1}]U_+^{-1}. \tag{4.22}$$

Although both (4.21) and (4.22) seem to be very different from ours for B and C in (4.15b) and (4.15c), they are actually the same in theory (see [39] for a proof).

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

(3) The matrices $\widehat{\Lambda}_{\pm}$ in (4.12) are two solutions of the matrix equation

$$AX^2 + BX + C = 0. \tag{4.23}$$

In fact,

$$A(U_+\Lambda_+U_+^{-1})^2+B(U_+\Lambda_+U_+^{-1})+C = (AU_+\Lambda_+^2+BU_+\Lambda_++CU_+)U_+^{-1} = 0,$$

and similarly for $A(U_-\Lambda_-U_-^{-1})^2 + B(U_-\Lambda_-U_-^{-1}) + C = 0$. It can be verified that any solution X to (4.23) gives rise to the factorization $\boldsymbol{Q}(\lambda) = (\lambda A + AX + B)(\lambda I - X)$, based on which Guo and Lancaster [**23**] proposed their solvent approach for solving HQEP (1.1) of modest size. More investigations on factorizing Hermitian quadratic matrix polynomials can be found in [**32**].

The inequalities in the next theorem bounds the condition numbers of the eigenvector matrices $U_{\pm}$ and the eigen-transformation matrix Z defined in Theorem 4.2. They appear in the perturbation bounds for eigenvalues of an HQEP later in Section 6.

THEOREM 4.4. *Let $U_{\pm}$ and Z be as defined in Theorem 4.2. Then*

$$\|U_+\|_2 = \|U_-\|_2 \leqslant \frac{\|A^{-1/2}\|_2}{\sqrt{\lambda_1^+ - \lambda_n^-}}, \tag{4.24a}$$

$$\|U_+^{-1}\|_2 = \|U_-^{-1}\|_2 \leqslant \|A^{1/2}\|_2\sqrt{\lambda_n^+ - \lambda_1^-}, \tag{4.24b}$$

$$\kappa(U_+) = \kappa(U_-) \leqslant \sqrt{\kappa(A)}\sqrt{\frac{\lambda_n^+ - \lambda_1^-}{\lambda_1^+ - \lambda_n^-}}, \tag{4.24c}$$

and

$$\|Z\|_2 \leqslant \xi\|U_{\pm}\|_2, \quad \|Z^{-1}\|_2 \leqslant \frac{\xi}{\lambda_1^+ - \lambda_n^-}\|U_{\pm}^{-1}\|_2, \tag{4.25}$$

where, with $\xi_{\pm} = \max\{|\lambda_1^{\pm}|, |\lambda_n^{\pm}|\}$,

$$\xi = \left(\frac{2 + \xi_+^2 + \xi_-^2 + \sqrt{[(\xi_+ - 1)^2 + (\xi_- + 1)^2][(\xi_+ + 1)^2 + (\xi_- - 1)^2]}}{2}\right)^{1/2}.$$

Proof. The equalities in (4.24) are consequences of $U_- = U_+\Upsilon$ and of Υ being unitary. We now prove the inequality parts in (4.24) for U_+. Use $(A^{1/2}U_+)^{\mathrm{H}}(A^{1/2}U_+) = \Theta$ to get

$$\|U_+\|_2 \leqslant \|A^{-1/2}\|_2\|A^{1/2}U_+\|_2 = \|A^{-1/2}\|_2\sqrt{\|\Theta\|_2} \leqslant \frac{\|A^{-1/2}\|_2}{\sqrt{\lambda_1^+ - \lambda_n^-}},$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

and use $(U_+^{-1}A^{-1/2})(U_+^{-1}A^{-1/2})^{\mathrm{H}} = \Theta^{-1}$ to get

$$\|U_+^{-1}\|_2 \leqslant \|U_+^{-1}A^{-1/2}\|_2\|A^{1/2}\|_2 = \sqrt{\|\Theta^{-1}\|_2}\|A^{1/2}\|_2 \leqslant \|A^{1/2}\|_2\sqrt{\lambda_n^+ - \lambda_1^-}.$$

They give (4.24a) and (4.24b) for U_+. Combine (4.24a) and (4.24b) to get (4.24c). For the first inequality in (4.25), we have

$$\|Z\|_2 \leqslant \left\| \begin{bmatrix} \|U_+\|_2 & \|U_-\|_2 \\ \|U_+\|_2\xi_+ & \|U_-\|_2\xi_- \end{bmatrix} \right\|_2 = \|U_+\|_2 \left\| \begin{bmatrix} 1 & 1 \\ \xi_+ & \xi_- \end{bmatrix} \right\|_2 = \|U_+\|_2\xi.$$

For the second inequality, we notice, by using $U_- = U_+\Upsilon$, that

$$Z = \begin{bmatrix} U_+ & 0 \\ 0 & U_+ \end{bmatrix} \begin{bmatrix} I & \Upsilon \\ \Lambda_+ & \Upsilon\Lambda_- \end{bmatrix} = \begin{bmatrix} U_+ & 0 \\ 0 & U_+ \end{bmatrix} \begin{bmatrix} I & 0 \\ \Lambda_+ & I \end{bmatrix} \begin{bmatrix} I & \Upsilon \\ 0 & S \end{bmatrix},$$

where $S = \Upsilon\Lambda_- - \Lambda_+\Upsilon = -\Theta^{-1}\Upsilon$. This expression, after some calculations, leads to

$$\begin{aligned} Z^{-1} &= \begin{bmatrix} I & -\Upsilon S^{-1} \\ 0 & S^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -\Lambda_+ & I \end{bmatrix} \begin{bmatrix} U_+^{-1} & 0 \\ 0 & U_+^{-1} \end{bmatrix} \\ &= \begin{bmatrix} \Upsilon S^{-1}\Upsilon\Lambda_-\Upsilon^{\mathrm{H}} & \Upsilon S^{-1} \\ -S^{-1}\Lambda_+ & S^{-1} \end{bmatrix} \begin{bmatrix} U_+^{-1} & 0 \\ 0 & U_+^{-1} \end{bmatrix}, \end{aligned}$$

and thus

$$\|Z^{-1}\|_2 \leqslant \|S^{-1}\|_2 \left\| \begin{bmatrix} \xi_- & 1 \\ \xi_+ & 1 \end{bmatrix} \right\|_2 \|U_+^{-1}\|_2 = \|U_+^{-1}\|_2\|\Theta\|_2\xi,$$

which implies the second inequality in (4.25). □

With item (3) of Theorem 4.3, it is now only logical to expect that A, B, and C defined by (4.15), given $U_\pm$, $\Lambda_\pm$, and Υ, should give rise to a hyperbolic quadratic polynomial. Indeed this is the case, as stated in the following theorem.

THEOREM 4.5. *Given diagonal matrices $\Lambda_\pm$ as in (4.5), and any two of $n \times n$ matrices U_+, U_-, and unitary Υ with the third determined by (4.7), if $\lambda_i^\pm$ can be arranged as in (3.3), then the quadratic matrix polynomial constructed by (4.15) is hyperbolic.*

Proof. First Θ is Hermitian and $\Theta \succ 0$ by (4.16). Obviously A, B, C in (4.15) are Hermitian, and $A \succ 0$. Choose $\lambda_0 = (\lambda_1^+ + \lambda_n^-)/2$; then $\Theta^{-1} \succ \Lambda_+ - \lambda_0 I \succ 0$ and $\Theta \prec (\Lambda_+ - \lambda_0 I)^{-1}$. Thus,

$$U_+^{\mathrm{H}}\boldsymbol{Q}(\lambda_0)U_+ = (\Lambda_+ - \lambda_0 I)\Theta(\Lambda_+ - \lambda_0 I) - (\Lambda_+ - \lambda_0 I) \prec 0,$$

which says that $\boldsymbol{Q}(\lambda_0) \prec 0$. By item (1) of Theorem 3.1, $\boldsymbol{Q}(\lambda)$ is hyperbolic. □

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

Theorem 4.5 solves one kind of inverse eigenvalue problem for HQEP. For more general inverse problems for Hermitian quadratic matrix polynomials, the reader is referred to [32].

5. Variational principles

Throughout this section, $\boldsymbol{Q}(\lambda) = \lambda^2 A + \lambda B + C \in \mathbb{C}^{n\times n}$ is assumed to be hyperbolic, and the notation used in Theorem 4.2 is kept. The scalar λ_0 is as in item (1) of Theorem 3.1 such that $\boldsymbol{Q}(\lambda_0) \prec 0$; that is, $\lambda_0 \in (\lambda_n^-, \lambda_1^+)$.

Consider the following equation in λ:

$$f(\lambda, x) := x^{\mathrm{H}} \boldsymbol{Q}(\lambda) x = \lambda^2 (x^{\mathrm{H}} A x) + \lambda (x^{\mathrm{H}} B x) + (x^{\mathrm{H}} C x) = 0, \tag{5.1}$$

given $x \neq 0$. Since $\boldsymbol{Q}(\lambda)$ is hyperbolic, this equation always has two distinct real roots (as functions of x):

$$\rho_{\pm}(x) = \frac{-x^{\mathrm{H}} B x \pm [(x^{\mathrm{H}} B x)^2 - 4(x^{\mathrm{H}} A x)(x^{\mathrm{H}} C x)]^{1/2}}{2(x^{\mathrm{H}} A x)}. \tag{5.2}$$

In Duffin [13], they were called the *primary and secondary functionals*, but here we shall call $\rho_+(x)$ the *pos-type Rayleigh quotient* of $\boldsymbol{Q}(\lambda)$ at x, and $\rho_-(x)$ the *neg-type Rayleigh quotient* of $\boldsymbol{Q}(\lambda)$ at x. They were also defined in [17, Ch. 13].

It is easy to verify that, for any $x \neq 0$, $\rho_{\pm}(x) \in \mathbb{R}$, and $\rho_{\pm}(\alpha x) = \rho_{\pm}(x)$ for any nonzero $\alpha \in \mathbb{C}$. By elementary knowledge of scalar quadratic polynomials, we have

$$\rho_+(x) + \rho_-(x) = -\frac{x^{\mathrm{H}} B x}{x^{\mathrm{H}} A x}, \quad \rho_+(x) \cdot \rho_-(x) = \frac{x^{\mathrm{H}} C x}{x^{\mathrm{H}} A x}. \tag{5.3}$$

Both will be used later in this paper. Two other important quantities are $\varsigma(x)$ and $\varsigma_0(x)$, defined in (4.4). Note that

$$\begin{aligned} 2\rho_{\pm}(x)\, x^{\mathrm{H}} A x + x^{\mathrm{H}} B x &= \left[-x^{\mathrm{H}} B x \pm \sqrt{(x^{\mathrm{H}} B x)^2 - 4(x^{\mathrm{H}} A x)(x^{\mathrm{H}} C x)} \right] + x^{\mathrm{H}} B x \\ &= \pm \varsigma(x), \end{aligned}$$

which yields the following alternative representation:

$$\varsigma(x) = \pm [2\rho_{\pm}(x)\, x^{\mathrm{H}} A x + x^{\mathrm{H}} B x], \tag{5.4}$$

where the $\pm$ sign before $[\cdots]$ is selected to make sure that the right-hand side comes out nonnegative.

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

THEOREM 5.1. *We have*

$$\rho_+(x) \in [\lambda_1^+, \lambda_n^+], \quad \rho_-(x) \in [\lambda_1^-, \lambda_n^-], \tag{5.5}$$

$$\varsigma_0(x) \in [(\lambda_1^+ - \lambda_n^-)\lambda_{\min}(A),\ (\lambda_n^+ - \lambda_1^-)\lambda_{\max}(A)]. \tag{5.6}$$

Moreover, $\lambda_i^+ = \rho_+(u_i^+)$ *for the eigenpair* (λ_i^+, u_i^+) *and* $\rho_-(u_j^-) = \lambda_j^-$ *for the eigenpair* (λ_j^-, u_j^-).

Proof. By item (3) of Theorem 3.1, for any fixed nonzero x, $f(\lambda, x) < 0$ for $\lambda \in (\lambda_n^-, \lambda_1^+)$ and $f(\lambda, x) > 0$ for $\lambda \in (-\infty, \lambda_1^-) \cup (\lambda_n^+, +\infty)$. Thus, the larger root of the scalar quadratic equation $f(\lambda, x) = 0$ in λ must lie in $[\lambda_1^+, \lambda_n^+]$, and the smaller one in $[\lambda_1^-, \lambda_n^-]$. This gives us (5.5). The inclusion (5.6) is a result of $\varsigma(x) = [\rho_+(x) - \rho_-(x)]\, x^{\mathrm{H}} A x$. Finally, by the definition of $\rho_\pm(u_i^+)$, we know that one of them is equal to λ_i^+. But $\rho_-(u_i^+) \leqslant \lambda_n^- < \lambda_i^+$ by (5.5), and thus $\lambda_i^+ = \rho_+(u_i^+)$. Similarly, $\rho_-(u_j^-) = \lambda_j^-$. □

5.1. Courant–Fischer type min–max principles. Theorem 5.2 below is a restatement of Theorems 32.10, 32.11, and Remark 32.13 in [45]. However, it is essentially due to Duffin [13, 1955], whose proof, although for overdamped $\boldsymbol{Q}$, works for the general hyperbolic case. Closely related ones for more general nonlinear eigenvalue problems (other than quadratic eigenvalue problems) can be found in [52, 53, 68, 69]. They can be considered as generalizations of Courant–Fischer min–max principles (see [50, page 206], [58, page 201]).

THEOREM 5.2 [13]. *Let* $\mathrm{typ} \in \{+, -\}$. *We have, for* $1 \leqslant i \leqslant n$,

$$\lambda_i^{\mathrm{typ}} = \max_{\substack{\mathcal{X} \subseteq \mathbb{C}^n \\ \operatorname{codim} \mathcal{X} = i-1}} \min_{\substack{x \in \mathcal{X} \\ x \neq 0}} \rho_{\mathrm{typ}}(x), \tag{5.7a}$$

$$\lambda_i^{\mathrm{typ}} = \min_{\substack{\mathcal{X} \subseteq \mathbb{C}^n \\ \dim \mathcal{X} = i}} \max_{\substack{x \in \mathcal{X} \\ x \neq 0}} \rho_{\mathrm{typ}}(x). \tag{5.7b}$$

In particular,

$$\lambda_1^{\mathrm{typ}} = \min_{x \neq 0} \rho_{\mathrm{typ}}(x), \quad \lambda_n^{\mathrm{typ}} = \max_{x \neq 0} \rho_{\mathrm{typ}}(x). \tag{5.8}$$

5.2. Wielandt–Lidskii type min–max principles. The min–max principles in Theorem 5.3, which can be considered as generalizations of Amir-Moéz type min–max principles [2], and in Theorem 5.4, which can be considered as generalizations of the Wielandt–Lidskii min–max principles (see [41, 72] and also [7, page 67], [58, page 199]), and the Fan type trace min/max principles [16] are

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

new. For the ease of stating them, let $\lambda_\pm \in \mathbb{R} \cup \{\pm\infty\}$ such that

$$\lambda_- \leqslant \lambda_1^- \leqslant \lambda_n^- \leqslant \lambda_0 \leqslant \lambda_1^+ \leqslant \lambda_n^+ \leqslant \lambda_+.$$

Such $\lambda_\pm$ exist: for example, $\lambda_- = \lambda_1^-$ or $-\infty$ and $\lambda_+ = \lambda_n^+$ or ∞. Set intervals

$$\mathscr{I}_+ = \begin{cases} [\lambda_0, \lambda_+] & \text{if } \lambda_+ < \infty, \\ [\lambda_0, \infty) & \text{otherwise,} \end{cases} \qquad \mathscr{I}_- = \begin{cases} [\lambda_-, \lambda_0] & \text{if } \lambda_- > -\infty, \\ (-\infty, \lambda_0] & \text{otherwise.} \end{cases} \tag{5.9}$$

The following lemma is also essentially due to Duffin [13], whose proof, although for overdamped Q, again works for the general hyperbolic case.

LEMMA 5.1. *Let* $\mathrm{typ} \in \{+, -\}$. *We have*

$$\lambda_i^{\mathrm{typ}} \geqslant \rho_{\mathrm{typ}}(x) \quad \textit{for any } x \in \mathrm{span}\{u_1^{\mathrm{typ}}, u_2^{\mathrm{typ}}, \ldots, u_i^{\mathrm{typ}}\}, \tag{5.10a}$$

$$\lambda_i^{\mathrm{typ}} \leqslant \rho_{\mathrm{typ}}(x) \quad \textit{for any } x \in \mathrm{span}\{u_i^{\mathrm{typ}}, u_{i+1}^{\mathrm{typ}}, \ldots, u_n^{\mathrm{typ}}\}, \tag{5.10b}$$

where u_j^{typ} *is the corresponding eigenvector to* λ_j^{typ} *for* $j = 1, \ldots, n$.

To generalize Amir-Moéz/Wielandt–Lidskii min–max principles, we introduce the following notation. For $X \in \mathbb{C}^{n\times k}$ with $\mathrm{rank}(X) = k$, $X^{\mathrm{H}}\boldsymbol{Q}(\lambda)X$ is a hyperbolic quadratic matrix polynomial of order k. Hence its eigenvalues are real and semisimple. Denote them by $\lambda_{i,X}^{\pm}$, and arrange them as

$$\lambda_{1,X}^- \leqslant \cdots \leqslant \lambda_{k,X}^- \leqslant \lambda_{1,X}^+ \leqslant \cdots \leqslant \lambda_{k,X}^+. \tag{5.11}$$

THEOREM 5.3. *Let* $1 \leqslant i_1 < \cdots < i_k \leqslant n$ *and* $\mathrm{typ} \in \{+, -\}$. *For any*

$$\Phi : \underbrace{\mathscr{I}_{\mathrm{typ}} \times \cdots \times \mathscr{I}_{\mathrm{typ}}}_{k} \to \mathbb{R}$$

that is nondecreasing in each of its arguments, we have

$$\min_{\substack{\mathfrak{X}_1 \subset \cdots \subset \mathfrak{X}_k \\ \dim \mathfrak{X}_j = i_j}} \sup_{\substack{x_j \in \mathfrak{X}_j,\ j=1,\ldots,k \\ X=[x_1,\ldots,x_k] \\ \mathrm{rank}(X)=k}} \Phi(\lambda_{1,X}^{\mathrm{typ}}, \ldots, \lambda_{k,X}^{\mathrm{typ}}) = \Phi(\lambda_{i_1}^{\mathrm{typ}}, \ldots, \lambda_{i_k}^{\mathrm{typ}}), \tag{5.12a}$$

$$\max_{\substack{\mathfrak{X}_1 \supset \cdots \supset \mathfrak{X}_k \\ \mathrm{codim}\, \mathfrak{X}_j = i_j - 1}} \inf_{\substack{x_j \in \mathfrak{X}_j,\ j=1,\ldots,k \\ X=[x_1,\ldots,x_k] \\ \mathrm{rank}(X)=k}} \Phi(\lambda_{1,X}^{\mathrm{typ}}, \ldots, \lambda_{k,X}^{\mathrm{typ}}) = \Phi(\lambda_{i_1}^{\mathrm{typ}}, \ldots, \lambda_{i_k}^{\mathrm{typ}}). \tag{5.12b}$$

(In (5.12a)*, it is not clear if the* sup *is attainable for any given* $\mathfrak{X}_j$ *satisfying the given assumptions, except for continuous* Φ. *The same comment applies to the* inf

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

in (5.12b).*) If also* Φ *is continuous, then the* sup *in* (5.12a) *and the* inf *in* (5.12b) *can be replaced by* max *and* min*, respectively. In particular, setting* $i_j = j$ *in* (5.12a) *and setting* $i_j = j + n - k$ *in* (5.12b)*, respectively, gives*

$$\min_{\operatorname{rank}(X)=k} \Phi(\lambda_{1,X}^{\mathrm{typ}}, \ldots, \lambda_{k,X}^{\mathrm{typ}}) = \Phi(\lambda_1^{\mathrm{typ}}, \ldots, \lambda_k^{\mathrm{typ}}), \tag{5.13a}$$

$$\max_{\operatorname{rank}(X)=k} \Phi(\lambda_{1,X}^{\mathrm{typ}}, \ldots, \lambda_{k,X}^{\mathrm{typ}}) = \Phi(\lambda_{n-k+1}^{\mathrm{typ}}, \ldots, \lambda_n^{\mathrm{typ}}). \tag{5.13b}$$

Proof. The following proof actually works for any $\mathrm{typ} \in \{+, -\}$ also, but for clarity, we present it for $\mathrm{typ} = +$ only. We also note that the results in this theorem for one $\mathrm{typ} \in \{+, -\}$ easily lead to ones for the other. For example, suppose that we already have proved (5.12) for $\mathrm{typ} = +$. Now consider $\widehat{\boldsymbol{Q}}(\lambda) = \lambda^2 A + \lambda(-B) + C$, whose eigenvalues are

$$\hat{\lambda}_1^- \leqslant \cdots \leqslant \hat{\lambda}_n^- < \hat{\lambda}_1^+ \leqslant \cdots \leqslant \hat{\lambda}_n^+,$$

where $\hat{\lambda}_i^- = -\lambda_{n-i+1}^+$ and $\hat{\lambda}_j^+ = -\lambda_{n-j+1}^-$. Apply (5.12b) for $\mathrm{typ} = +$ to $\widehat{\boldsymbol{Q}}(\lambda)$ and $-\Phi(-\xi_k, \ldots, -\xi_1)$ to get (5.12a) for $\mathrm{typ} = -$, and apply (5.12a) for $\mathrm{typ} = +$ to $\widehat{\boldsymbol{Q}}(\lambda)$ and $-\Phi(-\xi_k, \ldots, -\xi_1)$ to get (5.12b) for $\mathrm{typ} = -$.

We now prove the theorem for $\mathrm{typ} = +$. We introduce, for a matrix $W = [w_1, \ldots, w_p]$,

$$\mathcal{S}_{j,W} := \operatorname{span}\{w_1, \ldots, w_j\}, \quad \mathcal{T}_{j,W} := \operatorname{span}\{w_j, \ldots, w_p\} \quad \text{for } j = 1, \ldots, p.$$

In particular, $\mathcal{S}_W = \mathcal{S}_{p,W}$, $\mathcal{T}_W = \mathcal{T}_{1,W}$, and thus $\mathcal{S}_W = \mathcal{T}_W$.

First we prove (5.12b). Recall the eigenvectors u_j^+ introduced in Theorem 4.2. Choose

$$\widehat{\mathcal{X}}_j = \operatorname{span}\{u_{i_j}^+, \ldots, u_n^+\} \quad \text{for } j = 1, 2, \ldots, k. \tag{5.14}$$

Then $\widehat{\mathcal{X}}_1 \supset \cdots \supset \widehat{\mathcal{X}}_k$ and $\operatorname{codim} \widehat{\mathcal{X}}_j = i_j - 1$. By Lemma 5.1, $\rho_+(x) \geqslant \lambda_{i_j}^+$ for any nonzero $x \in \widehat{\mathcal{X}}_j$. Therefore

$$\min_{\substack{x \in \widehat{\mathcal{X}}_j \\ x \neq 0}} \rho_+(x) = \rho_+(i_j^+) = \lambda_{i_j}^+.$$

For any $X = [x_1, \ldots, x_k]$ with $x_j \in \widehat{\mathcal{X}}_j$ for $j = 1, \ldots, k$ such that $\operatorname{rank}(X) = k$, consider $X^{\mathrm{H}} \boldsymbol{Q}(\lambda) X$, which is a hyperbolic quadratic matrix polynomial of order k. For $j = 1, \ldots, k$, noticing that $\mathcal{T}_{j,X} \subset \widehat{\mathcal{X}}_j$, we have, by Theorem 5.2,

$$\lambda_{j,X}^+ = \max_{\substack{\mathcal{X} \subset \mathcal{T}_X \\ \dim \mathcal{X} = k-j+1}} \min_{\substack{x \in \mathcal{X} \\ x \neq 0}} \rho_+(x) \geqslant \min_{\substack{x \in \mathcal{T}_{j,X} \\ x \neq 0}} \rho_+(x) \geqslant \min_{\substack{x \in \widehat{\mathcal{X}}_j \\ x \neq 0}} \rho_+(x) = \lambda_{i_j}^+.$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

Since $\Phi(\cdot)$ is nondecreasing in each of its arguments,

$$\Phi(\lambda_{1,X}^+, \ldots, \lambda_{k,X}^+) \geqslant \Phi(\lambda_{i_1}^+, \ldots, \lambda_{i_k}^+),$$

which gives

$$\min_{\substack{x_j \in \widehat{\mathfrak{X}}_j,\ j=1,\ldots,k \\ X=[x_1,\ldots,x_k] \\ \operatorname{rank}(X)=k}} \Phi(\lambda_{1,X}^+, \ldots, \lambda_{k,X}^+) \geqslant \Phi(\lambda_{i_1}^+, \ldots, \lambda_{i_k}^+),$$

because $x_j \in \widehat{\mathfrak{X}}_j$ for $1 \leqslant i \leqslant k$ are arbitrary, subject to $\operatorname{rank}(X) = k$. Therefore

$$\sup_{\substack{\mathfrak{X}_1 \supset \cdots \supset \mathfrak{X}_k \\ \operatorname{codim} \mathfrak{X}_j = i_j - 1}} \inf_{\substack{x_j \in \mathfrak{X}_j,\ j=1,\ldots,k \\ X=[x_1,\ldots,x_k] \\ \operatorname{rank}(X)=k}} \Phi(\lambda_{1,X}^+, \ldots, \lambda_{k,X}^+) \geqslant \Phi(\lambda_{i_1}^+, \ldots, \lambda_{i_k}^+). \tag{5.15}$$

On the other hand, let $\mathfrak{X}_j$ for $j = 1, \ldots, k$ be any subspaces that satisfy the following assumptions: $\mathfrak{X}_1 \supset \cdots \supset \mathfrak{X}_k$ and $\operatorname{codim} \mathfrak{X}_j = i_j - 1$. Define $\mathcal{Y}_j = \operatorname{span}\{u_1^+, \ldots, u_{i_j}^+\}$. Then $\mathcal{Y}_1 \subset \cdots \subset \mathcal{Y}_k$ and $\dim \mathcal{Y}_j = i_j$. By [2, Corollary 2.2] (see also [38, Lemma 3.2]), there exist two A-orthonormal sets $\{x_1, \ldots, x_k\}$ and $\{y_1, \ldots, y_k\}$ with $x_j \in \mathfrak{X}_j$ for $1 \leqslant j \leqslant k$ and $y_j \in \mathcal{Y}_j$ for $1 \leqslant j \leqslant k$ such that

$$\mathcal{T}_X := \operatorname{span}\{x_1, \ldots, x_k\} = \operatorname{span}\{y_1, \ldots, y_k\} =: \mathcal{S}_Y,$$

where $X = [x_1, \ldots, x_k]$ and $Y = [y_1, \ldots, y_k]$ satisfy $X^{\mathrm{H}}AX = Y^{\mathrm{H}}AY = I_k$. Then $Y^{\mathrm{H}}\boldsymbol{Q}(\lambda)Y$ is a hyperbolic quadratic matrix polynomial whose pos-type eigenvalues are $\lambda_{1,Y}^+ \leqslant \cdots \leqslant \lambda_{k,Y}^+$. Since $\mathcal{S}_Y = \mathcal{T}_X$, $\lambda_{j,Y}^+ = \lambda_{j,X}^+$ for $j = 1, \ldots, k$. By Lemma 5.1, $\rho_+(y) \leqslant \lambda_{i_j}^+$ for any nonzero $y \in \mathcal{Y}_j$. Therefore

$$\max_{\substack{y \in \mathcal{Y}_j \\ y \neq 0}} \rho_+(y) = \lambda_{i_j}^+.$$

By Theorem 5.2, and noticing that $\mathcal{S}_{j,Y} \subset \mathcal{Y}_j$, we have, for $j = 1, \ldots, k$,

$$\lambda_{j,X}^+ = \lambda_{j,Y}^+ = \min_{\substack{\mathcal{Y} \subset \mathcal{S}_Y \\ \dim \mathcal{Y} = j}} \max_{\substack{y \in \mathcal{Y} \\ y \neq 0}} \rho_+(y) \leqslant \max_{\substack{y \in \mathcal{S}_{j,Y} \\ y \neq 0}} \rho_+(y) \leqslant \max_{\substack{y \in \mathcal{Y}_j \\ y \neq 0}} \rho_+(y) = \lambda_{i_j}^+.$$

Since $\Phi(\cdot)$ is nondecreasing in each of its arguments,

$$\Phi(\lambda_{1,X}^+, \ldots, \lambda_{k,X}^+) \leqslant \Phi(\lambda_{i_1}^+, \ldots, \lambda_{i_k}^+),$$

which gives

$$\inf_{\substack{x_j \in \mathfrak{X}_j,\ j=1,\ldots,k \\ X=[x_1,\ldots,x_k] \\ \operatorname{rank}(X)=k}} \Phi(\lambda_{1,X}^+, \ldots, \lambda_{k,X}^+) \leqslant \Phi(\lambda_{i_1}^+, \ldots, \lambda_{i_k}^+).$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

Since $\mathcal{X}_j$ are arbitrary, we conclude that

$$\sup_{\substack{\mathcal{X}_1\supset\cdots\supset\mathcal{X}_k\\ \operatorname{codim}\mathcal{X}_j=i_j-1}} \inf_{\substack{x_j\in\mathcal{X}_j,\ j=1,\ldots,k\\ X=[x_1,\ldots,x_k]\\ \operatorname{rank}(X)=k}} \Phi(\lambda^+_{1,X},\ldots,\lambda^+_{k,X}) \leqslant \Phi(\lambda^+_{i_1},\ldots,\lambda^+_{i_k}). \tag{5.16}$$

Combine (5.15) and (5.16) to get

$$\sup_{\substack{\mathcal{X}_1\supset\cdots\supset\mathcal{X}_k\\ \operatorname{codim}\mathcal{X}_j=i_j-1}} \inf_{\substack{x_j\in\mathcal{X}_j,\ j=1,\ldots,k\\ X=[x_1,\ldots,x_k]\\ \operatorname{rank}(X)=k}} \Phi(\lambda^+_{1,X},\ldots,\lambda^+_{k,X}) = \Phi(\lambda^+_{i_1},\ldots,\lambda^+_{i_k}). \tag{5.12b$'$}$$

But the sup here is achievable by the selection in (5.14). Thus we have (5.12b).

Now we claim that the inf can be replaced by min for a continuous Φ. Let $\mathcal{X}_j$ for $j = 1,\ldots,k$ be given, and let them satisfy the following assumptions: $\mathcal{X}_1 \supset \cdots \supset \mathcal{X}_k$ and $\operatorname{codim}\mathcal{X}_j = i_j - 1$. There exists a sequence of $X^{(i)} \in \mathbb{C}^{n\times k}$ with $\operatorname{rank}(X^{(i)}) = k$ and its jth column in $\mathcal{X}_j$ such that

$$\lim_{i\to\infty} \Phi(\lambda^+_{1,X^{(i)}},\ldots,\lambda^+_{k,X^{(i)}}) = \inf_{\substack{x_j\in\mathcal{X}_j,\ j=1,\ldots,k\\ X=[x_1,\ldots,x_k]\\ \operatorname{rank}(X)=k}} \Phi(\lambda^+_{1,X},\ldots,\lambda^+_{k,X}). \tag{5.17}$$

Without loss of generality, we may assume that $X^{(i)}$ has A-orthonormal columns; that is,

$$(X^{(i)})^{\mathrm{H}} A X^{(i)} = I_k;$$

otherwise we can perform the Gram–Schimdt A-orthogonalization on the columns of $X^{(i)}$ from the last column backwards, and the new $X^{(i)}$ has the same property as the old $X^{(i)}$: $\operatorname{rank}(X^{(i)}) = k$ and its jth column is in $\mathcal{X}_j$, and also $\lambda^{\pm}_{j,X^{(i)}}$ remain the same. Since $\{X^{(i)}\}$ is a bounded set in $\mathbb{C}^{n\times k}$, it has a convergent subsequence. Through renaming, we may assume that $\{X^{(i)}\}$ itself is convergent, and let $Y \in \mathbb{C}^{n\times k}$ be the limit. It is not hard to see that $Y^{\mathrm{H}}AY = I_k$, which implies that $\operatorname{rank}(Y) = k$ and that the jth column of Y is in $\mathcal{X}_j$. Since $(X^{(i)})^{\mathrm{H}}\boldsymbol{Q}(\lambda)X^{(i)}$ goes to $Y^{\mathrm{H}}\boldsymbol{Q}(\lambda)Y$, by the continuity of eigenvalues with respect to the coefficient matrices we conclude that

$$\lim_{i\to\infty} \lambda^{\pm}_{j,X^{(i)}} = \lambda^{\pm}_{j,Y} \quad \text{for } 1 \leqslant j \leqslant k.$$

Therefore the left-hand side of (5.17) is equal to $\Phi(\lambda^+_{1,Y},\ldots,\lambda^+_{k,Y})$, and thus the inf in (5.17) is attainable.

For (5.12a), a proof similar to what we did above for (5.12b) works: choosing $\widehat{\mathcal{X}}_j = \operatorname{span}\{u^+_1,\ldots,u^+_{i_j}\}$ will lead to the left-hand side being no bigger than its right-hand side, and choosing $\mathcal{Y}_j = \operatorname{span}\{u^+_{i_j},\ldots,u^+_n\}$ will give the opposite. □

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

Specializing Theorem 5.3 to the case where Φ and Ψ are the sums of its arguments leads to Wielandt–Lidskii type min–max principles, as summarized in the following theorem, and Fan type trace min/max principles.

THEOREM 5.4. *Let* $1 \leqslant i_1 < \cdots < i_k \leqslant n$ *and* $\mathrm{typ} \in \{+, -\}$. *Then*

$$\min_{\substack{\mathcal{X}_1 \subset \cdots \subset \mathcal{X}_k \\ \dim \mathcal{X}_j = i_j}} \max_{\substack{x_j \in \mathcal{X}_j \\ X=[x_1,\ldots,x_k] \\ \mathrm{rank}(X)=k}} \sum_{j=1}^{k} \lambda_{j,X}^{\mathrm{typ}} = \sum_{j=1}^{k} \lambda_{i_j}^{\mathrm{typ}}, \tag{5.18a}$$

$$\max_{\substack{\mathcal{X}_1 \supset \cdots \supset \mathcal{X}_k \\ \mathrm{codim}\, \mathcal{X}_j = i_j - 1}} \min_{\substack{x_j \in \mathcal{X}_j \\ X=[x_1,\ldots,x_k] \\ \mathrm{rank}(X)=k}} \sum_{j=1}^{k} \lambda_{j,X}^{\mathrm{typ}} = \sum_{j=1}^{k} \lambda_{i_j}^{\mathrm{typ}}. \tag{5.18b}$$

In particular, setting $i_j = j$ *in* (5.18a) *and setting* $i_j = j + n - k$ *in* (5.18b) *gives*

$$\min_{\mathrm{rank}(X)=k} \sum_{j=1}^{k} \lambda_{j,X}^{\mathrm{typ}} = \sum_{j=1}^{k} \lambda_{j}^{\mathrm{typ}}, \qquad \max_{\mathrm{rank}(X)=k} \sum_{j=1}^{k} \lambda_{j,X}^{\mathrm{typ}} = \sum_{j=1}^{k} \lambda_{n-k+j}^{\mathrm{typ}}. \tag{5.19}$$

5.3. Cauchy type interlacing inequalities. The Cauchy type interlacing inequalities in (5.20) were recently obtained by Veselić [66]. Here we present a simple proof, using our generalizations of Amir-Moéz type min–max principles in Theorem 5.3.

THEOREM 5.5 (Cauchy type interlacing inequalities [66]). *Suppose that* $X \in \mathbb{C}^{n\times k}$ *with* $\mathrm{rank}(X) = k$. *Denote the eigenvalues of* $X^{\mathrm{H}} \boldsymbol{Q}(\lambda) X$ *by*

$$\mu_1^- \leqslant \cdots \leqslant \mu_k^- < \mu_1^+ \leqslant \cdots \leqslant \mu_k^+.$$

Let $\mathrm{typ} \in \{+, -\}$. *Then*

$$\lambda_i^{\mathrm{typ}} \leqslant \mu_i^{\mathrm{typ}} \leqslant \lambda_{i+n-k}^{\mathrm{typ}} \quad \textit{for } i = 1, \ldots, k. \tag{5.20}$$

Proof. Let

$$\Phi(\alpha_1, \ldots, \alpha_k) = \text{the } i\text{th largest } \alpha_j.$$

Then this Φ satisfies the condition of Theorem 5.3. Making use of (5.13a) and (5.13b) gives $\mu_i^{\mathrm{typ}} \geqslant \lambda_i^{\mathrm{typ}}$ and $\mu_i^{\mathrm{typ}} \leqslant \lambda_{i+n-k}^{\mathrm{typ}}$, respectively. That is (5.20). □

REMARK 5.1. The Cauchy type interlacing inequalities in Theorem 5.5 are sharper than those possibly derived by linearization. Actually, through linearization and by item 1 of [40, Theorem 1.1] (which is, in fact, [30, Theorem 2.1]),

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

we can only obtain

$$\lambda_i^+ \leqslant \mu_i^+ \leqslant \lambda_{i+2n-2k}^+, \quad i = 1, \ldots, k,$$
$$\lambda_{j-(n-k)}^- \leqslant \mu_j^- \leqslant \lambda_{j+n-k}^-, \quad j = 1, \ldots, k,$$

where we set $\lambda_i^+ = +\infty$ for $i > n$ and $\lambda_j^- = -\infty$ for $j < 1$.

6. Perturbation analysis

Throughout this section, we suppose that Hermitian matrices A, B, and C are perturbed to Hermitian matrices $\widetilde{A}$, $\widetilde{B}$, and $\widetilde{C}$, and set

$$\Delta A = \widetilde{A} - A, \quad \Delta B = \widetilde{B} - B, \quad \Delta C = \widetilde{C} - C. \tag{6.1}$$

This notational convention of placing a $\sim$ over a symbol for the corresponding perturbed quantity and a Δ before a symbol for the change in the quantity will be generalized to all quantities that depend on A, B, and C. For example, $\boldsymbol{Q}(\lambda) = \lambda^2 A + \lambda B + C$ is perturbed to $\widetilde{\boldsymbol{Q}}(\lambda) = \lambda^2 \widetilde{A} + \lambda \widetilde{B} + \widetilde{C}$, as a result, and

$$\Delta\rho_\pm(x) = \frac{-(x^{\mathrm{H}}\widetilde{B}x) \pm \left[(x^{\mathrm{H}}\widetilde{B}x)^2 - 4(x^{\mathrm{H}}\widetilde{A}x)(x^{\mathrm{H}}\widetilde{C}x)\right]^{1/2}}{2(x^{\mathrm{H}}\widetilde{A}x)} - \frac{-(x^{\mathrm{H}}Bx) \pm \left[(x^{\mathrm{H}}Bx)^2 - 4(x^{\mathrm{H}}Ax)(x^{\mathrm{H}}Cx)\right]^{1/2}}{2(x^{\mathrm{H}}Ax)}.$$

Given a shift $\lambda_0 \in \mathbb{R}$, define

$$\boldsymbol{Q}_{\lambda_0}(\lambda) := \boldsymbol{Q}(\lambda + \lambda_0) = \lambda^2 A + \lambda(2\lambda_0 A + B) + \boldsymbol{Q}(\lambda_0) \tag{6.2a}$$
$$= \lambda^2 A + \lambda B_{\lambda_0} + C_{\lambda_0}, \tag{6.2b}$$

where

$$B_{\lambda_0} = 2\lambda_0 A + B, \quad C_{\lambda_0} = \boldsymbol{Q}(\lambda_0). \tag{6.2c}$$

It can be verified that (μ, x) *is an eigenpair of* $\boldsymbol{Q}_{\lambda_0}(\lambda)$ *if and only if* $(\mu + \lambda_0, x)$ *is an eigenpair of* $\boldsymbol{Q}(\lambda)$.

6.1. Asymptotical analysis. It is a common technique to perform an asymptotical analysis in numerical analysis for at least three reasons:

(1) it is mathematically sound, provided it is known that the interesting quantities are continuous with respect to what is being perturbed;

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

(2) it is relatively easy because it is a first-order analysis; and

(3) it is powerful in revealing the intrinsic sensitivity of the interesting quantities.

Let (μ, x) be a simple eigenpair of HQEP (1.1) for $\boldsymbol{Q}(\lambda)$. Since HQEP (1.1) is equivalent to the eigenvalue problem for the regular matrix pencil $\mathscr{L}_{\boldsymbol{Q}}(\lambda)$ in (4.1), and since the eigenvalues of a regular matrix pencil and the eigenvectors associated with simple eigenvalues are continuous with respect to the entries of the involved matrices [58], $\widetilde{\boldsymbol{Q}}(\lambda)$ has a simple eigenpair $(\tilde{\mu}, \tilde{x}) = (\mu + \Delta\mu, x + \Delta x)$ such that $\Delta\mu \to 0$ and $\Delta x \to 0$ as ΔA, ΔB, $\Delta C \to 0$. Now suppose that $\|\Delta A\|$, $\|\Delta B\|$, and $\|\Delta C\|$ are sufficiently tiny, and so are $\Delta\mu$ and $\|\Delta x\|$. Ignoring terms of order 2 or higher, and noticing that $\boldsymbol{Q}(\mu)x = 0$, we have, from $\widetilde{\boldsymbol{Q}}(\mu + \Delta\mu)(x + \Delta x) = 0$,

$$\Delta\mu[2\mu A + B]x + [\mu^2\Delta A + \mu\Delta B + \Delta C]x + [\mu^2 A + \mu B + C]\Delta x \approx 0, \quad (6.3)$$

where the $\approx$ means that the equation is true after ignoring terms of order 2 or higher. Premultiply (6.3) by x^{H} and use $x^{\mathrm{H}}\boldsymbol{Q}(\mu) = 0$ to get

$$\Delta\mu \approx -\frac{x^{\mathrm{H}}[\mu^2\Delta A + \mu\Delta B + \Delta C]x}{x^{\mathrm{H}}[2\mu A + B]x} \quad (6.4)$$

$$= -\frac{x^{\mathrm{H}}[\mu^2\Delta A + \mu\Delta B + \Delta C]x}{\varsigma(x)} \quad (6.5)$$

$$= -\frac{\mu^2}{\pm\varsigma(x)} \cdot x^{\mathrm{H}}\Delta A x - \frac{\mu}{\pm\varsigma(x)} \cdot x^{\mathrm{H}}\Delta B x - \frac{1}{\pm\varsigma(x)} \cdot x^{\mathrm{H}}\Delta C x, \quad (6.6)$$

where the equality in (6.5) is due to (5.4). There is a clear interpretation of (6.6): the change $\Delta\mu$ is proportional to ΔA, ΔB, ΔC with multiplying factors $|\mu^2/\varsigma(x)|$, $|\mu/\varsigma(x)|$, and $1/|\varsigma(x)|$, respectively. Our following strict bounds reflect this interpretation.

The expression (6.4) is not new, and its derivation follows a rather standard technique (see, for example, [63]). What is new here is the use of (5.4) to relate its denominator $x^{\mathrm{H}}[2\mu A + B]x$ to $\varsigma(x)$, a quantity that determines the hyperbolicity of $\boldsymbol{Q}$.

6.2. Perturbation bounds in the spectral norm. Throughout the rest of this section, we assume that $\boldsymbol{Q}(\lambda)$ and $\widetilde{\boldsymbol{Q}}(\lambda)$ are hyperbolic and that

$$\|A^{-1/2}\Delta A A^{-1/2}\|_2 < 1, \quad (6.7)$$

which guarantees that $\widetilde{A} \succ 0$. We will adopt the notation introduced in Theorem 4.2. Our goal is to bound the norms of

$$\Delta\Lambda_+ = \mathrm{diag}(\tilde{\lambda}_1^+ - \lambda_1^+, \ldots, \tilde{\lambda}_n^+ - \lambda_n^+), \quad \Delta\Lambda_- = \mathrm{diag}(\tilde{\lambda}_1^- - \lambda_1^-, \ldots, \tilde{\lambda}_n^- - \lambda_n^-).$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

Bounds on norms of the change to $\Lambda = \mathrm{diag}(\Lambda_-, \Lambda_+)$ are easily derivable through

$$\|\Delta\Lambda\|_2 = \max_{\pm} \|\Delta\Lambda_\pm\|_2, \quad \|\Delta\Lambda\|_{\mathrm{F}} = \sqrt{\|\Delta\Lambda_+\|_{\mathrm{F}}^2 + \|\Delta\Lambda_-\|_{\mathrm{F}}^2},$$
$$\|\Delta\Lambda\|_{\mathrm{ui}} \leqslant 2 \max_{\pm} \|\Delta\Lambda_\pm\|_{\mathrm{ui}},$$

where $\|\cdot\|_{\mathrm{ui}}$ denotes a general unitarily invariant norm. For the definition and properties of unitarily invariant norms, the reader is referred to [7, 58] for details. In this article, for convenience, any $\|\cdot\|_{\mathrm{ui}}$ we use is generic to matrix sizes in the sense that it applies to matrices of all sizes. Examples include the matrix spectral norm $\|\cdot\|_2$ and the Frobenius norm $\|\cdot\|_{\mathrm{F}}$. Two important properties of unitarily invariant norms are

$$\|X\|_2 \leqslant \|X\|_{\mathrm{ui}}, \quad \|XYZ\|_{\mathrm{ui}} \leqslant \|X\|_2 \cdot \|Y\|_{\mathrm{ui}} \cdot \|Z\|_2 \tag{6.8}$$

for any matrices X, Y, and Z of compatible sizes.

In this subsection, we will focus on the spectral norm, and leave the case for more generally unitarily invariant norms to the next subsection. Our main results of this subsection are summarized in Theorem 6.1, which is reminiscent of the well-known result of Weyl [71]. We will comment more on it after stating the theorem.

THEOREM 6.1. *Let* $\mathrm{typ} \in \{+, -\}$, *and*

$$\epsilon_a = \|A^{-1/2}\Delta A A^{-1/2}\|_2, \quad \epsilon_b = \frac{\|\Delta B\|_2}{\|B\|_2}, \quad \epsilon_c = \frac{\|\Delta C\|_2}{\|C\|_2}, \tag{6.9}$$
$$\lambda_{\max}^{\mathrm{typ}} = \max\{|\lambda_1^{\mathrm{typ}}|, |\lambda_n^{\mathrm{typ}}|\}, \quad \tilde{\lambda}_{\max}^{\mathrm{typ}} = \max\{|\tilde{\lambda}_1^{\mathrm{typ}}|, |\tilde{\lambda}_n^{\mathrm{typ}}|\}, \tag{6.10}$$
$$\chi_\varsigma = \min_{x \neq 0}\{\varsigma_0(x), \tilde{\varsigma}_0(x)\}, \quad \chi_{\lambda^{\mathrm{typ}}} = \max\{\lambda_{\max}^{\mathrm{typ}}, \tilde{\lambda}_{\max}^{\mathrm{typ}}\}. \tag{6.11}$$

(1) *If* $\Delta A = \Delta B = 0$ *and*

$$\epsilon_c < \frac{\chi_\varsigma^2}{4\|A\|_2\|C\|_2}, \tag{6.12}$$

then

$$\|\Delta\Lambda_{\mathrm{typ}}\|_2 \leqslant \frac{1}{\chi_\varsigma}\|\Delta C\|_2. \tag{6.13}$$

(2) *If* $\Delta B = \Delta C = 0$ *and*

$$\epsilon_a < \min\left\{1, \frac{\chi_\varsigma^2}{4\|A\|_2\|C\|_2}\right\}, \tag{6.14}$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

then

$$\|\Delta\Lambda_{\rm typ}\|_2 \leqslant \frac{\chi^2_{\lambda^{\rm typ}}}{(1-\epsilon_a)\chi_\varsigma}\,\|\Delta A\|_2. \tag{6.15}$$

(3) *If* $\Delta A = \Delta C = 0$ *and*

$$\epsilon_b < \frac{\chi^2_\varsigma}{\|B\|_2(\|B\|_2 + 2\sqrt{\|A\|_2\|C\|_2})}, \tag{6.16}$$

then

$$\|\Delta\Lambda_{\rm typ}\|_2 \leqslant \frac{\chi_{\lambda^{\rm typ}}}{\chi_\varsigma}\,\|\Delta B\|_2 + \frac{\|C\|_2}{\chi^3_\varsigma}\,\|\Delta B\|_2^2. \tag{6.17}$$

(4) *If* $\Delta A = \Delta C = 0$ *and*

$$\|\Delta B\|_2 < \frac{\chi^2_\varsigma}{\|2\lambda_0 A + B\|_2 + 2\sqrt{\|A\|_2\|\boldsymbol{Q}(\lambda_0)\|_2}}, \tag{6.18}$$

where $\lambda_0 \in (-\infty, \min\{\lambda_1^-, \tilde{\lambda}_1^-\}] \cup [\max\{\lambda_n^+, \tilde{\lambda}_n^+\}, +\infty)$, *then*

$$\|\Delta\Lambda_{\rm typ}\|_2 \leqslant \frac{\chi_{\lambda^{\rm typ}} + |\lambda_0|}{\chi_\varsigma}\,\|\Delta B\|_2. \tag{6.19}$$

(5) *In general, without assuming that two of* ΔA, ΔB, *and* ΔC *are zero, if*

$$\epsilon_a < \gamma \min\left\{1, \frac{\chi^2_\varsigma}{4\|A\|_2\|C\|_2}\right\}, \tag{6.20a}$$

$$\epsilon_b < \gamma \frac{\chi^2_\varsigma}{\|B\|_2(\|B\|_2 + 2\sqrt{\|A\|_2\|C\|_2})}, \tag{6.20b}$$

$$\epsilon_c < \gamma \frac{\chi^2_\varsigma}{4\|A\|_2\|C\|_2}, \tag{6.20c}$$

where

$$\gamma = \frac{\chi^2_\varsigma}{\|B\|_2^2 + \chi^2_\varsigma + \sqrt{(\|B\|_2^2 + \chi^2_\varsigma)(\|B\|_2^2 + 2\chi^2_\varsigma)}} < \sqrt{2} - 1, \tag{6.21}$$

then

$$\begin{aligned}\|\Delta\Lambda_{\rm typ}\|_2 \leqslant{}& \frac{4}{(1-\epsilon_a)\chi^3_\varsigma}\|C\|_2[\|A\|_2\|C\|_2(\epsilon_a+\epsilon_c)^2 + \|B\|_2^2(\epsilon_b+\epsilon_a)(\epsilon_b+\epsilon_c)] \\ &+ \frac{1}{(1-\epsilon_a)\chi_\varsigma}[(\chi_{\lambda^{\rm typ}})^2\|\Delta A\|_2 + \chi_{\lambda^{\rm typ}}\|\Delta B\|_2 + \|\Delta C\|_2].\end{aligned} \tag{6.22}$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

All bounds by this theorem are strict. They resemble the well-known result of Weyl [71] for the Hermitian eigenvalue problem. Let $H \in \mathbb{C}^{n\times n}$ be a Hermitian matrix which is perturbed to another Hermitian matrix $\widetilde{H} \in \mathbb{C}^{n\times n}$, and denote their eigenvalues by ω_i and $\widetilde{\omega}_j$, respectively, which are arranged in the ascending order as

$$\omega_1 \leqslant \omega_2 \leqslant \cdots \leqslant \omega_n, \quad \tilde{\omega}_1 \leqslant \tilde{\omega}_2 \leqslant \cdots \leqslant \tilde{\omega}_n.$$

Let $\Omega = \operatorname{diag}(\omega_1, \ldots, \omega_n)$ and $\widetilde{\Omega} = \operatorname{diag}(\tilde{\omega}_1, \ldots, \tilde{\omega}_n)$. The well-known result of Weyl [71] (see also [7, page 63], [50, page 208], and [58, page 203]) says that

$$\|\widetilde{\Omega} - \Omega\|_2 \leqslant \|\widetilde{H} - H\|_2. \tag{6.23}$$

Our results in Theorem 6.1 resemble Weyl's result (6.23) in a way that they serve the purpose of bounding the largest possible deviations in the corresponding eigenvalues in terms of the perturbations in the matrices involved. However, ours here contain the quantities defined in (6.10) and (6.11), and these quantities make our bounds look less elegant than (6.23). But we argue that for an HQEP it is in general unavoidable because the results in Theorem 6.1 are consistent with the asymptotic expression (6.6) after dropping terms of order 2 or higher in ϵ_a, ϵ_b, and ϵ_c. For example, (6.22) yields

$$\|\Delta\Lambda_{\mathrm{typ}}\|_2 \lesssim \frac{1}{\chi_\varsigma}[(\chi_{\lambda^{\mathrm{typ}}})^2\|\Delta A\|_2 + \chi_{\lambda^{\mathrm{typ}}}\|\Delta B\|_2 + \|\Delta C\|_2]. \tag{6.24}$$

This inequality is rather sharp asymptotically in general, since the asymptotic expression (6.6) is an equality up to the first order.

Weyl's bound (6.23) is a special case of a much more general perturbation result for the Hermitian eigenvalue problem. In fact, we have

$$\|\widetilde{\Omega} - \Omega\|_{\mathrm{ui}} \leqslant \|\widetilde{H} - H\|_{\mathrm{ui}} \tag{6.25}$$

for any unitarily invariant norm $\|\cdot\|_{\mathrm{ui}}$. This inequality, which we will refer to as the *Wielandt–Lidskii–Mirsky inequality* (or *perturbation theorem*), is a direct consequence of any one of the following: Wielandt's min–max principle [72], Lidskii's theorem on the relationship among eigenvalues of two Hermitian matrices and their sums [41], and Mirsky's perturbation result for singular values [46] (see also [7, page 71], [58, page 205]).

Our proof of Theorem 6.1 is long and involves complicated computations. We defer it to Appendix A.

6.3. Perturbation bounds in unitarily invariant norms. Our main results of this subsection are Theorems 6.2 and 6.3. These results can be viewed as

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

extensions of the Wielandt–Lidskii–Mirsky inequality (6.25) to an HQEP. The proof of Theorem 6.2 is based on our new Wielandt–Lidskii min–max principles. Since it is rather long, we defer it also to Appendix A.

THEOREM 6.2. *Suppose that* $\Delta A = \Delta B = 0$ *and that* (6.12) *holds, and let*

$$\gamma = (\lambda_1^+ - \lambda_n^-)\lambda_{\min}(A), \quad \widetilde{\gamma} = (\tilde{\lambda}_1^+ - \tilde{\lambda}_n^-)\lambda_{\min}(A). \tag{6.26}$$

Then

$$\|\Delta\Lambda_\pm\|_{\mathrm{ui}} \leqslant c \cdot \frac{\|\Delta C\|_{\mathrm{ui}}}{\min\{\gamma, \widetilde{\gamma}\}}, \tag{6.27}$$

where the constant $c = 1$ *if* ΔC *is semidefinite and* $c = 2$ *in general.*

The inequality (6.27) can be considered as an extension of (6.13), but it is a little bit less satisfying in that it does not become (6.13) after specializing the unitarily invariant norm to the spectral norm in two aspects: (1) c is not always 1, and (2)

$$\min_{x\neq 0} \varsigma_0(x) \geqslant \gamma,$$

which can be a strict inequality. It makes us wonder if the stronger version of (6.27) upon setting $c = 1$ always and replacing $\min\{\gamma, \widetilde{\gamma}\}$ by χ_ς holds. But how to settle this question eludes us for now.

Recall the eigen-decomposition in Theorem 4.2 for the linearization $\mathscr{A} - \lambda\mathscr{B}$ of $\boldsymbol{Q}(\lambda)$. The next theorem is a straightforward application of Theorem B.2, where $\|Z\|_2$ and $\|\widetilde{Z}\|_2$ can be bounded, using Theorem 4.4.

THEOREM 6.3. *Let* $\mathscr{A} - \lambda\mathscr{B} = \mathscr{L}_{\boldsymbol{Q}}(\lambda)$ *and* $\widetilde{\mathscr{A}} - \lambda\widetilde{\mathscr{B}} = \mathscr{L}_{\widetilde{\boldsymbol{Q}}}(\lambda)$, *admitting the eigen-decomposition in* (4.8). *Then*

$$\|\widetilde{\Lambda} - \Lambda\|_{\mathrm{ui}} \leqslant \|Z\|_2\|\widetilde{Z}\|_2\big(\|\widetilde{\mathscr{A}} - \mathscr{A}\|_{\mathrm{ui}} + \xi\|\widetilde{\mathscr{B}} - \mathscr{B}\|_{\mathrm{ui}}\big), \tag{6.28}$$

where $\xi = \max\{|\lambda_{\max}^+|, |\lambda_{\max}^-|, |\tilde{\lambda}_{\max}^+|, |\tilde{\lambda}_{\max}^-|\}$, *and* $\lambda_{\max}^\pm$ *and* $\tilde{\lambda}_{\max}^\pm$ *are defined by* (6.10).

7. Best approximations from a subspace and the Rayleigh–Ritz procedure

Two most important aspects in solving a large scale eigenvalue problem are:

(1) building subspaces to which the desired eigenvectors (or invariant subspaces) are close; and

(2) seeking *best possible* approximations from the suitably built subspaces.

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

In this section, we shall address the second aspect for our current problem at hand, that is, seeking *best possible* approximations to a few eigenvalues of $\boldsymbol{Q}(\lambda)$ and their associated eigenvectors from a given subspace of $\mathbb{C}^n$. We leave the first aspect to the later sections when we present our computational algorithms.

The concept of *best possible* comes with a quantitative measure as to what constitutes *best possible*. There may not be such a measure in general. In [**50**, Section 11.4], Parlett uses three different ways to justify the use of the Rayleigh–Ritz procedure for the symmetric eigenvalue problem. For the HQEP here, each of the minimization principles in Section 5 provides a quantitative measure.

Let $\boldsymbol{Q}(\lambda) = \lambda^2 A + \lambda B + C \in \mathbb{C}^{n\times n}$ be a hyperbolic quadratic matrix polynomial, and let $\mathcal{Y} \subset \mathbb{C}^n$ be a subspace of dimension m. We are seeking *best possible* approximations to a few eigenvalues of $\boldsymbol{Q}(\lambda)$ using $\mathcal{Y}$. Let $Y \in \mathbb{C}^{n\times m}$ be a basis matrix of $\mathcal{Y}$.

According to (5.7a), which says (upon substituting $i = n - j + 1$) that

$$\lambda^+_{n-j+1} = \max_{\substack{\mathcal{X}\subseteq\mathbb{C}^n \\ \dim \mathcal{X}=j}} \min_{\substack{x\in\mathcal{X} \\ x\neq 0}} \rho_+(x), \tag{5.7a$'$}$$

it is natural to approximate λ^+_{n-j+1}, given $\mathcal{Y} \subset \mathbb{C}^n$, by

$$\mu^+_{m-j+1} := \max_{\substack{\mathcal{X}\subseteq\mathcal{Y} \\ \dim \mathcal{X}=j}} \min_{\substack{x\in\mathcal{X} \\ x\neq 0}} \rho_+(x), \tag{7.1}$$

via replacing $\mathcal{X} \subseteq \mathbb{C}^n$ in (5.7a′) by $\mathcal{X} \subseteq \mathcal{Y}$. Any nonzero $x \in \mathcal{X} \subseteq \mathcal{Y}$ can be written as $x = Yy$ for some nonzero $y \in \mathbb{C}^m$, and thus

$$\rho_+(x) = \rho_+(Yy) = \frac{-(y^{\mathrm{H}}Y^{\mathrm{H}}BYy) + [(y^{\mathrm{H}}Y^{\mathrm{H}}BYy)^2 - 4(y^{\mathrm{H}}Y^{\mathrm{H}}AYy)(y^{\mathrm{H}}Y^{\mathrm{H}}CYy)]^{1/2}}{2(y^{\mathrm{H}}Y^{\mathrm{H}}AYy)}.$$

Combined with (5.7a′) and this expression for $\rho_+(x)$, (7.1) implies that $\mu^+_1, \ldots, \mu^+_m$ are the m pos-type eigenvalues of $Y^{\mathrm{H}}\boldsymbol{Q}(\lambda)Y$. What this means is that μ^+_j for $1 \leqslant j \leqslant m$ provide the best approximations to the m largest λ^+_j, given $\mathcal{Y}$, in the sense of (5.7a). Of course, some approximations $\mu^+_j \approx \lambda^+_{n-m+j}$ are more accurate than others.

Similarly, given $\mathcal{Y}$, μ^+_j for $1 \leqslant j \leqslant m$ provide the best approximations to the m smallest λ^+_j in the sense of (5.7b).

Let $\mu^-_1, \ldots, \mu^-_m$ be the m neg-type eigenvalues of $Y^{\mathrm{H}}\boldsymbol{Q}(\lambda)Y$. The same argument shows that, given $\mathcal{Y}$, μ^-_j for $1 \leqslant j \leqslant m$ provide the best approximations to the m largest λ^-_j in the sense of (5.7a), and the best approximations to the m smallest λ^-_j in the sense of (5.7b).

In summary, we have justified that the eigenvalues of $Y^{\mathrm{H}}\boldsymbol{Q}(\lambda)Y$ yield the best approximations to some of the largest or smallest pos-type or neg-type

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

Algorithm 7.1 Rayleigh–Ritz procedure

Given $Y \in \mathbb{C}^{n\times m}$ which is a basis matrix of $\mathcal{Y} \subset \mathbb{C}^n$, this algorithm returns approximations to k extreme eigenpairs (of pos-type or neg-type) of $\boldsymbol{Q}(\lambda)$.

1: solve the HQEP for $Y^{\mathrm{H}}\boldsymbol{Q}(\lambda)Y$ to get its eigenvalues $\mu_j^{\pm}$ and associated eigenvectors $y_j^{\pm}$.
2: **return**

- $(\mu_i^{\pm}, Yy_i^{\pm})$ for $1 \leqslant i \leqslant k$ as approximations to $(\lambda_i^{\pm}, u_i^{\pm})$ for $1 \leqslant i \leqslant k$, or
- $(\mu_i^{\pm}, Yy_i^{\pm})$ for $m-k+1 \leqslant i \leqslant m$ as approximations to $(\lambda_i^{\pm}, u_i^{\pm})$ for $n-k+1 \leqslant i \leqslant n$,

depending on what kind of extreme eigenpairs are desired.

eigenvalues of $\boldsymbol{Q}(\lambda)$ in certain respective senses. This statement may sound confusing: how could the same set of values be the best approximations to some of both largest and smallest eigenvalues at the same time? But we point out that this is not what the statement is saying. The key to understanding the subtlety is not to forget that they provide the best approximations under the mentioned senses, and being the best approximations (under a particular sense) does not necessarily imply that the approximations are good, just that they are the best (under that particular sense). In practice, $\mathcal{Y}$ is built to approximate either the largest or smallest eigenvalues well, as in the case of optimization methods in Sections 8–11.

Theorems 5.3 and 5.4, generalizing Amir-Moéz's min–max principles and Wielandt–Lidskii min–max principles, can also be used to justify that the eigenvalues of $Y^{\mathrm{H}}\boldsymbol{Q}(\lambda)Y$ are candidates for best approximating the largest or smallest pos-type or neg-type eigenvalues of $\boldsymbol{Q}(\lambda)$, too. For example, according to (5.13a) with any prechosen Φ, we should seek best approximations to λ_i^+ for $1 \leqslant i \leqslant k$ by

$$\text{minimizing } \Phi(\lambda_{1,X}^+, \ldots, \lambda_{k,X}^+) \text{ subject to } \mathcal{R}(X) \subseteq \mathcal{Y} \text{ and } \operatorname{rank}(X) = k. \quad (5.13a')$$

Noticing that any $X \in \mathbb{C}^{n\times k}$ satisfying $\mathcal{R}(X) \subseteq \mathcal{Y}$ and $\operatorname{rank}(X) = k$ can be written as $X = Y\hat{X}$ for some $\hat{X} \in \mathbb{C}^{m\times k}$ with $\operatorname{rank}(\hat{X}) = k$, we see that $\lambda_{j,X}^+$ are pos-type eigenvalues of $[Y\widehat{X}]^{\mathrm{H}}\boldsymbol{Q}(\lambda)[Y\widehat{X}] = \widehat{X}^{\mathrm{H}}Y^{\mathrm{H}}\boldsymbol{Q}(\lambda)Y\widehat{X}$. Varying X subject to $\mathcal{R}(X) \subseteq \mathcal{Y}$ and $\operatorname{rank}(X) = k$ is transferred to varying $\hat{X} \in \mathbb{C}^{m\times k}$ subject to $\operatorname{rank}(\hat{X}) = k$. Consequently,

$$\min_X \Phi(\lambda_{1,X}^+, \ldots, \lambda_{k,X}^+) = \min_{\widehat{X}} \Phi(\mu_{1,\widehat{X}}^+, \ldots, \mu_{k,\widehat{X}}^+), \quad (7.2)$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

where $\mu^+_{j,\widehat{X}}$ are pos-type eigenvalues of $\widehat{X}^{\mathrm{H}} Y^{\mathrm{H}} \boldsymbol{Q}(\lambda) Y \widehat{X}$. Apply Theorem 5.3 to see that the right-hand side of (7.2) is $\Phi(\mu^+_1, \ldots, \mu^+_k)$, indicating that μ^+_j for $1 \leqslant j \leqslant k$ provide the best approximations to the k smallest λ^+_j, as expected.

The same statement can be made about μ^+_j as approximations to the largest λ^+_j, and μ^-_j as approximations to the smallest λ^-_j or as approximations to the largest λ^-_j, using other min–max principles in Theorems 5.3 and 5.4.

In summary, our discussion so far has led to a Rayleigh–Ritz type procedure detailed in Algorithm 7.1 to compute the best approximations to the desired eigenpairs of $\boldsymbol{Q}(\lambda)$, given a prebuilt subspace $\mathcal{Y}$.

8. The steepest descent/ascent method

A common approach to solve a quadratic eigenvalue problem in general, as well as any polynomial eigenvalue problem, is through *linearization*, which converts the problem into a linear generalized eigenvalue problem of a matrix pencil [25, 43, 44]. The latter can be solved either by some iterative methods for a large scale problem or by the QZ algorithm [3, 47] for a problem of small to modest size (n up to around a few thousands, for example). This approach is usually adopted for general QEPs that have no favorable structure to exploit. For an HQEP, however, it is a different story—there is much to exploit. Most recent development includes the solvent approach [11, 21, 24, 64] for certain kinds of QEP among which the HQEP [23] is one. Numerical evidence indicates that this solvent approach is rather efficient for QEPs of small to modest size.

In this paper, we focus on optimization approaches based on various min–max principles previously established and the new ones established here. They are iterative methods and intended for solving large-scale HQEPs.

The equations in (5.8),

$$\lambda^{\mathrm{typ}}_1 = \min_{x \neq 0} \rho_{\mathrm{typ}}(x), \quad \lambda^{\mathrm{typ}}_n = \max_{x \neq 0} \rho_{\mathrm{typ}}(x) \tag{5.8}$$

where $\mathrm{typ} \in \{+, -\}$, naturally suggest using some optimization techniques, including the steepest descent/ascent or CG type methods, to compute the first or last eigenpair $(\lambda^{\mathrm{typ}}_j, u^{\mathrm{typ}}_j)$ as in the case of the standard Hermitian eigenvalue problem [4, 15]. Block variations can also be devised to simultaneously compute the first or last few eigenpairs $(\lambda^{\mathrm{typ}}_j, u^{\mathrm{typ}}_j)$, again as in the case of the standard Hermitian eigenvalue problem [4, 42].

8.1. Gradients. To apply any of optimization techniques, we need to compute the gradients of $\rho_\pm(x)$. To this end, we use $\rho(x)$ for either $\rho_+(x)$ or $\rho_-(x)$. As x is perturbed to $x + p$, where p is assumed small in magnitude, $\rho(x + p)$ is changed

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

to $\rho(x+p)=\rho(x)+\eta$, where the magnitude η is comparable to $\|p\|$. We have, by (5.1),

$$[\rho(x)+\eta]^2\,(x+p)^{\mathrm{H}}A(x+p)+[\rho(x)+\eta]\,(x+p)^{\mathrm{H}}B(x+p)+(x+p)^{\mathrm{H}}C(x+p)=0,$$

which gives, upon noticing that $f(\rho(x),x)=0$,

$$\begin{aligned}[2\rho(x)\,x^{\mathrm{H}}Ax+x^{\mathrm{H}}Bx]\eta+p^{\mathrm{H}}[\rho(x)^2Ax+\rho(x)Bx+Cx]\\ +[\rho(x)^2Ax+\rho(x)Bx+Cx]^{\mathrm{H}}p+O(\|p\|^2)=0,\end{aligned}$$

and thus

$$\eta=-\frac{p^{\mathrm{H}}[\rho(x)^2Ax+\rho(x)Bx+Cx]+[\rho(x)^2Ax+\rho(x)Bx+Cx]^{\mathrm{H}}p}{2\rho(x)\,x^{\mathrm{H}}Ax+x^{\mathrm{H}}Bx}+O(\|p\|^2).$$

Therefore the gradient of $\rho(x)$ at x is

$$\nabla\rho(x)=-\frac{2[\rho(x)^2A+\rho(x)B+C]x}{2\rho(x)\,x^{\mathrm{H}}Ax+x^{\mathrm{H}}Bx},$$

or equivalently

$$\nabla\rho_{\pm}(x)=\mp\frac{2\,\boldsymbol{Q}(\rho_{\pm}(x))x}{\varsigma(x)}, \tag{8.1}$$

where we have used (5.4).

It is important to notice that the gradient $\nabla\rho_{\pm}(x)$ is parallel to the residual vector

$$r_{\pm}(x):=[\rho_{\pm}(x)^2A+\rho_{\pm}(x)B+C]x=\boldsymbol{Q}(\rho_{\pm}(x))x, \tag{8.2}$$

whose normalized norm is commonly used to determine if the approximate eigenpair $(\rho_{\pm}(x),x)$ meets a preset tolerance `rtol`:

$$\frac{\|r_{\pm}(x)\|}{|\rho_{\pm}(x)|^2\|Ax\|+|\rho_{\pm}(x)|\,\|Bx\|+\|Cx\|}\leqslant \texttt{rtol}. \tag{8.3}$$

If (8.3) holds for $(\rho_+(x),x)$, then it is accepted as a converged pos-type eigenpair, and similarly for $(\rho_-(x),x)$. Here which vector norm $\|\cdot\|$ to use is usually inconsequential. More conservatively, $\|Ax\|$ in the denominator should be replaced by $\|A\|\,\|x\|$, and likewise for $\|Bx\|$ and $\|Cx\|$ there. For large sparse matrices, the use of $\|Ax\|$, $\|Bx\|$, and $\|Cx\|$ is more economical because of their availability.

Beside being easily implementable, the use of (8.3) can also be rationalized by the existing backward error analysis of approximate eigenpairs for polynomial eigenvalue problems [**25**, **37**, **63**].

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

8.2. The steepest descent/ascent method. Now the steepest descent/ascent method for computing one of $\lambda_\ell^\pm$ for $\ell \in \{1, n\}$ can be readily given. For this purpose, we fix two parameters typ and ℓ with varying ranges as

$$\text{typ} \in \{+, -\}, \quad \ell \in \{1, n\} \tag{8.4}$$

to mean that we are to compute the eigenpair $(\lambda_\ell^{\text{typ}}, u_\ell^{\text{typ}})$. A key step of the method is the following line-search problem:

$$t_{\text{opt}} = \arg\,\text{opt}_{t\in\mathbb{C}}\ \rho_{\text{typ}}(x + t\ p), \tag{8.5}$$

where x is the current approximation to u_ℓ^{typ} (thus there is no reason to let $x = 0$), p is the search direction, and

$$\arg\,\text{opt} = \begin{cases} \arg\min & \text{for } \ell = 1, \\ \arg\max & \text{for } \ell = n. \end{cases} \tag{8.6}$$

The next approximate eigenvector is

$$y = \begin{cases} x + t_{\text{opt}}\ p & \text{if } t_{\text{opt}} \text{ is finite}, \\ p & \text{otherwise.} \end{cases} \tag{8.7}$$

But the line-search problem (8.5) does not seem to be solvable straightforwardly by simple calculus as for the standard symmetric eigenvalue problem (see, for example, [**4**, **15**, **42**, **73**]), given the (complicated) expressions for ρ_{typ} in (5.2). Fortunately, the theory we developed in Section 7 gives us another way to look at it and thus solve it with ease. In fact, the problem is equivalent to finding the best possible approximation within the subspace $\mathcal{Y} = \mathcal{R}([x, p])$. Suppose that x and p are linearly independent (otherwise, no improvement is expected by optimizing $\rho_{\text{typ}}(x+tp)$ because then $\rho_{\text{typ}}(x+tp) \equiv \rho_{\text{typ}}(x)$ for all scalar t), and let $Y = [x, p]$. Solve the order-2 HQEP for $Y^{\text{H}}\boldsymbol{Q}(\lambda)Y$ to get its eigenvalues

$$\mu_1^- \leqslant \mu_2^- < \mu_1^+ \leqslant \mu_2^+ \tag{8.8}$$

and corresponding eigenvectors $y_j^\pm \in \mathbb{C}^2$. We then have the following table for selecting the next approximate eigenpair, according to the parameter pair (typ, ℓ).

(typ, ℓ)	current approx.	next approx.
(typ, 1)	$(\rho_{\text{typ}}(x), x)$	$(\mu_1^{\text{typ}}, Y y_1^{\text{typ}})$
(typ, n)	$(\rho_{\text{typ}}(x), x)$	$(\mu_2^{\text{typ}}, Y y_2^{\text{typ}})$

(8.9)

In light of this alternative way to solve (8.5), the resulting steepest descent/ascent method is summarized in Algorithm 8.1.

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

Algorithm 8.1 Steepest descent/ascent method

Given an initial approximation $\boldsymbol{x}_0$ to u_ℓ^{typ}, and a relative tolerance `rtol`, the algorithm computes an approximate pair to $(\lambda_\ell^{\text{typ}}, u_\ell^{\text{typ}})$ with the prescribed `rtol`.

$\boldsymbol{x}_0 = \boldsymbol{x}_0/\|\boldsymbol{x}_0\|$, $\boldsymbol{\rho}_0 = \rho_{\text{typ}}(\boldsymbol{x}_0)$, $\boldsymbol{r}_0 = r_{\text{typ}}(\boldsymbol{x}_0)$;
for $i = 0, 1, \ldots$ **do**
  if $\|\boldsymbol{r}_i\|/(|\boldsymbol{\rho}_i|^2\|A\boldsymbol{x}_i\| + |\boldsymbol{\rho}_i|\,\|B\boldsymbol{x}_i\| + \|C\boldsymbol{x}_i\|) \leqslant$ `rtol` **then**
    BREAK;
  else
    solve the HQEP for $Y_i^{\text{H}}\boldsymbol{Q}(\lambda)Y_i$, where $Y_i = [\boldsymbol{x}_i, \boldsymbol{r}_i]$ to get its eigenvalues $\mu_j^{\pm}$ as in (8.8) and corresponding eigenvectors $y_j^{\pm}$;
    select the next approximate eigenpair $(\mu, y) = (\mu_j^{\text{typ}}, Y_i y_j^{\text{typ}})$ according to (8.9);
    $\boldsymbol{x}_{i+1} = y/\|y\|$, $\boldsymbol{\rho}_{i+1} = \mu$, $\boldsymbol{r}_{i+1} = r_{\text{typ}}(\boldsymbol{x}_{i+1})$;
  end if
end for
return $(\boldsymbol{\rho}_i, \boldsymbol{x}_i)$ as an approximate eigenpair to $(\lambda_\ell^{\text{typ}}, u_\ell^{\text{typ}})$.

LEMMA 8.1. *For* (8.5)–(8.7), $p^{\text{H}} r_{\text{typ}}(y) = 0$.

Proof. If x and p are linearly dependent (the trivial case $p = 0$ included), then $p = \alpha x$ and $y = \beta x$ for some scalars α and β. Thus $\rho_{\text{typ}}(y) = \rho_{\text{typ}}(x)$, $r_{\text{typ}}(y) = \beta r_{\text{typ}}(x)$, and $p^{\text{H}} r_{\text{typ}}(y) = \alpha\beta x^{\text{H}} r_{\text{typ}}(x) = 0$ by the definition of $\rho_{\text{typ}}(x)$.

Suppose that x and p are linearly independent. If $|t_{\text{opt}}| = \infty$, then $y = p$. Thus $p^{\text{H}} r_{\text{typ}}(y) = y^{\text{H}} r_{\text{typ}}(y) = 0$. Consider the case that t_{opt} is finite. Let $t = t_{\text{opt}} + s$. For tiny s, we have

$$\rho(y + sp) = \rho(y) - \frac{2\text{RE}\,(s[\rho(y)^2 Ay + \rho(y)By + Cy]^{\text{H}} p)}{2\rho(y)\, y^{\text{H}} Ay + y^{\text{H}} By} + O(s^2),$$

where we drop the subscript typ in $\rho_{\text{typ}}(\,\cdot\,)$ for convenience. Since $\min_s \rho(y + sp)$ over $s \in \mathbb{C}$ is attained at $s = 0$, it must hold that $[\rho(y)^2 Ay + \rho(y)By + Cy]^{\text{H}} p = 0$, as was to be shown. □

8.3. The extended steepest descent/ascent method. In Algorithm 8.1, the search space is spanned by

$$\boldsymbol{x}_i, \quad \boldsymbol{r}_i = \boldsymbol{Q}(\boldsymbol{\rho}_i)\boldsymbol{x}_i.$$

Thus it is the second-order Krylov subspace $\mathcal{K}_2(\boldsymbol{Q}(\boldsymbol{\rho}_i), \boldsymbol{x}_i)$ of $\boldsymbol{Q}(\boldsymbol{\rho}_i)$ on $\boldsymbol{x}_i$. Inspired by the inverse free Krylov subspace method [19], which seeks to improve

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

the steepest descent method for the Hermitian generalized eigenvalue problem by extending the search space to a higher-order Krylov subspace, we may improve Algorithm 8.1 in the same way, that is, using a high-order Krylov subspace

$$\mathcal{K}_m(\boldsymbol{Q}(\rho_i), \boldsymbol{x}_i) = \operatorname{span}\{\boldsymbol{x}_i, \boldsymbol{Q}(\rho_i)\boldsymbol{x}_i, \ldots, [\boldsymbol{Q}(\rho_i)]^{m-1}\boldsymbol{x}_i\} \tag{8.10}$$

as the search space. Let Y_i be a basis matrix of this Krylov subspace. We then solve the order-m HQEP for $Y_i^{\mathrm{H}}\boldsymbol{Q}(\lambda)Y_i$ to get its eigenvalues

$$\mu_1^- \leqslant \cdots \leqslant \mu_m^- < \mu_1^+ \leqslant \cdots \leqslant \mu_m^+ \tag{8.11}$$

and corresponding eigenvectors $y_j^{\pm}$. (Often $Y_i \in \mathbb{C}^{n\times m}$, but there is a possibility that $\dim \mathcal{K}_m(\boldsymbol{Q}(\rho_i), \boldsymbol{x}_i) < m$. When this occurs, Y_i will have fewer columns than m, and the rest of the development is still valid with minor changes. This is rare, especially in actual computations. For simplicity of presentation, we will assume that Y_i has m columns.) We then have the following table for selecting the next approximate eigenpair, according to the parameter pair (typ, ℓ).

(typ, ℓ)	currentapprox.	nextapprox.
(typ, 1)	$(\rho_{\mathrm{typ}}(\boldsymbol{x}_i), \boldsymbol{x}_i)$	$(\mu_1^{\mathrm{typ}}, Y_i y_1^{\mathrm{typ}})$
(typ, n)	$(\rho_{\mathrm{typ}}(\boldsymbol{x}_i), \boldsymbol{x}_i)$	$(\mu_m^{\mathrm{typ}}, Y_i y_m^{\mathrm{typ}})$

(8.12)

We summarize the resulting method, called the *Extended Steepest Descent/Ascent method*, in Algorithm 8.2.

When $m = 2$, Algorithm 8.2 reduces to the steepest descent/ascent method given in Algorithm 8.1.

8.4. Convergence analysis. While our convergent results are stated for all four possible (typ, ℓ) $\in \{(\pm, 1), (\pm, n)\}$, our proofs will be presented mostly for one (typ, ℓ),

$$(\mathrm{typ}, \ell) = (+, 1), \quad \text{and thus } \arg\operatorname{opt} = \arg\min \text{ in (8.6)}, \tag{8.13}$$

to save space. Proofs for other (typ, ℓ) can be obtained with minor changes accordingly. For convenience, in our proofs we will drop the pos-type sign $+$ in $r_+(\cdot)$, $\rho_+(\cdot)$, and u_j^+ with an understanding that they are all for the pos-type, even though, occasionally, the sign is still written out at critical places.

By Theorem 4.2, $\boldsymbol{Q}(\lambda)$ has n linearly independent pos-type eigenvectors u_j^+ for $1 \leqslant j \leqslant n$ and n linearly independent neg-type eigenvectors u_j^- for $1 \leqslant j \leqslant n$. Define for each (pos-type/neg-type) eigenvalue μ its corresponding eigenspace

$$\mathcal{U}_\mu = \{x \in \mathbb{C}^n \mid \boldsymbol{Q}(\mu)x = 0\} = \bigoplus_{\lambda_i^{\mathrm{typ}}=\mu} \operatorname{span}\{u_i^{\mathrm{typ}}\}.$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

Algorithm 8.2 Extended steepest descent/ascent method

Given an initial approximation $\boldsymbol{x}_0$ to u_ℓ^{typ}, a relative tolerance `rtol`, and the search space dimension m, the algorithm computes an approximate pair to $(\lambda_\ell^{\text{typ}}, u_\ell^{\text{typ}})$ with the prescribed `rtol`.

1: $\boldsymbol{x}_0 = \boldsymbol{x}_0/\|\boldsymbol{x}_0\|$, $\boldsymbol{\rho}_0 = \rho_{\text{typ}}(\boldsymbol{x}_0)$, $\boldsymbol{r}_0 = r_{\text{typ}}(\boldsymbol{x}_0)$;
2: **for** $i = 0, 1, \ldots$ **do**
3: **if** $\|\boldsymbol{r}_i\|/(|\boldsymbol{\rho}_i|^2\|A\boldsymbol{x}_i\| + |\boldsymbol{\rho}_i|\,\|B\boldsymbol{x}_i\| + \|C\boldsymbol{x}_i\|) \leqslant$ `rtol` **then**
4: **BREAK**;
5: **else**
6: compute a basis matrix Y_i for the Krylov subspace $\mathcal{K}_m(\boldsymbol{Q}(\boldsymbol{\rho}_i), \boldsymbol{x}_i)$ in (8.10);
7: solve the HQEP for $Y_i^{\text{H}}\boldsymbol{Q}(\lambda)Y_i$ to get its eigenvalues $\mu_j^{\pm}$ as in (8.11) and corresponding eigenvectors $y_j^{\pm}$;
8: select the next approximate eigenpair $(\mu, y) = (\mu_j^{\text{typ}}, Yy_j^{\text{typ}})$ according to (8.12);
9: $\boldsymbol{x}_{i+1} = y/\|y\|$, $\boldsymbol{\rho}_{i+1} = \mu$, $\boldsymbol{r}_{i+1} = r_{\text{typ}}(\boldsymbol{x}_{i+1})$;
10: **end if**
11: **end for**
12: **return** $(\boldsymbol{\rho}_i, \boldsymbol{x}_i)$ as an approximate eigenpair to $(\lambda_\ell^{\text{typ}}, u_\ell^{\text{typ}})$.

We will use the angle $\theta(\boldsymbol{x}_i, \mathcal{U}_\mu)$ from $\boldsymbol{x}_i$ to an eigenspace $\mathcal{U}_\mu$,

$$\cos\theta(\boldsymbol{x}_i, \mathcal{U}_\mu) := \min_{0\neq u\in\mathcal{U}_\mu} \frac{|u^{\text{H}}\boldsymbol{x}_i|}{\|\boldsymbol{x}_i\|_2\|u\|_2},$$

to measure the convergence of $\boldsymbol{x}_i$ toward $\mathcal{U}_\mu$. Note that $0 \leqslant \theta(\boldsymbol{x}_i, \mathcal{U}_\mu) \leqslant \pi/2$.

For the sake of our convergence analysis, it is convenient for us to execute Algorithms 8.1 and 8.2 without their Lines 3 and 4 so that $\boldsymbol{x}_i$, $\boldsymbol{r}_i$, and $\boldsymbol{\rho}_i$ are defined for all $i \geqslant 0$. But without the two lines, we need to be clear about the case when $\boldsymbol{r}_i = 0$ for some i. When it occurs, $\mathcal{K}_m(\boldsymbol{Q}(\boldsymbol{\rho}_i), \boldsymbol{x}_i) = \text{span}\{\boldsymbol{x}_i\}$ for any $m \geqslant 2$. For Algorithm 8.2, all subsequent $\boldsymbol{x}_j$, $\boldsymbol{\rho}_j$, and $\boldsymbol{r}_j$ for $j > i$ are well defined. In fact, we will have

$$\boldsymbol{\rho}_i = \boldsymbol{\rho}_{i+1} = \cdots, \quad \boldsymbol{x}_i = \boldsymbol{x}_{i+1} = \cdots, \quad \boldsymbol{r}_i = \boldsymbol{r}_{i+1} = \cdots = 0. \tag{8.14}$$

But, for Algorithm 8.1, all we have to do is to modify its Line 6 to $Y_i = \boldsymbol{x}_i$ if $\boldsymbol{r}_i = 0$, and then $\boldsymbol{x}_j$, $\boldsymbol{\rho}_j$, and $\boldsymbol{r}_j$ for $j > i$ are again well defined and they again satisfy (8.14).

THEOREM 8.1. *Let the sequences* $\{\boldsymbol{\rho}_i\}, \{\boldsymbol{r}_i\}, \{\boldsymbol{x}_i\}$ *be produced by* Algorithm 8.1/8.2.

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

(1) *Only one of the following two mutually exclusive situations can occur:*

  (a) *for some* i, (8.14) *holds, and* $(\boldsymbol{\rho}_i, \boldsymbol{x}_i)$ *is an eigenpair of* $\boldsymbol{Q}(\lambda)$*;*

  (b) $\boldsymbol{\rho}_i$ *is strictly monotonically decreasing for* $(\text{typ}, \ell) \in \{(\pm, 1)\}$ *or strictly monotonically increasing for* $(\text{typ}, \ell) \in \{(\pm, n)\}$, $\boldsymbol{r}_i \neq 0$ *for all* i*, and no two* $\boldsymbol{x}_i$ *are linearly dependent.*

(2) $\boldsymbol{x}_i^{\mathrm{H}} \boldsymbol{r}_i = 0$, $\boldsymbol{r}_i^{\mathrm{H}} \boldsymbol{r}_{i+1} = 0$, $\boldsymbol{x}_i^{\mathrm{H}} \boldsymbol{r}_{i+1} = 0$ *for* Algorithm 8.1.

(3) $\boldsymbol{x}_i^{\mathrm{H}} \boldsymbol{r}_i = 0$, $Y_i^{\mathrm{H}} \boldsymbol{r}_{i+1} = 0$ *for* Algorithm 8.2.

(4) *In the case of* 1(b),

  (a) $\boldsymbol{\rho}_i \to \hat{\rho} \in [\lambda_1^{\text{typ}}, \lambda_n^{\text{typ}}]$ *as* $i \to \infty$,

  (b) $\boldsymbol{r}_i \neq 0$ *for all* i *but* $\boldsymbol{r}_i \to 0$ *as* $i \to \infty$,

  (c) $\hat{\rho}$ *is an eigenvalue of* $\boldsymbol{Q}(\lambda)$*, and any limit point* $\hat{x}$ *of* $\{\boldsymbol{x}_i\}$ *is a corresponding eigenvector, that is,* $\boldsymbol{Q}(\hat{\rho})\hat{x} = 0$,

  (d) $\theta(\boldsymbol{x}_i, \mathcal{U}_{\hat{\rho}}) \to 0$ *as* $i \to \infty$.

Proof. As we remarked at the beginning of this subsection, we will prove the claims only for $(\text{typ}, \ell) = (+, 1)$.

There are only two possibilities: either $\boldsymbol{r}_i = 0$ for some i or $\boldsymbol{r}_i \neq 0$ for all i. If $\boldsymbol{r}_i = 0$ for some i, then $\boldsymbol{\rho}_i = \boldsymbol{\rho}_{i+1}$ and $\boldsymbol{x}_i = \boldsymbol{x}_{i+1}$ because $\rho(\boldsymbol{x}_i + t\boldsymbol{r}_i) \equiv \rho(\boldsymbol{x}_i)$. Consequently $\boldsymbol{r}_{i+1} = 0$, and the equations in (8.14) hold. Consider now $\boldsymbol{r}_i \neq 0$ for all i. Note that $\boldsymbol{r}_i \neq 0$ implies that $\nabla \boldsymbol{\rho}_i \neq 0$, and so $\rho(\boldsymbol{x}_i - s\nabla \boldsymbol{\rho}_i) < \rho(\boldsymbol{x}_i)$ for some s with sufficiently tiny $|s|$. This in turn implies that $\rho(\boldsymbol{x}_i + t\boldsymbol{r}_i) < \rho(\boldsymbol{x}_i)$ for some t with sufficiently tiny $|t|$, and thus

$$\boldsymbol{\rho}_{i+1} = \inf_t \rho(\boldsymbol{x}_i + t\boldsymbol{r}_i) < \rho(\boldsymbol{x}_i).$$

Therefore $\boldsymbol{\rho}_i$ is strictly monotonically decreasing. No two $\boldsymbol{x}_i$ are linearly dependent because linearly dependent $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ produce $\boldsymbol{\rho}_i = \boldsymbol{\rho}_j$. This proves item (1).

For item (2), $\boldsymbol{x}_i^{\mathrm{H}} \boldsymbol{r}_i = \boldsymbol{x}_i^{\mathrm{H}} \boldsymbol{Q}(\boldsymbol{\rho}_i) \boldsymbol{x}_i = 0$. Since $\rho(\boldsymbol{x}_{i+1}) = \min_t \rho(\boldsymbol{x}_i + t\boldsymbol{r}_i)$, by Lemma 8.1, $\boldsymbol{r}_i^{\mathrm{H}} \boldsymbol{r}_{i+1} = 0$. We now prove that $\boldsymbol{x}_i^{\mathrm{H}} \boldsymbol{r}_{i+1} = 0$. If $\boldsymbol{r}_i = 0$, then all $\boldsymbol{r}_j = 0$ for $j > i$, and thus no proof is necessary. Consider $\boldsymbol{r}_i \neq 0$. Then $\boldsymbol{\rho}_{i+1} < \boldsymbol{\rho}_i$. Note that $\boldsymbol{x}_{i+1}$ is a linear combination of $\boldsymbol{x}_i$ and $\boldsymbol{r}_i$; so we write $\boldsymbol{x}_{i+1} = \alpha_i \boldsymbol{x}_i + \beta_i \boldsymbol{r}_i$ for some scalars α_i and β_i. We know that $\beta_i \neq 0$; otherwise $\boldsymbol{x}_{i+1} = \alpha_i \boldsymbol{x}_i$ to yield $\boldsymbol{\rho}_{i+1} = \boldsymbol{\rho}_i$, which contradicts $\boldsymbol{\rho}_{i+1} < \boldsymbol{\rho}_i$. Therefore

$$\boldsymbol{\rho}_{i+1} = \rho(\boldsymbol{r}_i + (\alpha_i/\beta_i)\boldsymbol{x}_i) = \inf_t \rho(\boldsymbol{r}_i + t\boldsymbol{x}_i).$$

Apply Lemma 8.1 with $x = \boldsymbol{r}_i$ and $p = \boldsymbol{x}_i$ to get $\boldsymbol{x}_i^{\mathrm{H}} \boldsymbol{r}_{i+1} = 0$.

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

For item (3), again $\boldsymbol{x}_i^{\mathrm{H}}\boldsymbol{r}_i = \boldsymbol{x}_i^{\mathrm{H}}\boldsymbol{Q}(\boldsymbol{\rho}_i)\boldsymbol{x}_i = 0$. Let $\boldsymbol{x}_{i+1} = Y_i y$. Then, for each column z of Y_i, we have

$$\boldsymbol{\rho}_{i+1} = \rho(Y_i y) = \inf_t \rho(Y_i y + tz).$$

Apply Lemma 8.1 with $x = Y_i y$ and $p = z$ to get $z^{\mathrm{H}}\boldsymbol{r}_{i+1} = 0$. Since z is any column of Y_i, we conclude that $Y_i^{\mathrm{H}}\boldsymbol{r}_{i+1} = 0$.

Now for item 4(a), since $\boldsymbol{\rho}_i$ is strictly monotonically decreasing and bounded from below since $\boldsymbol{\rho}_i \geqslant \lambda_1^+$, it is convergent, and $\boldsymbol{\rho}_i \to \hat{\rho} \in [\lambda_1^+, \lambda_n^+]$, because $\boldsymbol{\rho}_i = \rho(\boldsymbol{x}_i) \in [\lambda_1^+, \lambda_n^+]$ for all i by Theorem 5.1.

For item 4(b), we have $\|\boldsymbol{r}_i\| = \|(A\boldsymbol{\rho}_i^2 + B\boldsymbol{\rho}_i + C)\boldsymbol{x}_i\| \leqslant \|A\|(\lambda_n^+)^2 + \|B\|\,|\lambda_n^+| + \|C\|$ since $\|\boldsymbol{x}_i\| = 1$; so both $\{\boldsymbol{r}_i\}$ and $\{\boldsymbol{x}_i\}$ are bounded sequences. It suffices to show that any limit point of $\{\boldsymbol{r}_i\}$ is the zero vector. Assume, to the contrary, that $\{\boldsymbol{r}_i\}$ has a nonzero limit point $\hat{r}$; that is, $\boldsymbol{r}_{i_j} \to \hat{r}$, where $\{\boldsymbol{r}_{i_j}\}$ is a subsequence of $\{\boldsymbol{r}_i\}$. Since $\{\boldsymbol{x}_{i_j}\}$ is bounded, it has a convergent subsequence. Without loss of generality, we may assume that $\boldsymbol{x}_{i_j}$ itself is convergent and that $\boldsymbol{x}_{i_j} \to \hat{x}$ as $j \to \infty$. We have $\hat{r}^{\mathrm{H}}\hat{x} = 0$ and $\|\hat{x}\| = 1$ because $\boldsymbol{r}_{i_j}^{\mathrm{H}}\boldsymbol{x}_{i_j} = 0$ and $\|\boldsymbol{x}_{i_j}\| = 1$. Now consider the quadratic eigenvalue problem for

$$\boldsymbol{Q}_{i_j}(\lambda) := Y_{i_j}^{\mathrm{H}}\boldsymbol{Q}(\lambda)Y_{i_j} = \begin{bmatrix} \boldsymbol{x}_{i_j}^{\mathrm{H}}\boldsymbol{Q}(\lambda)\boldsymbol{x}_{i_j} & \boldsymbol{x}_{i_j}^{\mathrm{H}}\boldsymbol{Q}(\lambda)\boldsymbol{r}_{i_j} \\ \boldsymbol{r}_{i_j}^{\mathrm{H}}\boldsymbol{Q}(\lambda)\boldsymbol{x}_{i_j} & \boldsymbol{r}_{i_j}^{\mathrm{H}}\boldsymbol{Q}(\lambda)\boldsymbol{r}_{i_j} \end{bmatrix}, \tag{8.15}$$

where $Y_{i_j} = [\boldsymbol{x}_{i_j}, \boldsymbol{r}_{i_j}]$. Since $\boldsymbol{r}_{i_j}^{\mathrm{H}}\boldsymbol{x}_{i_j} = 0$, $\operatorname{rank}(Y_{i_j}) = 2$, and thus $\boldsymbol{Q}_{i_j}(\lambda)$ is hyperbolic. Denote by $\mu_{j;k}^{\pm}$ its eigenvalues. It can be seen that

$$\lambda_1^- \leqslant \mu_{j;1}^- \leqslant \mu_{j;2}^- \leqslant \lambda_n^- < \lambda_1^+ \leqslant \mu_{j;1}^+ \leqslant \mu_{j;2}^+ \leqslant \lambda_n^+. \tag{8.16}$$

Then (for Algorithm 8.1, $\boldsymbol{\rho}_{i_j+1} = \mu_{j;1}^+$) $\lambda_1^+ \leqslant \boldsymbol{\rho}_{i_j+1} \leqslant \mu_{j;1}^+$. Let

$$\widehat{\boldsymbol{Q}}(\lambda) = \lim_{j\to\infty} \boldsymbol{Q}_{i_j}(\lambda),$$

whose eigenvalues are denoted by $\hat{\mu}_i^{\pm}$. By the continuity of the eigenvalues with respect to the entries of coefficient matrices of a quadratic polynomial with a nonsingular leading coefficient matrix, we know that $\mu_{j;i}^{\pm} \to \hat{\mu}_i^{\pm}$ as $j \to \infty$, and thus

$$\lambda_1^- \leqslant \hat{\mu}_1^- \leqslant \hat{\mu}_2^- \leqslant \lambda_n^- < \lambda_1^+ \leqslant \hat{\mu}_1^+ \leqslant \hat{\mu}_2^+ \leqslant \hat{\lambda}_n^+. \tag{8.17}$$

Notice that, by (8.16) and (8.17),

$$\lambda_1^+ \leqslant \boldsymbol{\rho}_{i_j+1} \leqslant \mu_{j;1}^+ \quad \Rightarrow \quad \hat{\mu}_2^- < \lambda_1^+ \leqslant \hat{\rho} \leqslant \hat{\mu}_1^+. \tag{8.18}$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

On the other hand, by (8.16), we have

$$\widehat{\boldsymbol{Q}}(\hat{\rho}) = \lim_{j\to\infty} \boldsymbol{Q}_{i_j}(\boldsymbol{\rho}_{i_j}) = \lim_{j\to\infty} \begin{bmatrix} 0 & \boldsymbol{r}_{i_j}^{\mathrm{H}}\boldsymbol{r}_{i_j} \\ \boldsymbol{r}_{i_j}^{\mathrm{H}}\boldsymbol{r}_{i_j} & \boldsymbol{r}_{i_j}^{\mathrm{H}}\boldsymbol{Q}(\boldsymbol{\rho}_{i_j})\boldsymbol{r}_{i_j} \end{bmatrix} = \begin{bmatrix} 0 & \hat{r}^{\mathrm{H}}\hat{r} \\ \hat{r}^{\mathrm{H}}\hat{r} & \hat{r}^{\mathrm{H}}\boldsymbol{Q}(\hat{\rho})\hat{r} \end{bmatrix},$$

which is indefinite because $\hat{r}^{\mathrm{H}}\hat{r} > 0$. But, by (8.18) and Theorem 3.1, $\widehat{\boldsymbol{Q}}(\hat{\rho}) \preceq 0$, a contradiction. So $\hat{r} = 0$, as was to be shown.

For item 4(c), since $\|\boldsymbol{x}_i\| = 1$, $\{\boldsymbol{x}_i\}$ has at least one limit point. Let $\hat{x}$ be any limit point of $\boldsymbol{x}_i$; that is, $\boldsymbol{x}_{i_j} \to \hat{x}$. Take limits on both sides of $\boldsymbol{Q}(\boldsymbol{\rho}_{i_j})\boldsymbol{x}_{i_j} = \boldsymbol{r}_{i_j}$ to get $\boldsymbol{Q}(\hat{\rho})\hat{x} = 0$; that is, $(\hat{\rho}, \hat{x})$ is an eigenpair.

For item 4(d), write $\theta_i = \theta(\boldsymbol{x}_i, \mathcal{U}_{\hat{\rho}})$ for convenience and write (without loss of generality, we may assume that $\|\cdot\|_2$ is used in the algorithms) $\boldsymbol{x}_i = \hat{u}_i \cos\theta_i + \hat{v}_i \sin\theta_i$, where $\hat{u}_i \in \mathcal{U}_{\hat{\rho}}$, $\hat{v}_i \in \mathcal{U}_{\hat{\rho}}^{\perp}$ (the orthogonal complement of $\mathcal{U}_{\hat{\rho}}$), and $\|\hat{u}_i\|_2 = \|\hat{v}_i\|_2 = 1$. Then

$$\boldsymbol{r}_i = \boldsymbol{Q}(\boldsymbol{\rho}_i)\boldsymbol{x}_i = (\boldsymbol{\rho}_i - \hat{\rho})[(\boldsymbol{\rho}_i + \hat{\rho})A + B]\hat{u}_i \cos\theta_i + \boldsymbol{Q}(\boldsymbol{\rho}_i)\hat{v}_i \sin\theta_i. \tag{8.19}$$

We claim that $\boldsymbol{Q}(\boldsymbol{\rho}_i)\hat{v}_i \sin\theta_i \to 0$. To see this, we notice that

$$\|(\boldsymbol{\rho}_i + \hat{\rho})A + B\|_2 \leqslant 2\max\{|\lambda_1^+|, |\lambda_n^+|\}\,\|A\|_2 + \|B\|_2,$$

$\boldsymbol{r}_i \to 0$, and $\boldsymbol{\rho}_i - \hat{\rho} \to 0$. Thus $\boldsymbol{Q}(\boldsymbol{\rho}_i)\hat{v}_i \sin\theta_i \to 0$ by (8.19). The null space of $\boldsymbol{Q}(\hat{\rho})$ is $\mathcal{U}_{\hat{\rho}}$. Since $\boldsymbol{Q}(\hat{\rho})$ is Hermitian,

$$\|\boldsymbol{Q}(\hat{\rho})v\|_2 \geqslant \gamma\|v\|_2 \quad \text{for any } v \in \mathcal{U}_{\hat{\rho}}^{\perp},$$

where $\gamma = \min|\xi|$ taken over all nonzero $\xi \in \mathrm{eig}(\boldsymbol{Q}(\hat{\rho}))$. Therefore $\|\boldsymbol{Q}(\hat{\rho})\hat{v}_i\|_2 \geqslant \gamma$. Because $\boldsymbol{\rho}_i \to \hat{\rho}$, for sufficiently large i we have $\|\boldsymbol{Q}(\boldsymbol{\rho}_i)\hat{v}_i\|_2 \geqslant \gamma/2$, and thus

$$\|\boldsymbol{Q}(\boldsymbol{\rho}_i)\hat{v}_i \sin\theta_i\|_2 \geqslant (\gamma/2)\sin\theta_i,$$

implying that $\sin\theta_i \to 0$, which leads to $\theta_i \to 0$, because $0 \leqslant \theta_i \leqslant \pi/2$. □

Theorem 8.1 ensures us the global convergence of Algorithm 8.1/8.2, but gives no indication as how fast the convergence may be. For that, we turn to our next theorem—Theorem 8.2—which provides an asymptotic rate of the sequences $\{\boldsymbol{\rho}_i\}$ generated by the algorithms. These theorems are reminiscent of [**19**, Theorem 3.2] and [**19**, Theorem 3.4], respectively. But Theorem 8.2 about the rate of convergence is much more difficult to prove than [**19**, Theorem 3.4]. Because of that, we defer its proof to Appendix C.

We introduce some new notation: for any $x \neq 0$,

$$a(x) = \frac{x^{\mathrm{H}}Ax}{x^{\mathrm{H}}x}, \quad b(x) = \frac{x^{\mathrm{H}}Bx}{x^{\mathrm{H}}x}, \quad c(x) = \frac{x^{\mathrm{H}}Cx}{x^{\mathrm{H}}x}. \tag{8.20}$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

Also recall that $\boldsymbol{Q}_{\lambda_0}(\lambda) := \boldsymbol{Q}(\lambda+\lambda_0)$ in (6.2a) for a given shift λ_0. Accordingly,

$$b_0(x) = \frac{x^{\mathrm{H}} B_{\lambda_0} x}{x^{\mathrm{H}} x} = \frac{x^{\mathrm{H}}(2\lambda_0 A + B)x}{x^{\mathrm{H}} x}, \quad c_0(x) = \frac{x^{\mathrm{H}} C_{\lambda_0} x}{x^{\mathrm{H}} x} = \frac{x^{\mathrm{H}} \boldsymbol{Q}(\lambda_0) x}{x^{\mathrm{H}} x}. \tag{8.21}$$

THEOREM 8.2. *Suppose that* $\lambda_1^{\mathrm{typ}} \leqslant \boldsymbol{\rho}_0 < \lambda_2^{\mathrm{typ}}$ *if* $\ell = 1$ *or* $\lambda_{n-1}^{\mathrm{typ}} < \boldsymbol{\rho}_0 \leqslant \lambda_n^{\mathrm{typ}}$ *if* $\ell = n$, *and let the sequences* $\{\boldsymbol{\rho}_i\}$, $\{\boldsymbol{r}_i\}$, $\{\boldsymbol{x}_i\}$ *be produced by Algorithm 8.2. Given a shift* $\lambda_0 \geqslant \lambda_n^+$, *define* B_{λ_0}, C_{λ_0} *by* (6.2a).

(1) *As* $i \to \infty$, $\boldsymbol{\rho}_i$ *monotonically converges to* $\hat{\rho} = \lambda_\ell^{\mathrm{typ}}$, *and* $\boldsymbol{x}_i$ *converges to* u_ℓ^{typ} *in direction; that is,* $\theta(\boldsymbol{x}_i, u_\ell^{\mathrm{typ}}) \to 0$.

(2) *The eigenvalues (their dependency upon* i *is suppressed for clarity)* ω_j *of the matrix* $\boldsymbol{Q}(\boldsymbol{\rho}_i)$ *can be ordered as*

$$\omega_1 > 0 > \omega_2 \geqslant \cdots \geqslant \omega_n \quad \text{if } (\mathrm{typ}, \ell) \in \{(+,1),(-,n)\}, \quad \text{or} \tag{8.22a}$$

$$\omega_1 < 0 < \omega_2 \leqslant \cdots \leqslant \omega_n \quad \text{if } (\mathrm{typ}, \ell) \in \{(+,n),(-,1)\}. \tag{8.22b}$$

Denote by v_1 *the eigenvector of* $\boldsymbol{Q}(\boldsymbol{\rho}_i)$ *associated with its eigenvalue* ω_1. *If* $\boldsymbol{\rho}_i$ *is sufficiently close to* $\lambda_\ell^{\mathrm{typ}}$, *then*

$$\begin{aligned} |\boldsymbol{\rho}_{i+1} - \lambda_\ell^{\mathrm{typ}}| \leqslant \varepsilon_{m-1}^2 |\boldsymbol{\rho}_i - \lambda_\ell^{\mathrm{typ}}| + (1-\varepsilon_{m-1}^2)\varepsilon_{m-1}\eta(v_1)|\boldsymbol{\rho}_i - \lambda_\ell^{\mathrm{typ}}|^{3/2} \\ + O(|\boldsymbol{\rho}_i - \lambda_\ell^{\mathrm{typ}}|^2), \end{aligned} \tag{8.23}$$

where

$$\varepsilon_{m-1} = \min_{g \in \mathbb{P}_{m-1}, g(\omega_1)\neq 0} \max_{i \neq 1} \frac{|g(\omega_i)|}{|g(\omega_1)|}, \tag{8.24}$$

$$\tau_A = \frac{1}{|\omega_2|}\frac{\|A\|_2}{a(v_1)}, \quad \tau_B = \frac{1}{|\omega_2|}\frac{\|B_{\lambda_0}\|_2}{b_0(v_1)}, \quad \tau_C = \frac{1}{|\omega_2|}\frac{\|C_{\lambda_0}\|_2}{c_0(v_1)}, \tag{8.25}$$

$$\eta(v_1) = 3\tau_A^{1/2} + 2\frac{(b_0(v_1))^2 \tau_B^{1/2} + 2a(v_1)c_0(v_1)(\tau_A^{1/2} + \tau_C^{1/2})}{\varsigma_0(v_1)^2}, \tag{8.26}$$

and $\mathbb{P}_{m-1}$ *is the set of polynomials of degree no higher than* $m-1$.

(3) *Denote (*$\boldsymbol{Q}(\lambda_\ell^{\mathrm{typ}})$ *is singular and, by Theorem 3.1, negative semidefinite if* $(\mathrm{typ}, \ell) \in \{(+,1),(-,n)\}$ *or positive semidefinite if* $(\mathrm{typ}, \ell) \in \{(+,n),(-,1)\}$*) by* γ *and* Γ *the smallest and largest positive eigenvalue of the matrix*

$$\begin{cases} -\boldsymbol{Q}(\lambda_\ell^{\mathrm{typ}}) & \text{for } (\mathrm{typ}, \ell) \in \{(+,1),(-,n)\}, \\ \boldsymbol{Q}(\lambda_\ell^{\mathrm{typ}}) & \text{for } (\mathrm{typ}, \ell) \in \{(+,n),(-,1)\}. \end{cases}$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

If $\boldsymbol{\rho}_i$ is sufficiently close to $\lambda_\ell^{\mathrm{typ}}$, then

$$|\boldsymbol{\rho}_{i+1}-\lambda_\ell^{\mathrm{typ}}| \leqslant \varepsilon^2|\boldsymbol{\rho}_i-\lambda_\ell^{\mathrm{typ}}|+(1-\varepsilon^2)\varepsilon\eta|\boldsymbol{\rho}_i-\lambda_\ell^{\mathrm{typ}}|^{3/2}+O(|\boldsymbol{\rho}_i-\lambda_\ell^{\mathrm{typ}}|^2), \quad (8.27)$$

where

$$\varepsilon = 2\left[\left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^{m-1}+\left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^{-(m-1)}\right]^{-1}, \quad \kappa = \frac{\Gamma}{\gamma}, \quad (8.28)$$

$$\eta = \sqrt{\frac{1}{|\gamma|}}\left[3\sqrt{\frac{\|A\|_2}{a(u)}}+2\frac{b_0(u)^2}{\varsigma_0(u)^2}\sqrt{\frac{\|B_{\lambda_0}\|_2}{b_0(u)}} + 4\frac{a(u)c_0(u)}{\varsigma_0(u)^2}\left(\sqrt{\frac{\|A\|_2}{a(u)}}+\sqrt{\frac{\|C_{\lambda_0}\|_2}{c_0(u)}}\right)\right] \quad (8.29)$$

$$\leqslant \sqrt{\frac{1}{|\gamma|}}\left[3\sqrt{\frac{\|A\|_2}{a(u)}}+2\frac{\|B_{\lambda_0}\|_2^2+4\|A\|_2\|C_{\lambda_0}\|_2}{b(u)^2-4a(u)c(u)}\right], \quad (8.30)$$

and $u = u_\ell^{\mathrm{typ}}$ for short.

9. Preconditioned steepest descent/ascent method

9.1. Preconditioning. We will explain the idea of preconditioning for computing (λ_1^+, u_1^+) only, via two different points of view. The same argument can be made for other extreme positive and negative eigenpairs.

It is well known that, when the contours of the objective function near its optimum are extremely elongated, at each step of the conventional steepest descent/ascent method, following the search direction which is the opposite of the gradient gets closer to the optimum on the line for a very short while and then starts to get away because the direction does not point 'toward the optimum', resulting in a long zigzag path of a large number of steps. The ideal search direction p is therefore the one such that, with its starting point at $\boldsymbol{x}$, p points to the optimum; that is, the optimum is on the line $\{\boldsymbol{x}+tp \,:\, t\in\mathbb{C}\}$. Specifically, expand $\boldsymbol{x}$ as a linear combination of u_j^+:

$$\boldsymbol{x} = \sum_{j=1}^{n}\alpha_j u_j^+ =: \alpha_1 u_1^+ + \boldsymbol{v}, \quad \boldsymbol{v} = \sum_{j=2}^{n}\alpha_j u_j^+. \quad (9.1)$$

Then the ideal search direction is

$$p = \alpha u_1^+ + \beta\boldsymbol{v}$$

for some scalars α and $\beta \neq 0$ such that $\alpha_1\beta - \alpha \neq 0$ (otherwise $p = \beta\boldsymbol{x}$). Of course, this is impractical because we do not know u_1^+ and $\boldsymbol{v}$. But we can construct

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

one that is close to it. One such p is

$$p = [\boldsymbol{Q}(\sigma)]^{-1} r_+(\boldsymbol{x}) = [\boldsymbol{Q}(\sigma)]^{-1} \boldsymbol{Q}(\boldsymbol{\rho}_+)\boldsymbol{x},$$

where $\boldsymbol{\rho}_+ = \rho_+(\boldsymbol{x})$ and (we reasonably assume also that $\sigma \neq \lambda_j^+$ for all j, too) σ is some shift near λ_1^+ but not equal to $\boldsymbol{\rho}_+$. Let us analyze this p. By (4.14a), we have

$$\begin{aligned}[\boldsymbol{Q}(\sigma)]^{-1} \boldsymbol{Q}(\boldsymbol{\rho}_+) = U_+(\sigma I - \Lambda_+)^{-1}(U_-^{\mathrm{H}} A U_+)^{-1}(\sigma I - \Lambda_-)^{-1} \\ \times (\boldsymbol{\rho}_+ I - \Lambda_-) U_-^{\mathrm{H}} A U_+ (\boldsymbol{\rho}_+ I - \Lambda_+) U_+^{-1}.\end{aligned}$$

Suppose now that both σ and $\boldsymbol{\rho}_+$ are near λ_1^+. Then

$$(\sigma I - \Lambda_-)^{-1}(\boldsymbol{\rho}_+ I - \Lambda_-) = I + (\boldsymbol{\rho}_+ - \sigma)(\sigma I - \Lambda_-)^{-1} \approx I.$$

Therefore $[\boldsymbol{Q}(\sigma)]^{-1} \boldsymbol{Q}(\boldsymbol{\rho}_+) \approx U_+(\sigma I - \Lambda_+)^{-1}(\boldsymbol{\rho}_+ I - \Lambda_+) U_+^{-1}$, and thus

$$p = [\boldsymbol{Q}(\sigma)]^{-1} \boldsymbol{Q}(\boldsymbol{\rho}_+)\boldsymbol{x} \approx \sum_{j=1}^{n} \mu_j \alpha_j u_j^+, \quad \mu_j := \frac{\lambda_j^+ - \boldsymbol{\rho}_+}{\lambda_j^+ - \sigma}. \tag{9.2}$$

Now if $\lambda_1^+ \leqslant \boldsymbol{\rho}_+ < \lambda_2^+$ and if the gap $\lambda_2^+ - \lambda_1^+$ is reasonably modest, then

$$\mu_j \approx 1 \quad \text{for } j > 1$$

to give a $p \approx \alpha u_1^+ + \boldsymbol{v}$, resulting in fast convergence. This rough but intuitive analysis suggests that $K = [\boldsymbol{Q}(\sigma)]^{-1}$ with a suitably chosen shift σ can be used to serve as a good preconditioner to improve the steepest descent/ascent method—Algorithm 8.1—by simply modifying $Y_i = [\boldsymbol{x}_i, \boldsymbol{r}_i]$ at Line 6 there to $Y_i = [\boldsymbol{x}_i, K\boldsymbol{r}_i]$. We caution the reader that implementing $K\boldsymbol{r}_i$ amounts to solving a linear system. This is usually done approximately by, for example, some iterative methods such as the linear conjugate gradient method or MINRES [12, 18, 20].

The second viewpoint is similar to the one proposed by Golub and Ye [19] for the generalized linear eigenvalue problem. Theorem 8.2 reveals that the rates of convergence for Algorithms 8.1 and 8.2 depend on the distribution of the eigenvalues ω_j of $\boldsymbol{Q}(\boldsymbol{\rho}_i)$, not the eigenvalues of $\boldsymbol{Q}(\lambda)$. In particular, if all $\omega_2 = \cdots = \omega_n$, then $\epsilon_m = 0$ for $m \geqslant 2$, and thus

$$\boldsymbol{\rho}_{i+1} - \lambda_1^+ = O(|\boldsymbol{\rho}_i - \lambda_1^+|^2),$$

suggesting quadratic convergence. Such an extreme case, though highly welcome, is unlikely to happen in practice, but it gives us an idea that if somehow we could transform an eigenvalue problem toward such an extreme case, the

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

transformed problem would be easier to solve. Specifically we should seek equivalent transformations that change the eigenvalues of the matrix $\boldsymbol{Q}(\boldsymbol{\rho}_i)$ as much as possible to

one isolated eigenvalue ω_1, and the rest ω_j $(2 \leqslant j \leqslant n)$ tightly clustered,	(9.3)

but leave the eigenvalues of $\boldsymbol{Q}(\lambda)$ unchanged.

We would like to equivalently transform the HQEP for $\boldsymbol{Q}(\lambda)$ to one for $L^{-1}\boldsymbol{Q}(\lambda)L^{-\mathrm{H}}$ by some nonsingular L (whose inverse or any linear system with L is easy to solve) so that the eigenvalues of $L^{-1}\boldsymbol{Q}(\boldsymbol{\rho}_i)L^{-\mathrm{H}}$ distribute more or less like (9.3). Then apply one step of Algorithm 8.1 or 8.2 to the pencil $L^{-1}\boldsymbol{Q}(\lambda)L^{-\mathrm{H}}$ to find the next approximation $\boldsymbol{\rho}_{i+1}$. The process repeats: that is, find a new L to transform the problem and apply one step of Algorithm 8.1 or 8.2 to the transformed problem.

Such an L may be constructed using the LDL^{H} decomposition of $\boldsymbol{Q}(\boldsymbol{\rho}_i)$ [**18**, page 139] if the decomposition exists: $\boldsymbol{Q}(\boldsymbol{\rho}_i) = LDL^{\mathrm{H}}$, where L is lower triangular and $D = \mathrm{diag}(\pm 1)$. Then $L^{-1}\boldsymbol{Q}(\boldsymbol{\rho}_i)L^{-\mathrm{H}} = D$ has the ideal eigenvalue distribution that gives $\epsilon_m = 0$ for any $m \geqslant 2$. Unfortunately, this simple solution is impractical in practice for the following reasons.

(1) The decomposition may not exist at all. In theory, the decomposition exists if all the leading principle submatrices of $\boldsymbol{Q}(\boldsymbol{\rho}_i)$ are nonsingular.

(2) If the decomposition does exist, it may not be numerically stable to compute, especially when $\boldsymbol{\rho}_i$ comes closer and closer to λ_1^+.

(3) The sparsity in $\boldsymbol{Q}(\boldsymbol{\rho}_i)$ is most likely destroyed, leaving L significantly denser than $\boldsymbol{Q}(\boldsymbol{\rho}_i)$. This makes all ensuing computations much more expensive.

A more practical solution is, however, through an incomplete LDL^{H} factorization (see [**54**, Ch. 10]), to get

$$\boldsymbol{Q}(\boldsymbol{\rho}_i) \approx LDL^{\mathrm{H}},$$

where $\approx$ includes not only the usual 'approximately equal', but also the case when $\boldsymbol{Q}(\boldsymbol{\rho}_i) - LDL^{\mathrm{H}}$ is approximately a low rank matrix, and $D = \mathrm{diag}(\pm 1)$. Such an L changes from one step of the algorithm to another. In practice, often we may use one fixed preconditioner for all or a number of consecutive iterative steps. Using a constant preconditioner is certainly not optimal: it likely does not give the best rate of convergence per step and thus it increases the number of total iterative steps, but it may reduce overall cost because it saves work in preconditioner constructions and thus it reduces the cost per step. The basic idea of using a step-independent preconditioner is to find a σ that is close to λ_1^+, perform an incomplete LDL^{H}

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

decomposition

$$\boldsymbol{Q}(\sigma) \approx LDL^{\mathrm{H}},$$

and transform $\boldsymbol{Q}(\lambda)$ accordingly before applying Algorithm 8.1 or 8.2. Now the rate of convergence is determined by the eigenvalues of

$$L^{-1}\boldsymbol{Q}(\boldsymbol{\rho}_i)L^{-\mathrm{H}} = L^{-1}\boldsymbol{Q}(\sigma)L^{-\mathrm{H}} + (\boldsymbol{\rho}_i - \sigma)L^{-1}\boldsymbol{Q}'(\sigma)L^{-\mathrm{H}} + O(|\boldsymbol{\rho}_i - \sigma|^2),$$

which would have a better spectral distribution so long as the last two terms are small relative to $L^{-1}\boldsymbol{Q}(\boldsymbol{\rho}_i)L^{-\mathrm{H}}$. When $\lambda_n^- < \sigma < \lambda_1^+$, $-\boldsymbol{Q}(\sigma) \succ 0$, and the incomplete LDL^{H} factorization becomes incomplete Cholesky factorization.

9.2. Preconditioned steepest descent/ascent method. We have insisted so far on applying Algorithm 8.1 or 8.2 straightforwardly to the transformed problem. There is another way, perhaps a better one: only symbolically applying Algorithm 8.1 or 8.2 to the transformed problem as a derivation tool for a preconditioned method that always projects the original pencil $\boldsymbol{Q}(\lambda)$ directly every step. The only difference is that now the projecting subspaces are preconditioned. Again we will explain it for the case of computing the first pos-type eigenpair (λ_1^+, u_1^+).

Suppose that $\boldsymbol{Q}(\lambda)$ is transformed to $\widehat{\boldsymbol{Q}}(\lambda) := L^{-1}\boldsymbol{Q}(\lambda)L^{-\mathrm{H}}$. Consider a typical step of Algorithm 8.2 applied to $\widehat{\boldsymbol{Q}}(\lambda)$. For the purpose of distinguishing notational symbols, we will put hats on all those for $\widehat{\boldsymbol{Q}}(\lambda)$. The typical step of Algorithm 8.2 on $\widehat{\boldsymbol{Q}}$ is

computing the smallest pos-type eigenvalue μ and corresponding eigenvector $\hat{v}$ of $\hat{Z}^{\mathrm{H}}\widehat{\boldsymbol{Q}}(\lambda)\hat{Z}$, where $\hat{Z} \in \mathbb{C}^{n\times m}$ is a basis matrix of Krylov subspace $\mathcal{K}_m(\widehat{\boldsymbol{Q}}(\hat{\boldsymbol{\rho}}), \hat{\boldsymbol{x}})$.	(9.4)

Notice that $[\widehat{\boldsymbol{Q}}(\hat{\boldsymbol{\rho}})]^j\hat{\boldsymbol{x}} = L^{\mathrm{H}}[(LL^{\mathrm{H}})^{-1}\boldsymbol{Q}(\hat{\boldsymbol{\rho}})]^j(L^{-\mathrm{H}}\hat{\boldsymbol{x}})$ to see that

$$L^{-\mathrm{H}}\mathcal{K}_m(\widehat{\boldsymbol{Q}}(\hat{\boldsymbol{\rho}}), \hat{\boldsymbol{x}}) = \mathcal{K}_m(K\boldsymbol{Q}(\hat{\boldsymbol{\rho}}), \boldsymbol{x}),$$

where $\boldsymbol{x} = L^{-\mathrm{H}}\hat{\boldsymbol{x}}$ and $K = (LL^{\mathrm{H}})^{-1}$. So $Z = L^{-\mathrm{H}}\hat{Z}$ is a basis matrix of Krylov subspace $\mathcal{K}_m(K\boldsymbol{Q}(\hat{\boldsymbol{\rho}}), \boldsymbol{x})$. Since also

$$\hat{Z}^{\mathrm{H}}\widehat{\boldsymbol{Q}}(\lambda)\hat{Z} = (L^{-\mathrm{H}}\hat{Z})^{\mathrm{H}}\boldsymbol{Q}(\lambda)(L^{-\mathrm{H}}\hat{Z}),$$
$$\hat{\boldsymbol{\rho}} = \hat{\rho}_+(\hat{\boldsymbol{x}}) = \rho_+(\boldsymbol{x}) = \boldsymbol{\rho},$$

the typical step (9.4) can be reformulated equivalently to

computing the smallest pos-type eigenvalue μ and corresponding eigenvector v of $Z^{\mathrm{H}}\boldsymbol{Q}(\lambda)Z$, where $Z \in \mathbb{C}^{n\times m}$ is a basis matrix of Krylov subspace $\mathcal{K}_m(K\boldsymbol{Q}(\boldsymbol{\rho}), \boldsymbol{x})$, where $K = (LL^{\mathrm{H}})^{-1}$.	(9.5)

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

Algorithm 9.1 Preconditioned extended steepest descent/ascent method

Given an initial approximation $\boldsymbol{x}_0$ to u_ℓ^{typ}, a relative tolerance `rtol`, and the search space dimension m, the algorithm computes an approximate pair to $(\lambda_\ell^{\text{typ}}, u_\ell^{\text{typ}})$ with the prescribed `rtol`.

$\boldsymbol{x}_0 = \boldsymbol{x}_0/\|\boldsymbol{x}_0\|$, $\boldsymbol{\rho}_0 = \rho_{\text{typ}}(\boldsymbol{x}_0)$, $\boldsymbol{r}_0 = r_{\text{typ}}(\boldsymbol{x}_0)$;
for $i = 0, 1, \ldots$ **do**
  if $\|\boldsymbol{r}_i\|/(|\boldsymbol{\rho}_i|^2\|A\boldsymbol{x}_i\| + |\boldsymbol{\rho}_i|\,\|B\boldsymbol{x}_i\| + \|C\boldsymbol{x}_i\|) \leqslant$ `rtol` **then**
    BREAK;
  else
    construct a preconditioner K_i;
    compute a basis matrix Y_i for the Krylov subspace $\mathcal{K}_m(K_i\boldsymbol{Q}(\boldsymbol{\rho}_i), \boldsymbol{x}_i)$;
    solve HQEP for $Y_i^{\mathrm{H}}\boldsymbol{Q}(\lambda)Y_i$ to get its eigenvalues $\mu_j^{\pm}$ as in (8.11) and eigenvectors $y_j^{\pm}$;
    select the next approximate eigenpair $(\mu, y) = (\mu_j^{\text{typ}}, Yy_j^{\text{typ}})$ according to (8.12);
    $\boldsymbol{x}_{i+1} = y/\|y\|$, $\boldsymbol{\rho}_{i+1} = \mu$, $\boldsymbol{r}_{i+1} = r_{\text{typ}}(\boldsymbol{x}_{i+1})$;
  end if
end for
return $(\boldsymbol{\rho}_i, \boldsymbol{x}_i)$ as an approximate eigenpair to $(\lambda_\ell^{\text{typ}}, u_\ell^{\text{typ}})$.

We are now ready to state a version of the *preconditioned extended steepest descent/ascent method*. To make it inclusive, in Algorithm 9.1 we use K_i to denote the preconditioner at the ith iterative step. Once again, they may all be the same or they may vary from one iterative step to another. Although the derivation of this algorithm was for the preconditioners obtained from the second viewpoint above, its final form includes the preconditioners from the first viewpoint.

9.3. Convergence analysis. If $K_i \succ 0$, the ith iterative step of Algorithm 9.1 is just one step of the extended steepest descent/ascent method applied to $K_i^{1/2}\boldsymbol{Q}(\lambda)K_i^{1/2}$. Therefore Theorem 8.2 implies the following theorem for Algorithm 9.1.

THEOREM 9.1. *Suppose that* $\lambda_1^{\text{typ}} \leqslant \boldsymbol{\rho}_0 < \lambda_2^{\text{typ}}$ *if* $\ell = 1$ *or* $\lambda_{n-1}^{\text{typ}} < \boldsymbol{\rho}_0 \leqslant \lambda_n^{\text{typ}}$ *if* $\ell = n$, *and let the sequences* $\{\boldsymbol{\rho}_i\}, \{\boldsymbol{r}_i\}, \{\boldsymbol{x}_i\}$ *be produced by* Algorithm 9.1. *Suppose that* $K_i \succ 0$.

(1) *As* $i \to \infty$, $\boldsymbol{\rho}_i$ *monotonically converges to* $\hat{\rho} = \lambda_\ell^{\text{typ}}$, *and* $\boldsymbol{x}_i$ *converges to* u_ℓ^{typ} *in direction; that is,* $\theta(\boldsymbol{x}_i, u_\ell^{\text{typ}}) \to 0$.

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

(2) *The eigenvalues (their dependency upon i is suppressed for clarity) ω_j of $K_i\boldsymbol{Q}(\boldsymbol{\rho}_i)$ can be ordered as*

$$\omega_1 > 0 > \omega_2 \geqslant \cdots \geqslant \omega_n \quad \text{if } (\mathrm{typ}, \ell) \in \{(+,1),(-,n)\}, \quad \text{or} \tag{9.6a}$$

$$\omega_1 < 0 < \omega_2 \leqslant \cdots \leqslant \omega_n \quad \text{if } (\mathrm{typ}, \ell) \in \{(+,n),(-,1)\}. \tag{9.6b}$$

If $\boldsymbol{\rho}_i$ is sufficiently close to $\lambda_\ell^{\mathrm{typ}}$, then

$$|\boldsymbol{\rho}_{i+1} - \lambda_\ell^{\mathrm{typ}}| \leqslant \varepsilon_{m-1}^2 |\boldsymbol{\rho}_i - \lambda_\ell^{\mathrm{typ}}| + O(\varepsilon_{m-1}|\boldsymbol{\rho}_i - \lambda_\ell^{\mathrm{typ}}|^{3/2} + |\boldsymbol{\rho}_i - \lambda_\ell^{\mathrm{typ}}|^2), \tag{9.7}$$

where ε_{m-1} is defined as in (8.24).

(3) *Denote (it is worth emphasizing that $K_i\boldsymbol{Q}(\lambda_\ell^{\mathrm{typ}})$ is singular and, by Theorem 3.1, $K_i^{1/2}\boldsymbol{Q}(\lambda_\ell^{\mathrm{typ}})K_i^{1/2}$ is negative semidefinite if $(\mathrm{typ}, \ell) \in \{(+,1),(-,n)\}$ and positive semidefinite if $(\mathrm{typ}, \ell) \in \{(+,n),(-,1)\}$) by γ and Γ the smallest and largest positive eigenvalue of the matrix*

$$\begin{cases} -K_i\boldsymbol{Q}(\lambda_\ell^{\mathrm{typ}}) & \text{for } (\mathrm{typ}, \ell) \in \{(+,1),(-,n)\}, \\ K_i\boldsymbol{Q}(\lambda_\ell^{\mathrm{typ}}) & \text{for } (\mathrm{typ}, \ell) \in \{(+,n),(-,1)\}. \end{cases}$$

If $\boldsymbol{\rho}_i$ is sufficiently close to $\lambda_\ell^{\mathrm{typ}}$, then

$$|\boldsymbol{\rho}_{i+1} - \lambda_\ell^{\mathrm{typ}}| \leqslant \varepsilon^2 |\boldsymbol{\rho}_i - \lambda_\ell^{\mathrm{typ}}| + O(\varepsilon|\boldsymbol{\rho}_i - \lambda_\ell^{\mathrm{typ}}|^{3/2} + |\boldsymbol{\rho}_i - \lambda_\ell^{\mathrm{typ}}|^2), \tag{9.8}$$

where ε is defined as in (8.28).

There is a convergence rate estimate, essentially due to Samokish [**55**, 1958], for the preconditioned steepest descent/ascent method in the case of the standard Hermitian eigenvalue problem. The reader is referred to [**29**, **49**] for details. Theorem 9.2 below is an extension of Samokish's result for an HQEP.

THEOREM 9.2. *Suppose that $K \succ 0$. Let $\ell \in \{1, n\}$ and $\mathrm{typ}, \mathrm{typ}' \in \{+,-\}$ such that typ and typ$'$ are opposite, and denote by γ and Γ the smallest and largest positive eigenvalue of the matrix*

$$\begin{cases} -K\boldsymbol{Q}(\lambda_\ell^{\mathrm{typ}}) & \text{for } (\mathrm{typ}, \ell) \in \{(+,1),(-,n)\}, \\ K\boldsymbol{Q}(\lambda_\ell^{\mathrm{typ}}) & \text{for } (\mathrm{typ}, \ell) \in \{(+,n),(-,1)\}, \end{cases}$$

and

$$\tau = \frac{2}{\gamma + \Gamma}, \quad \kappa = \frac{\Gamma}{\gamma}, \quad \varepsilon = \frac{\kappa - 1}{\kappa + 1}.$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

Let arg opt *be as given in* (8.6)*, and let*

$$t_{\mathrm{opt}} = \arg\operatorname{opt}_{t\in\mathbb{C}}\, \rho_{\mathrm{typ}}(x + tKr_{\mathrm{typ}}(x)), \quad y = x + t_{\mathrm{opt}}Kr_{\mathrm{typ}}(x),$$
$$z = \begin{cases} x + \tau Kr_{+}(x) & \text{for } (\mathrm{typ},\ell) \in \{(+,1),(-,n)\}, \\ x - \tau Kr_{+}(x) & \text{for } (\mathrm{typ},\ell) \in \{(+,n),(-,1)\}. \end{cases}$$

We have

$$\begin{aligned} |\rho_{\mathrm{typ}}(y) - \lambda_\ell^{\mathrm{typ}}| &\leqslant |\rho_{\mathrm{typ}}(z) - \lambda_\ell^{\mathrm{typ}}| \\ &\leqslant \frac{1}{|\lambda_\ell^{\mathrm{typ}} - \rho_{\mathrm{typ}'}(z)|}\left[\frac{\varepsilon\sqrt{|\lambda_\ell^{\mathrm{typ}} - \rho_{\mathrm{typ}'}(x)|} + \tau\sqrt{\Gamma}\,\delta_1}{1 - \tau\left(\sqrt{\Gamma}\delta_2 + \delta_3^2\right)}\right]^2 \\ &\quad \times |\rho_{\mathrm{typ}}(x) - \lambda_\ell^{\mathrm{typ}}|, \end{aligned} \tag{9.9}$$

provided that $\tau(\sqrt{\Gamma}\delta_2 + \delta_3^2) < 1$*, where*

$$\delta_1 = \sqrt{|\rho_{\mathrm{typ}}(x) - \lambda_\ell^{\mathrm{typ}}|}\,\|K^{1/2}\{A[\rho_{\mathrm{typ}}(x) + \lambda_\ell^{\mathrm{typ}}] + B\}A^{-1/2}\|_2,$$
$$\delta_2 = \sqrt{\|K^{1/2}AK^{1/2}\|_2\,|\rho_{\mathrm{typ}}(x) - \lambda_\ell^{\mathrm{typ}}| \cdot |\lambda_\ell^{\mathrm{typ}} - \rho_{\mathrm{typ}'}(x)|},$$
$$\delta_3 = \sqrt{\|A^{1/2}K\{A[\rho_{\mathrm{typ}}(x) + \lambda_\ell^{\mathrm{typ}}] + B\}A^{-1/2}\|_2\,|\rho_{\mathrm{typ}}(x) - \lambda_\ell^{\mathrm{typ}}|}.$$

Proof. We will prove the case when $(\mathrm{typ},\ell) = (+,1)$ only. The other cases can be handled in the same way.

Note that $z = x + \tau Kr_+(x) = x + \tau K\boldsymbol{Q}(\rho_+(x))\,x$. We have $\lambda_1^+ \leqslant \rho_+(y) \leqslant \rho_+(z)$, and thus $\rho_+(y) - \lambda_1^+ \leqslant \rho_+(z) - \lambda_1^+$. So it remains to show that $\rho_+(z) - \lambda_1^+$ is no bigger than the right-hand side of (9.9).

Let $M = -\boldsymbol{Q}(\lambda_1^+) \succeq 0$. For any vector w, we have

$$\begin{aligned} \|w\|_M^2 &= -w^{\mathrm{H}}\boldsymbol{Q}(\lambda_1^+)w \\ &= [\rho_+(w) - \lambda_1^+][\lambda_1^+ - \rho_-(w)]\|w\|_A^2, \end{aligned} \tag{9.10}$$
$$\begin{aligned} \|[I + \tau K\boldsymbol{Q}(\lambda_1^+)]w\|_M &= \|[I - \tau KM]w\|_M \\ &\leqslant \varepsilon\|w\|_M. \end{aligned} \tag{9.11}$$

Write

$$\begin{aligned} z &= [I + \tau K\boldsymbol{Q}(\lambda_1^+)]x - \tau K[\boldsymbol{Q}(\lambda_1^+) - \boldsymbol{Q}(\rho_+(x))]x \\ &= [I + \tau K\boldsymbol{Q}(\lambda_1^+)]x + \tau[\rho_+(x) - \lambda_1^+]K[A(\rho_+(x) + \lambda_1^+) + B]x. \end{aligned}$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

Without loss of generality, we may assume that $\|x\|_A = 1$. We have

$$\begin{aligned}
\|z\|_M &= \sqrt{[\rho_+(z) - \lambda_1^+][\lambda_1^+ - \rho_-(z)]}\,\|z\|_A, \quad \text{by (9.10)} \\
\|z\|_M &\leqslant \|[I + \tau K\,\boldsymbol{Q}(\lambda_1^+)]x\|_M \\
&\quad + \tau[\rho_+(x) - \lambda_1^+]\|K[A(\rho_+(x) + \lambda_1^+) + B]x\|_M \\
&\leqslant \varepsilon\|x\|_M + \tau[\rho_+(x) - \lambda_1^+]\sqrt{\Gamma}\,\|[A(\rho_+(x) + \lambda_1^+) + B]x\|_K \\
&\leqslant \varepsilon\sqrt{[\rho_+(x) - \lambda_1^+][\lambda_1^+ - \rho_-(x)]} \\
&\quad + \tau[\rho_+(x) - \lambda_1^+]\sqrt{\Gamma}\,\|K^{1/2}[A(\rho_+(x) + \lambda_1^+) + B]A^{-1/2}\|_2 \\
&= \left[\varepsilon\sqrt{\lambda_1^+ - \rho_-(x)} + \tau\sqrt{\Gamma}\,\delta_1\right]\sqrt{\rho_+(x) - \lambda_1^+}, \qquad (9.12) \\
\|z\|_A &\geqslant \|x\|_A - \tau\|Kr_+(x)\|_A \\
&= 1 - \tau\|Kr_+(x)\|_A, \\
\|Kr_+(x)\|_A &= \|K\,\boldsymbol{Q}(\lambda_1^+)x - K[\boldsymbol{Q}(\lambda_1^+) - \boldsymbol{Q}(\rho_+(x))]x\|_A \\
&\leqslant \|K\,\boldsymbol{Q}(\lambda_1^+)x\|_A + [\rho_+(x) - \lambda_1^+]\|K[A(\rho_+(x) + \lambda_1^+) + B]x\|_A \\
&\leqslant \sqrt{\|K^{1/2}AK^{1/2}\|_2\Gamma}\,\|x\|_M \\
&\quad + [\rho_+(x) - \lambda_1^+]\|A^{1/2}K[A(\rho_+(x) + \lambda_1^+) + B]A^{-1/2}\|_2\|x\|_A \\
&= \sqrt{\Gamma}\delta_2 + \delta_3^2. \qquad (9.13)
\end{aligned}$$

Finally, use

$$\rho_+(z) - \lambda_1^+ = \frac{\|z\|_M^2}{[\lambda_1^+ - \rho_-(z)]\|z\|_A^2} \leqslant \frac{\|z\|_M^2}{[\lambda_1^+ - \rho_-(z)]\cdot[1 - \tau\|Kr_+(x)\|_A]^2}$$

and (9.12) and (9.13) to complete the proof. □

10. Block preconditioned steepest descent/ascent method

The convergence of any of the previous steepest descent/ascent methods can be very slow if $\lambda_1^{\text{typ}} \approx \lambda_2^{\text{typ}}$ or $\lambda_{n-1}^{\text{typ}} \approx \lambda_n^{\text{typ}}$. This is reflected by $\omega_1 \approx \omega_2$ in Theorems 8.2 and 9.1. Often in practice, there are needs to compute the first few extreme eigenpairs, not just the first one. For that purpose, block variations of the methods become particularly attractive for at least the following reasons:

(1) they can simultaneously compute the first k extreme eigenpairs $(\lambda_j^{\text{typ}}, u_j^{\text{typ}})$;

(2) they run more efficiently on modern computer architecture because more computations can be organized into the matrix–matrix multiplication type;

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

(3) they have better rates of convergence to the desired eigenpairs and save overall cost by using a block size that is slightly bigger than the number of asked eigenpairs.

In summary, the benefits of using a block variation are similar to those of using the simultaneous subspace iteration versus the power method [57].

In what follows, we will explain a block steepest descent/ascent method for computing the first few (λ_j^+, u_j^+). The same reasoning applies to other extreme eigenpairs.

Any block variation starts with a given $X_0 \in \mathbb{C}^{n\times n_b}$ with $\operatorname{rank}(X_0) = n_b$, instead of just one vector $\boldsymbol{x}_0 \in \mathbb{C}^n$ previously for the single-vector steepest descent type methods. Here either the jth column of X_0 is already an approximation to u_j^+, or the subspace $\mathcal{R}(X_0)$ contains a good approximation to the subspace spanned by u_j^+ for $1 \leqslant j \leqslant k$, or the canonical angles from $\mathcal{R}([u_1^+, \ldots, u_k^+])$ to $\mathcal{R}(X_0)$ are nontrivial, where $k \leqslant n_b$ is the number of desired eigenpairs. In the latter two cases, a preprocessing is needed to turn the case into the first case:

(1) solve the HQEP for $X_0^{\mathrm{H}} \boldsymbol{Q}(\lambda) X_0$ to get its pos-type eigenpairs $(\boldsymbol{\rho}_{0;j}^+, y_j^+)$;

(2) reset $X_0 := X_0[y_1^+, \ldots, y_{n_b}^+]$.

So we will assume henceforth that the jth column of the given X_0 is an approximation to u_j^+. Now consider generalizing the steepest descent method to a block version. Its typical ith iterative step may well look like the following. Suppose that we have already computed

$$X_i = [x_{i;1}, x_{i;2}, \ldots, x_{i;n_b}] \in \mathbb{C}^{n\times n_b},$$

whose jth column $x_{i;j}$ approximates u_j^+, and

$$\Omega_i = \operatorname{diag}(\boldsymbol{\rho}_{i;1}^+, \boldsymbol{\rho}_{i;2}^+, \ldots, \boldsymbol{\rho}_{i;n_b}^+),$$

whose jth diagonal entry $\boldsymbol{\rho}_{i,j}^+ = \rho_+(x_{i;j})$ approximates λ_j^+. Define the residual matrix

$$R_i = [r_+(x_{i;1}), r_+(x_{i;2}), \ldots, r_+(x_{i;n_b})] = AX_i\Omega_i^2 + BX_i\Omega_i + CX_i.$$

The next set of approximations is computed as follows.

(1) Compute a basis matrix Z of $\mathcal{R}([X_i, R_i])$ by, for example, MGS.

(2) Solve the HQEP for $Z^{\mathrm{H}} \boldsymbol{Q}(\lambda) Z$ to get its pos-type eigenpairs $(\boldsymbol{\rho}_{i+1;j}^+, y_j^+)$, and let $\Omega_{i+1} = \operatorname{diag}(\boldsymbol{\rho}_{i+1;1}^+, \boldsymbol{\rho}_{i+1;2}^+, \ldots, \boldsymbol{\rho}_{i+1;n_b}^+)$.

(3) Set $X_{i+1} = Z[y_1^+, \ldots, y_{n_b}^+]$.

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

In the same way as we explained before, this block steepest descent method can be improved in two directions—extending the search space is one and incorporating preconditioners is the other.

Note that $r_+(x_{i;j}) = \boldsymbol{Q}(\boldsymbol{\rho}_{i;j}^+)x_{i;j}$, and thus

$$\begin{aligned}\mathcal{R}([X_i, R_i]) &= \sum_{j=1}^{n_b} \mathcal{R}([x_{i;j}, \boldsymbol{Q}(\boldsymbol{\rho}_{i;j}^+)x_{i;j}]) \\ &= \sum_{j=1}^{n_b} \mathcal{K}_2(\boldsymbol{Q}(\boldsymbol{\rho}_{i;j}^+), x_{i;j}).\end{aligned}$$

So it is natural to extend the search space $\mathcal{R}([X_\ell, R_\ell])$ through extending each Krylov subspace $\mathcal{K}_2(\boldsymbol{Q}(\boldsymbol{\rho}_{i;j}^+), x_{\ell;j})$ to a high-order one, and of course different Krylov subspaces can be extended to different orders. For simplicity, we will extend each to the mth order. The new extended search subspace now is

$$\sum_{j=1}^{n_b} \mathcal{K}_m(\boldsymbol{Q}(\boldsymbol{\rho}_{i;j}^+), x_{i;j}). \tag{10.1}$$

Define the linear operator

$$\mathscr{R}_i : X \in \mathbb{C}^{n\times n_b} \to \mathscr{R}_i(X) = AX\Omega_i^2 + BX\Omega_i + CX \in \mathbb{C}^{n\times n_b}.$$

Then the subspace in (10.1) can be compactly written as

$$\mathcal{K}_m(\mathscr{R}_i, X_i) = \operatorname{span}\{X_i, \mathscr{R}_i(X_i), \ldots, \mathscr{R}_i^{m-1}(X_i)\}, \tag{10.2}$$

where $\mathscr{R}_i^j(\cdot)$ is understood as successively applying the operator $\mathscr{R}_i$ j times; for example, $\mathscr{R}_i^2(X) = \mathscr{R}_i(\mathscr{R}_i(X))$.

As to incorporate suitable preconditioners, in light of our extensive discussions in Section 9.1, the search subspace should be modified to

$$\sum_{j=1}^{n_b} \mathcal{K}_m(K_{i;j}\boldsymbol{Q}(\boldsymbol{\rho}_{i;j}^+), x_{i;j}), \tag{10.3}$$

where $K_{i;j}$ are the preconditioners, one for each approximate eigenpair $(\boldsymbol{\rho}_{i;j}^+, x_{i;j})$ for $1 \leqslant j \leqslant n_b$ in the ith iterative step. As before, $K_{i;j}$ can be constructed in one of the following two ways.

- $K_{i;j}$ is an approximate inverse of $\boldsymbol{Q}(\tilde{\boldsymbol{\rho}}_{i;j}^+)$ for some $\tilde{\boldsymbol{\rho}}_{i;j}^+$ different from $\boldsymbol{\rho}_{i;j}^+$, ideally closer to λ_j^+ than to any other eigenvalue of $\boldsymbol{Q}(\lambda)$. But this requirement on $\tilde{\boldsymbol{\rho}}_{i;j}^+$ is impractical because the eigenvalue λ_j^+ of $\boldsymbol{Q}(\lambda)$ is unknown. A compromise would be to make $\tilde{\boldsymbol{\rho}}_{i;j}^+$ closer but not equal to $\boldsymbol{\rho}_{i;j}^+$ than to any other $\boldsymbol{\rho}_{i;j}^+$.

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

- Perform an incomplete LDL^{H} factorization (see [54, Ch. 10]) $\boldsymbol{Q}(\tilde{\boldsymbol{\rho}}^{+}_{i;j}) \approx L_{i;j}D_{i;j}L^{\mathrm{H}}_{i;j}$, where $\approx$ includes not only the usual 'approximately equal', but also the case when $\boldsymbol{Q}(\tilde{\boldsymbol{\rho}}^{+}_{i;j}) - L_{i;j}D_{i;j}L^{\mathrm{H}}_{i;j}$ is approximately a low rank matrix, and $D_{i;j} = \mathrm{diag}(\pm 1)$. Finally, set $K_{i;j} = L_{i;j}L^{\mathrm{H}}_{i;j}$.

Algorithm 10.1 is the general framework of a Block Preconditioned Extended Steepest Descent method (BPeSD) which embeds four methods into one:

(1) Block Steepest Descent method: $m = 2$ and all preconditioners $K_{i;j} = I$;

(2) Block Preconditioned Steepest Descent method: $m = 2$ and nontrivial $K_{i;j}$;

(3) Block Extended Steepest Descent method: $m > 2$ and all preconditioners $K_{i;j} = I$;

(4) Block Preconditioned Extended Steepest Descent method: $m > 2$ and nontrivial $K_{i;j}$.

There are two important implementation issues to worry about in turning this general framework into a piece of working code.

(1) In (10.3), a different preconditioner is used for each and every approximate eigenpair $(\boldsymbol{\rho}^{+}_{i;j}, x_{i;j})$ for $1 \leqslant j \leqslant n_b$. While, conceivably, doing so will speed up convergence for each approximate eigenpair because each preconditioner can be constructed to make that approximate eigenpair converge faster, the cost in constructing these preconditioners may likely be too heavy to bear. A more practical approach would be to use one preconditioner K_i for all $K_{i;j}$ aiming at speeding up the convergence of $(\rho^{+}_{i;1}, x_{i;1})$ (or the first few approximate eigenpairs for tightly clustered eigenvalues). Once it (or the first few in the case of a tightly cluster) is determined to be sufficiently accurate, the converged eigenpairs are locked up and deflated, and a new preconditioner is computed to aim at the next nonconverged eigenpairs, and the process continues.

(2) Consider implementing Line 5, that is, generating a basis matrix for the subspace (10.4). In the most general case, Z can be gotten by packing the basis matrices of all

$$\mathcal{K}_m(K_{i;j}\boldsymbol{Q}(\boldsymbol{\rho}^{+}_{i;j}), x_{\ell;j}) \quad \text{for } 1 \leqslant j \leqslant n_b$$

together. There could be two problems with this: (1) such Z could be ill conditioned, that is, the columns of Z may not be sufficiently numerically linearly independent, and (2) the arithmetic operations in building a basis for each $\mathcal{K}_m(K_{i;j}\boldsymbol{Q}(\boldsymbol{\rho}^{+}_{i;j}), x_{i;j})$ are mostly matrix–vector multiplications, straying from one of the purposes: performing most arithmetic operations through matrix–matrix multiplications in order to achieve high performance on modern

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

Algorithm 10.1 Block preconditioned extended steepest descent/ascent method

Given an initial approximation $X_0 \in \mathbb{C}^{n\times n_b}$ with $\operatorname{rank}(X_0) = n_b$, and an integer $m \geqslant 2$, the algorithm computes approximate eigenpairs to $(\lambda_j^{\text{typ}}, u_j^{\text{typ}})$ for $j \in \mathbb{J}$, where $\mathbb{J} = \{1 \leqslant j \leqslant n_b\}$ for computing the few smallest eigenpairs of the given type or $\{n - n_b + 1 \leqslant j \leqslant n\}$ for computing the few largest eigenpairs of the given type.

1: solve the HQEP for $X_0^{\mathrm{H}} \boldsymbol{Q}(\lambda) X_0$ to get its eigenpairs $(\boldsymbol{\rho}_{0;j}^{\text{typ}}, y_j^{\text{typ}})$;
2: $X_0 = X_0[y_1^{\text{typ}}, \ldots, y_{n_b}^{\text{typ}}]$, $\hat{\mathbb{J}} = \{1 \leqslant j \leqslant n_b\}$;
3: **for** $i = 0, 1, \ldots$ **do**
4: construct preconditioners $K_{i;j}$ for $j \in \hat{\mathbb{J}}$;
5: compute a basis matrix Z of the subspace

$$\sum_{j\in\hat{\mathbb{J}}} \mathcal{K}_m(K_{i;j}\, \boldsymbol{Q}(\boldsymbol{\rho}_{i;j}^{\text{typ}}), x_{i;j}), \tag{10.4}$$

and let n_Z be its dimension and $\hat{\mathbb{J}} = \{1 \leqslant j \leqslant n_b\}$ for computing the few smallest eigenpairs of the given type or $\{n_Z - n_b + 1 \leqslant j \leqslant n_Z\}$ for computing the few largest eigenpairs of the given type;
6: compute the n_b eigenpairs of $Z^{\mathrm{H}} \boldsymbol{Q}(\lambda) Z$: $(\boldsymbol{\rho}_{i+1;j}^{\text{typ}}, y_j^{\text{typ}})$ for $j \in \hat{\mathbb{J}}$ and let $\Omega_{i+1} = \operatorname{diag}(\ldots, \boldsymbol{\rho}_{i+1;j}^{\text{typ}}, \ldots)$ whose diagonal entries are those for $j \in \hat{\mathbb{J}}$;
7: $X_{i+1} = ZW$, where $W = [\ldots, y_j^{\text{typ}}, \ldots]$ whose columns are those for $j \in \hat{\mathbb{J}}$;
8: **end for**
9: **return** approximate eigenpairs to $(\lambda_j^{\text{typ}}, u_j^{\text{typ}})$ for $j \in \mathbb{J}$.

computers. To address these two problems, we may do a tradeoff by using $K_{i;j} \equiv K_i$ for all j. This may likely degrade the effectiveness of the preconditioner per step in terms of rates of convergence for all approximate eigenpairs $(\boldsymbol{\rho}_{i;j}^{+}, x_{i;j})$ but may achieve overall gain in using less time because then the code will run much faster in matrix–matrix operations, not to mention the saving in constructing just one preconditioner K_i instead of n_b different preconditioners $K_{i;j}$. To simplify our discussion below, we will drop the subscript i for readability. Since $K_{i;j} \equiv K$ for all j, (10.4) is the same as

$$\mathcal{K}_m(K\mathscr{R}, X) = \operatorname{span}\{X, K\mathscr{R}(X), \ldots, [K\mathscr{R}]^{m-1}(X)\}, \tag{10.5}$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

where $[K\mathscr{R}]^j(\cdot)$ is understood as successively applying the operator $K\mathscr{R}$ j times; for example, $[K\mathscr{R}]^2(X) = K\mathscr{R}_\ell(K\mathscr{R}(X))$. A basis matrix

$$Z = [Z_1, Z_2, \ldots, Z_m]$$

can be computed by the following block Arnoldi-like process.

1: $Z_1 T = X$ (MGS);
2: **for** $i = 2$ to m **do**
3: $\quad Y = K(AZ_{i-1}\Omega^2 + BZ_{i-1}\Omega + CZ_{i-1})$;
4: $\quad$ **for** $j = 1$ to $i-1$ **do**
5: $\quad\quad T = Z_j^{\mathrm{H}} Y$; $Y = Y - Z_j T$;
6: $\quad$ **end for**
7: $\quad Z_i T = Y$ (MGS);
8: **end for**

There is a possibility that at Line 7 Y is numerically not of full column rank. If it happens, it poses no difficulty at all. In running MGS on Y's columns, anytime if a column is deemed linearly dependent on previous columns, that column should be deleted, along with the corresponding $\boldsymbol{\rho}_j^+$ from Ω to shrink its size by 1 as well. At the completion of MGS, Z_i will have fewer columns than Y, and the size of Ω is shrunk accordingly. Ultimately, at the end, the columns of Z are orthonormal; that is, $Z^{\mathrm{H}}Z = I$ (of apt size), which may fail to an unacceptable level due to roundoff; so some form of reorthogonalization should be incorporated.

11. Conjugate gradient method

Again because of the equations in (5.8), the nonlinear CG type method [**48**, **61**] and its variations are natural candidates for computing the first or last eigenpair $(\lambda_j^{\mathrm{typ}}, u_j^{\mathrm{typ}})$, and their block variations can also be devised to simultaneously compute the first or last few eigenpairs $(\lambda_j^{\mathrm{typ}}, u_j^{\mathrm{typ}})$. Since much of the machinery including gradients and preconditioning has already been built up, what remain are more or less simple adaptations of CG type methods [**36**] for the generalized Hermitian eigenvalue problem to the current case.

11.1. Preconditioned conjugate gradient method. Single-vector CG type methods heavily rely on the line-search problem (8.5)–(8.7) which was solved by projecting the original order-n HQEP for $\boldsymbol{Q}(\lambda)$ to an order-2 HQEP for $Y^{\mathrm{H}}\boldsymbol{Q}(\lambda)Y$ without actually computing the optimal parameter t_{opt}, and thus the next approximation y as in (8.7) for the steepest descent/ascent method and

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

Algorithm 11.1 Preconditioned conjugate gradient method

Given an initial approximation $\boldsymbol{x}_0$ to u_ℓ^{typ}, a (positive definite) preconditioner K, and a relative tolerance `rtol`, the algorithm computes an approximate pair to $(\lambda_\ell^{\text{typ}}, u_\ell^{\text{typ}})$ with the prescribed `rtol`.

$\boldsymbol{x}_0 = \boldsymbol{x}_0/\|\boldsymbol{x}_0\|_2$, , $\boldsymbol{\rho}_0 = \rho_{\text{typ}}(\boldsymbol{x}_0)$, $\boldsymbol{r}_0 = r_{\text{typ}}(\boldsymbol{x}_0)$, $\boldsymbol{x}_0 = -K\boldsymbol{r}_0$;
for $i = 0, 1, \ldots$ **do**
  if $\|\boldsymbol{r}_i\|_2/(|\boldsymbol{\rho}_i|^2\|A\boldsymbol{x}_i\| + |\boldsymbol{\rho}_i|\,\|B\boldsymbol{x}_i\| + \|C\boldsymbol{x}_i\|) \leqslant$ `rtol` **then**
    BREAK;
  else
    solve the HQEP for $Y_i^{\text{H}}\boldsymbol{Q}(\lambda)Y_i$, where $Y_i = [\boldsymbol{x}_i, \boldsymbol{x}_i]$ to get its eigenvalues $\mu_j^{\pm}$ as in (8.8) and eigenvectors $y_j^{\pm}$;
    select the next approximate eigenpair $(\mu, Y_i v)$ according to the table (8.9);
    compute $\alpha_i = t_{\text{opt}}$ as in (11.2) and then y as in (8.7) with $x = \boldsymbol{x}_i$ and $p = \boldsymbol{x}_i$;
    $\boldsymbol{x}_{i+1} = y/\|y\|_2$;
    set $\boldsymbol{\rho}_{i+1} = \rho_{\text{typ}}(\boldsymbol{x}_{i+1})$, $\boldsymbol{r}_{i+1} = r_{\text{typ}}(\boldsymbol{x}_{i+1})$, $\boldsymbol{x}_{i+1} = -K\boldsymbol{r}_{i+1} + \beta_i\boldsymbol{x}_i$, where β_i is commonly given by either one of

$$\text{either} \quad \beta_i = \frac{\boldsymbol{r}_{i+1}^{\text{H}} K \boldsymbol{r}_{i+1}}{\boldsymbol{r}_i^{\text{H}} K \boldsymbol{r}_i} \quad \text{or} \quad \beta_i = \frac{\boldsymbol{r}_{i+1}^{\text{H}} K (\boldsymbol{r}_{i+1} - \boldsymbol{r}_i)}{\boldsymbol{r}_i^{\text{H}} K \boldsymbol{r}_i}; \tag{11.1}$$

  end if
end for
return $(\boldsymbol{\rho}_i, \boldsymbol{x}_i)$ as an approximate eigenpair to $(\lambda_\ell^{\text{typ}}, u_\ell^{\text{typ}})$.

its variations. The outcome of it is that the computed next approximation is a (complex) scalar multiple of y in (8.7). This is good enough for the steepest descent/ascent method, but not for the CG method, for which y in (8.7) needs to be computed. We now show how this y can be recovered from the approximation given in table (8.9). Let (μ, Yv) be selected according to the table, and write $v = [v_1, v_2]^{\text{T}}$ and $\hat{y} = Yv = v_1 x + v_2 p$. Thus

$$t_{\text{opt}} = v_2/v_1 \quad \text{if } v_1 \neq 0, \text{ and } \infty \text{ otherwise.} \tag{11.2}$$

With this, set y as in (8.7).

Our discussions on selecting a good preconditioner in Section 9.1 should be followed. Algorithm 11.1 presents the framework for the single-vector preconditioned conjugate gradient method for $\boldsymbol{Q}(\lambda)$.

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

Algorithm 11.2 Locally optimal block preconditioned extended conjugate gradient method

Given an initial approximation $X_0 \in \mathbb{C}^{n\times n_b}$ with $\operatorname{rank}(X_0) = n_b$, and an integer $m \geqslant 2$, the algorithm computes approximate eigenpairs to $(\lambda_j^{\text{typ}}, u_j^{\text{typ}})$ for $j \in \mathbb{J}$, where $\mathbb{J} = \{1 \leqslant j \leqslant n_b\}$ for computing the few smallest eigenpairs of the given type or $\{n - n_b + 1 \leqslant j \leqslant n\}$ for computing the few largest eigenpairs of the given type.

1: solve the HQEP for $X_0^{\mathrm{H}} \boldsymbol{Q}(\lambda) X_0$ to get its eigenpairs $(\boldsymbol{\rho}_{0;j}^{\text{typ}}, y_j^{\text{typ}})$;
2: $X_0 = X_0[y_1^{\text{typ}}, \ldots, y_{n_b}^{\text{typ}}]$, $X_{-1} = 0$, $\hat{\mathbb{J}} = \{1 \leqslant j \leqslant n_b\}$;
3: **for** $i = 0, 1, \ldots$ **do**
4: construct preconditioners $K_{i;j}$ for $j \in \hat{\mathbb{J}}$;
5: compute a basis matrix Z of the subspace

$$\sum_{j\in\hat{\mathbb{J}}} \mathcal{K}_m(K_{i;j}\boldsymbol{Q}(\boldsymbol{\rho}_{i;j}), x_{i;j}) + \mathcal{R}(X_{i-1}), \tag{11.3}$$

and let n_Z be its dimension and $\hat{\mathbb{J}} = \{1 \leqslant j \leqslant n_b\}$ for computing the few smallest eigenpairs of the given type or $\{n_Z - n_b + 1 \leqslant j \leqslant n_Z\}$ for computing the few largest eigenpairs of the given type;
6: compute the n_b eigenpairs of $Z^{\mathrm{H}}\boldsymbol{Q}(\lambda)Z$: $(\boldsymbol{\rho}_{i+1;j}^{\text{typ}}, y_j^{\text{typ}})$ for $j \in \hat{\mathbb{J}}$ and let $\Omega_{i+1} = \operatorname{diag}(\ldots, \boldsymbol{\rho}_{i+1;j}^{\text{typ}}, \ldots)$ whose diagonal entries are those for $j \in \hat{\mathbb{J}}$;
7: $X_{i+1} = ZW$, where $W = [\ldots, y_j^{\text{typ}}, \ldots]$ whose columns are those for $j \in \hat{\mathbb{J}}$;
8: **end for**
9: **return** approximate eigenpairs to $(\lambda_j^{\text{typ}}, u_j^{\text{typ}})$ for $j \in \mathbb{J}$.

11.2. Locally optimal block preconditioned extended conjugate gradient method. When it comes to eigenvalue computations by CG type methods, CG's locally optimal variations [**51**, **62**] combined with preconditioning and blocking are more preferable than the usual single-vector CG method as in Algorithm 11.1 [**4**, **28**, **36**]. In Algorithm 11.2, we present a framework of the so-called *Locally Optimal Block Preconditioned Extended Conjugate Gradient Method* (LOBPeCG), whose different implementation choice gives rise to a list of CG type methods which we will not elaborate.

The two important implementation issues we discussed for Algorithm 10.1 (Block Preconditioned Extended Steepest Descent method) after its introduction essentially apply here, except that some changes are needed in the computation of Z at Line 5 here.

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

First X_{i-1} can be replaced by something else. Specifically, we modify Lines 2, 6, and 8 of Algorithm 11.2 to

---

2: $X_0 = X_0 W$, and $Y_0 = 0$, $\hat{\mathbb{J}} = \{1 \leqslant j \leqslant n_b\}$;

5: compute a basis matrix Z of the subspace

$$\sum_{j \in \hat{\mathbb{J}}} \mathcal{K}_m(K_{i:j}\, \boldsymbol{Q}(\boldsymbol{\rho}_{i:j}), x_{i:j}) + \mathcal{R}(Y_i), \tag{11.4}$$

such that $\mathcal{R}(Z_{(:,1:n_b)}) = \mathcal{R}(X_i)$. Let n_Z be its dimension and $\hat{\mathbb{J}} = \{1 \leqslant j \leqslant n_b\}$ for computing the few smallest eigenpairs of the given type or $\{n_Z - n_b + 1 \leqslant j \leqslant n_Z\}$ for computing the few largest eigenpairs of the given type;

7: X_{i+1}=ZW, where W=$[\ldots, y_j^{\text{typ}}, \ldots]$ whose columns are those for $j \in \hat{\mathbb{J}}$, $Y_{i+1} = Z_{(:,n_b+1:(m+1)n_b)} W_{(n_b+1:(m+1)n_b,:)}$;

Next we will compute a basis matrix for the subspace (11.3) or (11.4). For better performance (by using more matrix–matrix multiplications), we will assume that $K_{i:j} \equiv K_i$ for all j for simplification. Dropping the subscript i for readability, we see that (11.4) is the same as

$$\mathcal{K}_m(K\mathscr{R}, X) + \mathcal{R}(Y) = \text{span}\{X, K\mathscr{R}(X), \ldots, [K\mathscr{R}]^{m-1}(X)\} + \mathcal{R}(Y). \tag{11.5}$$

We will first compute a basis matrix $[Z_1, Z_2, \ldots, Z_m]$ for $\mathcal{K}_m(K\mathscr{R}, X)$ by the block Arnoldi-like process outlined at the end of Section 10. In particular, $\mathcal{R}(Z_1) = \mathcal{R}(X)$. Then orthogonalize Y against $[Z_1, Z_2, \ldots, Z_m]$ to get Z_{m+1} satisfying $Z_{m+1}^{\text{H}} Z_{m+1} = I$. Finally, take $Z = [Z_1, Z_2, \ldots, Z_{m+1}]$.

Our understanding for precise convergence behaviors of these CG type methods is very limited, despite overwhelming numerical evidence that CG type methods are superior to steepest descent/ascent type methods. This is an area that needs further research, even in the case of using similar CG type methods in the linear eigenvalue problem [**36**]. But we point out that per step Algorithm 11.2 produces better approximations than Algorithm 10.1 does because the former uses a search subspace that contains the one used by the latter. In view of this, the convergence estimates in Theorems 8.2, 9.1, and 9.2 are mathematically correct for locally optimal preconditioned extended conjugate gradient method, that is, Algorithm 11.2 with $n_b = 1$. Nonetheless, we believe that the actual convergence rate should be much better than these estimates suggest.

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

12. Numerical examples

In this section, we will present a couple of examples to demonstrate the numerical behavior of Algorithm 11.2, which often performs much better than the steepest descent/ascent type methods. In presenting numerical results, we will use the normalized residuals

$$\frac{\|\boldsymbol{Q}(\mu_i)x_i\|_2}{(\|A\|_1\mu_i^2 + \|B\|_1|\mu_i| + \|C\|_1)\|x_i\|_2}$$

to show the convergence progress for approximations (μ_i, x_i) to a particular eigenpair versus the iteration index, where the matrix ℓ_1 operator norms $\|A\|_1$, $\|B\|_1$, and $\|C\|_1$ are used, more for computational convenience than anything else, as any other norm would serve the same purpose just as well.

EXAMPLE 12.1. This is the problem **Wiresaw1** in the collection [6]. It is actually a gyroscopic QEP arising in the vibration analysis of a wiresaw [70], but it leads to an HQEP. Here

$$A = \frac{1}{2}I_n, \quad C = \frac{(\nu^2-1)\pi^2}{2}\,\mathrm{diag}(1^2, 2^2, \ldots, n^2),$$
$$B = \iota\,(b_{ij}) \quad \text{with } b_{ij} = \begin{cases} \nu\dfrac{4ij}{i^2-j^2} & \text{if } i+j \text{ is odd,} \\ 0 & \text{otherwise,} \end{cases}$$

where $\iota = \sqrt{-1}$ is the imaginary unit and ν is a real nonnegative parameter corresponding to the speed of the wire. For $0 < \nu < 1$, $\boldsymbol{Q}(0) = C$ is negative definite, and thus $\boldsymbol{Q}(\lambda) = \lambda^2 A + \lambda B + C$ is hyperbolic by Theorem 3.1. Moreover

$$\lambda_i^- < 0 < \lambda_j^+ \quad \text{for all } i, j.$$

Therefore it is rather natural to use $K = -C^{-1}$ as a preconditioner when it comes to computing the few smallest λ_j^+ or largest λ_i^-, or for testing purposes some approximations to C^{-1} such as those corresponding to the linear conjugate gradient methods.

We ran Algorithm 11.2 with $n_b = 10$, $m = 2$ and random $X_0 = \texttt{randn}(n, n_b)$ on this example for $n = 1000$ and $\nu = 0.8$ without or with preconditioners

$$K \approx \begin{cases} [\boldsymbol{Q}(\pm 6.0 \cdot 10^3)]^{-1} & \text{for largest } \lambda_j^+ \text{ or smallest } \lambda_j^-, \\ -[\boldsymbol{Q}(0)]^{-1} = -C^{-1} & \text{for smallest } \lambda_j^+ \text{ or largest } \lambda_j^-, \end{cases} \tag{12.1}$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

implemented through the linear conjugate gradient method with stopping criteria of normalized residuals for the involved linear systems being no bigger than 10^{-1} or reaching the maximum number of CG steps, which is 10. We have already explained the use of $-C^{-1}$ or its approximations as possible preconditioners. After running Algorithm 11.2 without any preconditioner, we noticed that all $\lambda_j^{\pm}$ lie in $(-6.0 \cdot 10^3, 6.0 \cdot 10^3)$, which leads to the use of $[\boldsymbol{Q}(\pm 6.0 \cdot 10^3)]^{-1}$ in (12.1).

Figure 12.1 plots the residual history for computing the largest or smallest few λ_i^{typ}, where the left column is for without any preconditioner while the right column is for with the preconditioners as given in (12.1). We notice that without using any preconditioner Algorithm 11.2 performed poorly for computing smallest λ_j^+ or largest λ_j^-, but reasonably well for largest λ_j^+ or smallest λ_j^-. The effectiveness of the preconditioners as in (12.1) is rather evident by comparing the plots in the two columns.

EXAMPLE 12.2. This is [**23**, Example 5], where $A = I_n$,

$$B = \xi \begin{bmatrix} 20 & -10 & & & & \\ -10 & 30 & -10 & & & \\ & \ddots & \ddots & \ddots & & \\ & & -10 & 30 & -10 & \\ & & & -10 & 20 & \end{bmatrix}, \quad C = \begin{bmatrix} 15 & -5 & & & & \\ -5 & 15 & -5 & & & \\ & \ddots & \ddots & \ddots & & \\ & & -5 & 15 & -5 & \\ & & & -5 & 15 & \end{bmatrix},$$

and ξ is a parameter. We take $n = 1000$ and $\xi = 1.1$. This is a pathological example in the sense that most eigenvalues are close to one another—they share about three significant decimal digits with their neighbors, except λ_1^+ and λ_2^+, which has a gap from the rest. When running Algorithm 11.2 with $m = 2$ and various different n_b, we noticed that the algorithm really had a hard time computing all extreme λ_j^{typ} even with some preconditioner $K = \pm[\boldsymbol{Q}(\mu)]^{-1}$ with $\mu \in (\lambda_n^-, \lambda_1^+)$ or $\mu > \lambda_n^+$ or $\mu < \lambda_1^-$ purposely selected, except for λ_1^+ and λ_2^+, which are rather easy to compute, actually. Figure 12.2 plots the residual history for computing λ_1^+ and λ_2^+, where the left plot is for without any preconditioner while the right plot is for with a preconditioner $K \approx [\boldsymbol{Q}(-8.0)]^{-1}$ implemented through the linear conjugate gradient method with the same stopping criteria as in the previous example.

13. Concluding remarks

We have performed a systematic study of the hyperbolic quadratic eigenvalue problem $\boldsymbol{Q}(\lambda) = \lambda^2 A + \lambda B + C$. Such a problem usually arises from dynamical systems with heavy friction. Such a system appears, for example, in an elevator

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

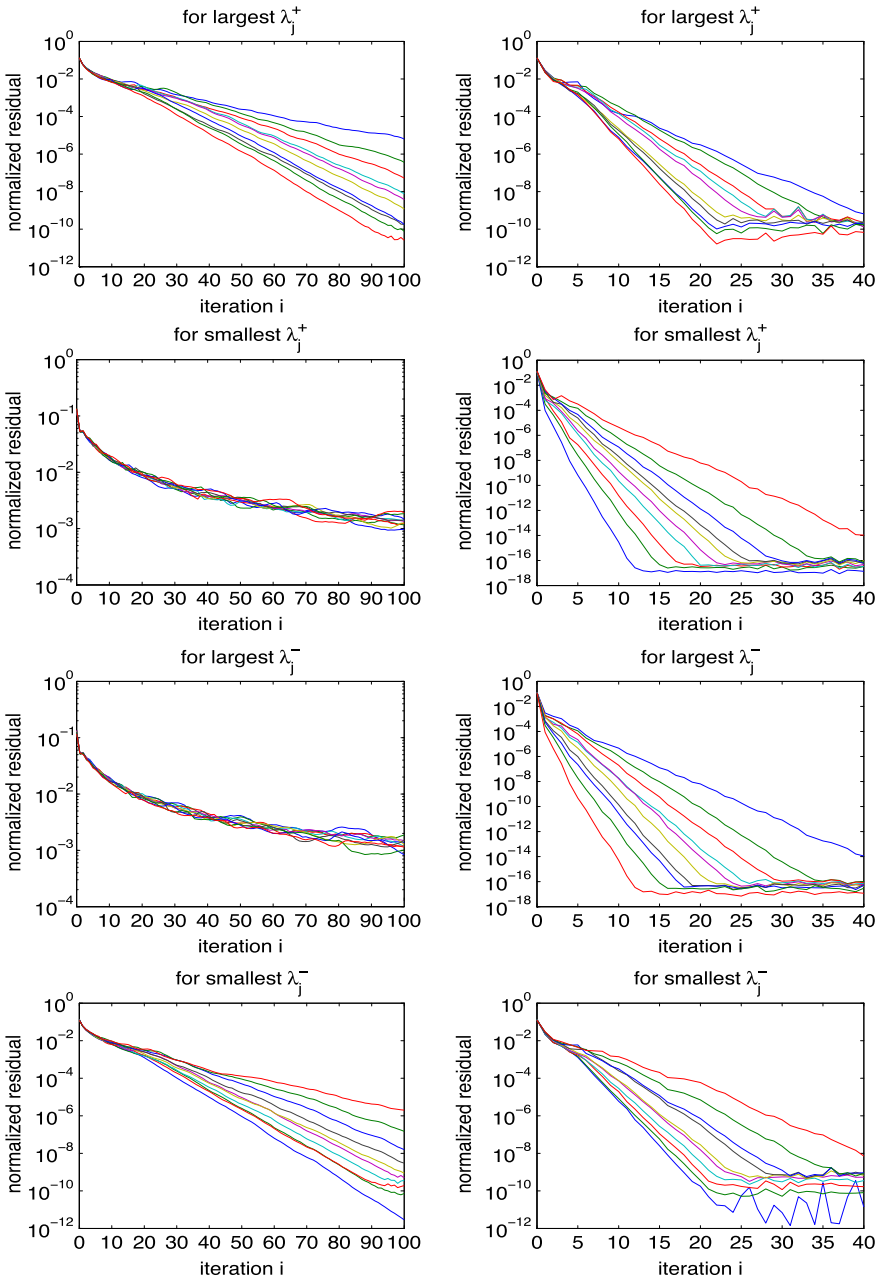

Figure 12.1. Residual history for running Algorithm 11.2 on Example 12.1.

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

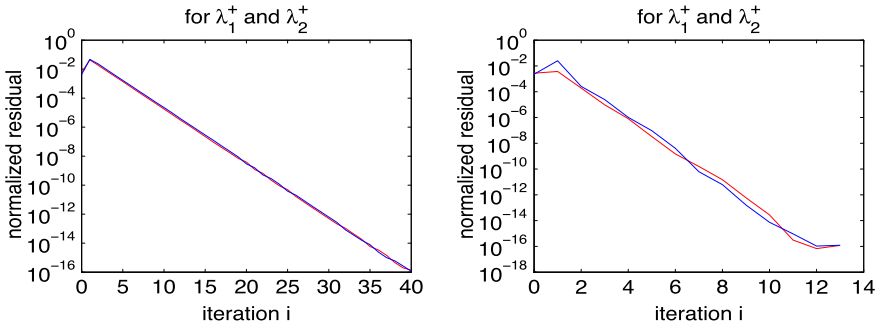

Figure 12.2. Residual history for running Algorithm 11.2 on Example 12.2 for computing λ_1^+ and λ_2^+.

or car braking system. It shares many characteristics with the standard Hermitian eigenvalue problem in the category of usual standard linear eigenvalue problems, and it has attracted considerable attention in the past. Most of the results were collected in [**17**, **45**, **67**].

Our contributions in this paper lie on two fronts. Theoretically, we have established Amir-Moéz/Wielandt–Lidskii type min–max principles for the sums of selected eigenvalues and, as corollaries, Fan trace min/max type principles, and also perturbation results in the spectral norm, as well as general unitarily invariant norms on how the eigenvalues will change if A, B, C are perturbed. Numerically, we have justified a naturally extended Rayleigh–Ritz type procedure, with the existing and newly established min–max principles, and why the procedure will produce the best approximations to eigenvalues/eigenvectors. We proposed steepest descent/ascent and CG type methods for computing extreme eigenpairs, and established convergence results, including the rate of convergence for the steepest descent/ascent methods, which shed light on preconditioning in what constitutes a good preconditioner and how to construct one.

Block steepest descent/ascent type methods often perform much better in practice than their single-vector counterparts. But their exact rates of convergence are hard to establish. Experience shows that their corresponding locally optimal CG type methods perform even better, but then again we do not know the exact rates of convergence for locally optimal CG type methods, either. It is recommended that locally optimal CG type methods should be preferred to their corresponding steepest descent/ascent type methods.

Despite the many successes we have had in this study in extending the important results (both theoretically and numerically) for the standard Hermitian eigenvalue problem, there is more work to be done. We list a few here for further work.

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

- We established perturbation bounds for eigenvalues, but did not do so for eigenvectors/eigenspaces. The latter is worth investigating, too. We expect that $\min_x \varsigma_0(x)$ will play a role.
- Many results in this paper should be extensible to hyperbolic matrix polynomials of degrees higher than 2 [45]. We are working on this, and results will be detailed in a separate paper.
- Higham *et al.* [26] expanded hyperbolic quadratic matrix polynomials to include the case when A is positive semidefinite, calling them definite matrix polynomials. Conceivably, many results in this paper may be extensible to quadratic definite matrix polynomials in the sense of [26], but care must be taken to deal with infinite eigenvalues.

Acknowledgements

The authors are indebted to Professor N. Higham for his careful reading of the manuscript and numerous handwritten corrections on the manuscript. The authors also wish to thank the three anonymous referees for their constructive comments and suggestions. These corrections, comments, and suggestions greatly improved the presentation. Liang was supported in part by China Scholarship Council and National Natural Science Foundation of China NSFC-61075119. This work is primarily done while this author was a visiting student, from August 2011 to September 2013, at the Department of Mathematics, University of Texas at Arlington, Arlington, TX 76019, USA. Li was supported in part by NSF grants DMS-1115834 and DMS-1317330, and a Research Gift Grant from Intel Corporation.

Appendix A. Proof of Theorems 6.1 and 6.2

Besides $A \succ 0$, the other key condition for $\boldsymbol{Q}(\lambda) = \lambda^2 A + \lambda B + C$ to be hyperbolic is

$$[\varsigma(x)]^2 = (x^{\mathrm{H}}Bx)^2 - 4(x^{\mathrm{H}}Ax)(x^{\mathrm{H}}Cx) > 0 \quad \text{for all } 0 \neq x \in \mathbb{C}^n. \tag{3.2}$$

We first establish a condition in Lemma A.1 under which a condition like (3.2) is weakly satisfied for all convex combination $(1-t)\,\boldsymbol{Q}(\lambda) + t\,\widetilde{\boldsymbol{Q}}(\lambda)$ in the sense that

$$g(t) := (x^{\mathrm{H}}[B + t\Delta B]x)^2 - 4(x^{\mathrm{H}}[A + t\Delta A]x)(x^{\mathrm{H}}[C + t\Delta C]x) \geqslant 0 \tag{A.1}$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

for all $0 \leqslant t \leqslant 1$. To this end, we define

$$\phi(x) := (x^{\mathrm{H}}\Delta Bx)^2 - 4(x^{\mathrm{H}}\Delta Ax)(x^{\mathrm{H}}\Delta Cx), \tag{A.2}$$

$$\psi(x) := (x^{\mathrm{H}}Bx)(x^{\mathrm{H}}\Delta Bx) - 2(x^{\mathrm{H}}Ax)(x^{\mathrm{H}}\Delta Cx) - 2(x^{\mathrm{H}}Cx)(x^{\mathrm{H}}\Delta Ax), \tag{A.3}$$

and define $\tilde{\phi}(x)$ and $\tilde{\psi}(x)$ in the same way, except by swapping the positions of A, B, and C with those of $\widetilde{A}$, $\widetilde{B}$, and $\widetilde{C}$. It can be verified that

$$\tilde{\phi}(x) = \phi(x), \quad \tilde{\psi}(x) = -\psi(x) - \phi(x).$$

Correspondingly,

$$\tilde{g}(t) := (x^{\mathrm{H}}[\widetilde{B} - t\Delta B]x)^2 - 4(x^{\mathrm{H}}[\widetilde{A} - t\Delta A]x)(x^{\mathrm{H}}[\widetilde{C} - t\Delta C]x) \equiv g(1-t). \tag{A.4}$$

By definition, if $A \succ 0$, then $\boldsymbol{Q}(\lambda)$ is hyperbolic if and only if $g(0) > 0$ for any nonzero $x \in \mathbb{C}^n$, and if $\widetilde{A} \succ 0$, then $\widetilde{\boldsymbol{Q}}(\lambda)$ is hyperbolic if and only if $g(1) > 0$ for any nonzero $x \in \mathbb{C}^n$.

LEMMA A.1. *Suppose that* $\min\{g(0), g(1)\} \geqslant 0$. *Then* $g(t) \geqslant 0$ *for all* $0 \leqslant t \leqslant 1$ *and nonzero* $x \in \mathbb{C}^n$ *if and only if*

$$\min\{\phi(x), -\psi(x), -\tilde{\psi}(x), \psi(x)^2 - \phi(x)\varsigma(x)^2\} \leqslant 0 \quad \textit{for all } x \neq 0. \tag{A.5}$$

Proof. The condition (A.5) is equivalent to that, for any nonzero x, at least one of

$$\phi(x) \leqslant 0, \quad \psi(x) \geqslant 0, \quad \tilde{\psi}(x) = -\psi(x) - \phi(x) \geqslant 0, \quad \psi(x)^2 - \phi(x)\varsigma(x)^2 \leqslant 0 \tag{A.6}$$

holds. Note that $g(0) \geqslant 0$ and $g(1) \geqslant 0$ by assumption.

We first prove that (A.5) implies that $g(t) \geqslant 0$ for all $0 \leqslant t \leqslant 1$ and for any nonzero $x \in \mathbb{C}^n$. To this end, we expand $g(t)$ in (A.1) and $\tilde{g}(t)$ in (A.4) to get

$$g(t) = \varsigma(x)^2 + 2\psi(x)t + \phi(x)t^2, \tag{A.7a}$$

$$\tilde{g}(t) = \tilde{\varsigma}(x)^2 + 2\tilde{\psi}(x)t + \phi(x)t^2, \tag{A.7b}$$

and let $0 \leqslant t \leqslant 1$ and $0 \neq x \in \mathbb{C}^n$.

(1) If $\phi(x) \leqslant 0$, then, by (A.7a), $g(t)$ is concave, and thus $g(t) \geqslant (1-t)g(0) + tg(1) \geqslant 0$.

(2) If $\psi(x) \geqslant 0$, then, by (A.7a),

$$\begin{aligned} g(t) &\geqslant \varsigma(x)^2 + 2\psi(x)t^2 + \phi(x)t^2 \\ &= (1-t^2)g(0) + t^2 g(1) \\ &\geqslant 0. \end{aligned}$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

(3) If $\tilde{\psi}(x) \geqslant 0$, then, similarly by (A.7b), $\tilde{g}(t) \geqslant (1-t^2)\tilde{g}(0) + t^2\tilde{g}(1) \geqslant 0$.

(4) Consider the case when $\psi(x)^2 - \phi(x)\varsigma(x)^2 \leqslant 0$. Suppose that $\phi(x) > 0$ (the case when $\phi(x) \leqslant 0$ has already been dealt with). Then $g(t)$ is a nontrivial quadratic function and it has at most one zero in $\mathbb{R}$. Then $g(t) \geqslant 0$ for all $0 \leqslant t \leqslant 1$.

Next, for the necessity of (A.5), suppose there were an $x \neq 0$ that violated all inequalities in (A.6); that is,

$$\phi(x) > 0, \quad \psi(x) < 0, \quad -\tilde{\psi}(x) = \psi(x)+\phi(x) > 0, \quad \psi(x)^2-\phi(x)\varsigma(x)^2 > 0.$$

Then

$$\min_t g(t) = -\frac{\psi(x)^2 - \phi(x)\varsigma(x)^2}{\phi(x)} < 0,$$

and $\min_t g(t)$ is attained at $t_{\min} = -\psi(x)/\phi(x) \in (0,1)$, contradicting the assumption that $g(t) \geqslant 0$ for $0 \leqslant t \leqslant 1$. □

LEMMA A.2. *Suppose that $\boldsymbol{Q}(\lambda)$ is hyperbolic, and adopt the notation introduced in Theorem 4.2.*

(1) *If $\lambda_0 \in (\lambda_n^-, \lambda_1^+)$, then* $\operatorname{diag}(-C_{\lambda_0}, A) = \operatorname{diag}(-\boldsymbol{Q}(\lambda_0), A) \succ 0$.

(2) *If $\lambda_0 \in [\lambda_n^+, +\infty)$, then $\boldsymbol{Q}_{\lambda_0}(\lambda)$ is overdamped; that is, $B_{\lambda_0} \succ 0$ and $C_{\lambda_0} \succeq 0$. Moreover,*

$$-(\lambda_n^- + \lambda_n^+ - 2\lambda_0)A \preceq B_{\lambda_0} \preceq -(\lambda_1^- + \lambda_1^+ - 2\lambda_0)A, \tag{A.8}$$
$$(\lambda_n^- - \lambda_0)(\lambda_n^+ - \lambda_0)A \preceq C_{\lambda_0} \preceq (\lambda_1^- - \lambda_0)(\lambda_1^+ - \lambda_0)A. \tag{A.9}$$

(3) *If $\|A^{-1/2}\Delta A A^{-1/2}\|_2 < 1$, then $\widetilde{A} \succ 0$.*

Proof. Item (1) is a consequence of Theorem 3.1 and (6.2c). For (A.8) of item (2), we have, for any nonzero x,

$$\begin{aligned} x^{\mathrm{H}} B_{\lambda_0} x &= 2\lambda_0 x^{\mathrm{H}} A x + x^{\mathrm{H}} B x \\ &= x^{\mathrm{H}} A x \left(2\lambda_0 + \frac{x^{\mathrm{H}} B x}{x^{\mathrm{H}} A x}\right) \\ &= x^{\mathrm{H}} A x \big(2\lambda_0 - [\rho_+(x) + \rho_-(x)]\big), \end{aligned}$$

which, together with (5.5), yields (A.8). For (A.9), we have, for any nonzero x,

$$x^{\mathrm{H}} C_{\lambda_0} x = x^{\mathrm{H}} \boldsymbol{Q}(\lambda_0) x = x^{\mathrm{H}} A x [\lambda_0 - \rho_+(x)][\lambda_0 - \rho_-(x)],$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

which, together with (5.5), yields (A.9). For item (3), we notice that the smallest eigenvalue of $A^{-1/2}\widetilde{A}A^{-1/2}$ satisfies

$$\lambda_{\min}(A^{-1/2}\widetilde{A}A^{-1/2}) = 1 + \lambda_{\min}(A^{-1/2}\Delta A A^{-1/2}) \geqslant 1 - \|A^{-1/2}\Delta A A^{-1/2}\|_2 > 0,$$

provided that $\|A^{-1/2}\Delta A A^{-1/2}\|_2 < 1$. □

Each of many expressions below is in its compact form for two. For example, (A.10) includes two displayed equations: one for $\Delta\rho_+$ and one for $\Delta\rho_+$, with all $\pm$ selected as either $+$ or $-$, accordingly.

LEMMA A.3. *If* (A.5) *and* (6.7) *hold, then for any* $x \neq 0$ *there exists* $0 \leqslant \xi \leqslant 1$ *such that*

$$\Delta\rho_\pm(x) = \delta^\pm(x,\xi) := \pm\left[\delta_3(x,\xi) - \frac{x^{\mathrm{H}}Ax}{x^{\mathrm{H}}\widetilde{A}x}\,\delta_2^\pm(x)\right], \tag{A.10}$$

where

$$\delta_2^\pm(x) = \frac{\rho_\pm(x)^2(x^{\mathrm{H}}\Delta A x) + \rho_\pm(x)(x^{\mathrm{H}}\Delta B x) + x^{\mathrm{H}}\Delta C x}{\varsigma(x)}, \tag{A.11a}$$

$$\delta_3(x,\xi) = \frac{\varsigma(x)^2\phi(x) - \psi(x)^2}{4(x^{\mathrm{H}}\widetilde{A}x)[\varsigma(x)^2 + 2\psi(x)\xi + \phi(x)\xi^2]^{3/2}}, \tag{A.11b}$$

and $\phi(x)$ *and* $\psi(x)$ *are defined in* (A.2) *and* (A.3). *In addition, we have*

$$\frac{1}{1 + \|A^{-1/2}\Delta A A^{-1/2}\|_2} \leqslant \frac{x^{\mathrm{H}}Ax}{x^{\mathrm{H}}\widetilde{A}x} \leqslant \frac{1}{1 - \|A^{-1/2}\Delta A A^{-1/2}\|_2}, \tag{A.12}$$

$$|\delta_2^\pm(x)| \leqslant \frac{\max\{|\lambda_1^\pm|^2, |\lambda_n^\pm|^2\}\|\Delta A\|_2 + \max\{|\lambda_1^\pm|, |\lambda_n^\pm|\}\|\Delta B\|_2 + \|\Delta C\|_2}{\min_{x\neq 0}\varsigma_0(x)}. \tag{A.13}$$

Proof. According to how the difference operator Δ is defined at the beginning of Section 6, we have

$$\pm\Delta\rho_\pm(x) = \frac{\Delta\varsigma(x) \mp x^{\mathrm{H}}\Delta B x}{2(x^{\mathrm{H}}Ax)} + \frac{\tilde{\varsigma}(x) \mp x^{\mathrm{H}}\widetilde{B}x}{2}\Delta\left(\frac{1}{x^{\mathrm{H}}Ax}\right) =: \epsilon_1 + \epsilon_2. \tag{A.14}$$

The rest of this proof is to calculate ϵ_1 and ϵ_2. By Lemma A.1,

$$f(t;x) := [\varsigma(x)^2 + 2\psi(x)t + \phi(x)t^2]^{1/2} \tag{A.15}$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

is well defined and differentiable for $0 \leqslant t \leqslant 1$. By the Taylor expansion, there exists $0 \leqslant \xi \leqslant 1$ such that

$$\begin{aligned}\tilde{\varsigma}(x) = f(1;x) &= f(0;x) + f'(0;x) + \frac{1}{2}f''(\xi;x) \\ &= \varsigma(x) + \frac{\psi(x)}{\varsigma(x)} + \frac{\varsigma(x)^2\phi(x) - \psi(x)^2}{2[f(\xi;x)]^3}. \end{aligned} \tag{A.16}$$

This ξ depends on x. Now we are ready to calculate ϵ_1 and ϵ_2. We have

$$\begin{aligned}\epsilon_1 &= \mp\frac{x^{\mathrm{H}}\Delta Bx}{2(x^{\mathrm{H}}Ax)} + \frac{1}{2(x^{\mathrm{H}}Ax)}\left(\frac{\psi(x)}{\varsigma(x)} + \frac{\varsigma(x)^2\phi(x) - \psi(x)^2}{2[f(\xi;x)]^3}\right) \\ &= \mp\frac{x^{\mathrm{H}}\Delta Bx}{2(x^{\mathrm{H}}Ax)} + \frac{(x^{\mathrm{H}}Bx)(x^{\mathrm{H}}\Delta Bx)}{2(x^{\mathrm{H}}Ax)\varsigma(x)} - \frac{x^{\mathrm{H}}\Delta Cx}{\varsigma(x)} - \frac{x^{\mathrm{H}}Cx}{\varsigma(x)}\frac{x^{\mathrm{H}}\Delta Ax}{x^{\mathrm{H}}Ax} \\ &\quad + \frac{\varsigma(x)^2\phi(x) - \psi(x)^2}{4(x^{\mathrm{H}}Ax)[f(\xi;x)]^3} \\ &= -\frac{\pm\varsigma(x) - (x^{\mathrm{H}}Bx)}{2(x^{\mathrm{H}}Ax)}\frac{x^{\mathrm{H}}\Delta Bx}{\varsigma(x)} - \frac{x^{\mathrm{H}}\Delta Cx}{\varsigma(x)} - \frac{x^{\mathrm{H}}Cx}{\varsigma(x)}\frac{x^{\mathrm{H}}\Delta Ax}{x^{\mathrm{H}}Ax} \\ &\quad + \frac{\varsigma(x)^2\phi(x) - \psi(x)^2}{4(x^{\mathrm{H}}Ax)[f(\xi;x)]^3} \\ &= -\frac{\rho_{\pm}(x)(x^{\mathrm{H}}\Delta Bx)}{\varsigma(x)} - \frac{x^{\mathrm{H}}\Delta Cx}{\varsigma(x)} - \frac{x^{\mathrm{H}}Cx}{\varsigma(x)}\frac{x^{\mathrm{H}}\Delta Ax}{x^{\mathrm{H}}Ax} \\ &\quad + \frac{x^{\mathrm{H}}\widetilde{A}x}{x^{\mathrm{H}}Ax}\frac{\varsigma(x)^2\phi(x) - \psi(x)^2}{4(x^{\mathrm{H}}\widetilde{A}x)[f(\xi;x)]^3} \\ &= -\delta_2^{\pm}(x) + \frac{\rho_{\pm}(x)^2(x^{\mathrm{H}}\Delta Ax)}{\varsigma(x)} - \frac{x^{\mathrm{H}}Cx}{\varsigma(x)}\frac{x^{\mathrm{H}}\Delta Ax}{x^{\mathrm{H}}Ax} + \frac{x^{\mathrm{H}}\widetilde{A}x}{x^{\mathrm{H}}Ax}\delta_3(x,\xi), \end{aligned}$$

and

$$\begin{aligned}\epsilon_2 &= -\frac{[\tilde{\varsigma}(x) \mp x^{\mathrm{H}}\widetilde{B}x](x^{\mathrm{H}}\Delta Ax)}{2(x^{\mathrm{H}}\widetilde{A}x)(x^{\mathrm{H}}Ax)} = \frac{\mp\tilde{\rho}_{\pm}(x)(x^{\mathrm{H}}\Delta Ax)}{x^{\mathrm{H}}Ax} \\ &= -[\pm\rho_{\pm}(x) \pm \Delta\rho_{\pm}(x)]\frac{x^{\mathrm{H}}\Delta Ax}{x^{\mathrm{H}}Ax}. \end{aligned}$$

Noticing that

$$\begin{aligned}\frac{x^{\mathrm{H}}Cx}{\varsigma(x)} \pm \rho_{\pm}(x) &= \frac{x^{\mathrm{H}}Cx}{\varsigma(x)} \pm \frac{-x^{\mathrm{H}}Bx \pm \varsigma(x)}{2(x^{\mathrm{H}}Ax)} \\ &= \frac{2(x^{\mathrm{H}}Ax)(x^{\mathrm{H}}Cx) \mp x^{\mathrm{H}}Bx\varsigma(x) + \varsigma(x)^2}{2\varsigma(x)(x^{\mathrm{H}}Ax)} \end{aligned}$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

$$= \frac{(x^{\mathrm{H}}Bx)^2 - \varsigma(x)^2 \mp 2(x^{\mathrm{H}}Bx)\varsigma(x) + 2\varsigma(x)^2}{4\varsigma(x)(x^{\mathrm{H}}Ax)}$$
$$= \frac{[x^{\mathrm{H}}Bx \mp \varsigma(x)]^2}{4\varsigma(x)(x^{\mathrm{H}}Ax)} = \frac{\rho_{\pm}(x)^2(x^{\mathrm{H}}Ax)}{\varsigma(x)},$$

we have

$$\pm\Delta\rho_{\pm}(x) = \epsilon_1 + \epsilon_2 = -\delta_2^{\pm}(x) + \frac{x^{\mathrm{H}}\widetilde{A}x}{x^{\mathrm{H}}Ax}\delta_3(x,\xi) - [\pm\Delta\rho_{\pm}(x)]\frac{x^{\mathrm{H}}\Delta Ax}{x^{\mathrm{H}}Ax},$$

solving which for $\pm\Delta\rho_{\pm}(x)$ leads to $\Delta\rho_{\pm}(x) = \delta^{\pm}(x,\xi)$ as given by (A.10). □

LEMMA A.4. *Suppose that* (A.5) *and* (6.7) *hold. Let* $\delta_{\mathrm{lb}}^{\pm}(x)$, $\delta_{\mathrm{ub}}^{\pm}(x)$, $\tilde{\delta}_{\mathrm{lb}}^{\pm}(x)$, *and* $\tilde{\delta}_{\mathrm{ub}}^{\pm}(x)$ *be functions satisfying*

$$\delta_{\mathrm{lb}}^{\pm}(x) \leqslant \delta^{\pm}(x,\xi) \leqslant \delta_{\mathrm{ub}}^{\pm}(x), \quad \tilde{\delta}_{\mathrm{lb}}^{\pm}(x) \leqslant \tilde{\delta}^{\pm}(x,\xi) \leqslant \tilde{\delta}_{\mathrm{ub}}^{\pm}(x) \tag{A.17}$$

for all nonzero $x \in \mathbb{C}^n$, $\xi \in [0,1]$, *where* $\delta^{\pm}(x,\xi)$ *is defined in Lemma* A.3. *Write*

$$\gamma_{\mathrm{uu}}^{\pm} = \max_{x\neq 0}\{\delta_{\mathrm{ub}}^{\pm}(x), \tilde{\delta}_{\mathrm{ub}}^{\pm}(x)\}, \qquad \gamma_{\mathrm{ll}}^{\pm} = \max_{x\neq 0}\{-\delta_{\mathrm{lb}}^{\pm}(x), -\tilde{\delta}_{\mathrm{lb}}^{\pm}(x)\},$$
$$\gamma_{\mathrm{lu}}^{\pm} = \max_{x\neq 0}\{-\delta_{\mathrm{lb}}^{\pm}(x), \delta_{\mathrm{ub}}^{\pm}(x)\}, \quad \widetilde{\gamma}_{\mathrm{lu}}^{\pm} = \max_{x\neq 0}\{-\tilde{\delta}_{\mathrm{lb}}^{\pm}(x), \tilde{\delta}_{\mathrm{ub}}^{\pm}(x)\}.$$

Then

$$\|\Delta\Lambda_{\pm}\|_2 = \max_{1\leqslant i\leqslant n}|\Delta\lambda_i^{\pm}| \leqslant \min\{\gamma_{\mathrm{uu}}^{\pm}, \gamma_{\mathrm{ll}}^{\pm}, \gamma_{\mathrm{lu}}^{\pm}, \widetilde{\gamma}_{\mathrm{lu}}^{\pm}\}. \tag{A.18}$$

Proof. We only consider the + case below; the − case is similar. In fact simply replacing + with − gives a proof for the − case.

By Lemma A.3,

$$\delta_{\mathrm{lb}}^{+}(x) \leqslant \Delta\rho_{+}(x) = \delta^{+}(x,\xi) \leqslant \delta_{\mathrm{ub}}^{+}(x).$$

Let $\mathcal{S}_i = \mathrm{span}\{u_1^+, \ldots, u_i^+\}$, $\mathcal{T}_i = \mathrm{span}\{u_i^+, \ldots, u_n^+\}$, and similarly define $\widetilde{\mathcal{S}}_i$ and $\widetilde{\mathcal{T}}_i$. By Theorem 5.1, the Courant–Fischer type min–max principles in Theorem 5.2, and Lemma 5.10,

$$\lambda_i^+ = \min_{\dim\mathcal{X}=i}\max_{0\neq x\in\mathcal{X}}\rho_+(x) = \max_{0\neq x\in\mathcal{S}_i}\rho_+(x) = \rho_+(u_i^+),$$
$$\tilde{\lambda}_i^+ = \min_{\dim\mathcal{X}=i}\max_{0\neq x\in\mathcal{X}}\tilde{\rho}_+(x) = \max_{0\neq x\in\widetilde{\mathcal{S}}_i}\tilde{\rho}_+(x) = \widetilde{\rho}_+(\widetilde{u}_i^+),$$
$$\lambda_i^+ = \max_{\mathrm{codim}\,\mathcal{X}=i-1}\min_{0\neq x\in\mathcal{X}}\rho_+(x) = \min_{0\neq x\in\mathcal{T}_i}\rho_+(x) = \rho_+(u_i^+),$$
$$\tilde{\lambda}_i^+ = \max_{\mathrm{codim}\,\mathcal{X}=i-1}\min_{0\neq x\in\mathcal{X}}\tilde{\rho}_+(x) = \min_{0\neq x\in\widetilde{\mathcal{T}}_i}\tilde{\rho}_+(x) = \widetilde{\rho}_+(\widetilde{u}_i^+).$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

Therefore,

$$\begin{aligned}
\tilde{\lambda}_i^+ = \min_{\dim \mathcal{X}=i} \max_{0\neq x\in\mathcal{X}} \tilde{\rho}_+(x) &\leqslant \max_{0\neq x\in\mathcal{S}_i} \tilde{\rho}_+(x) \\
&\leqslant \max_{0\neq x\in\mathcal{S}_i} \left[\rho_+(x) + \delta_{\mathrm{ub}}^+(x)\right] \\
&\leqslant \max_{0\neq x\in\mathcal{S}_i} \rho_+(x) + \max_{0\neq x\in\mathcal{S}_i} \delta_{\mathrm{ub}}^+(x) \\
&= \lambda_i^+ + \max_{0\neq x\in\mathcal{S}_i} \delta_{\mathrm{ub}}^+(x), \\
\tilde{\lambda}_i^+ = \max_{\operatorname{codim} \mathcal{X}=i-1} \min_{0\neq x\in\mathcal{X}} \tilde{\rho}_+(x) &\geqslant \min_{0\neq x\in\mathcal{T}_i} \tilde{\rho}_+(x) \\
&\geqslant \min_{0\neq x\in\mathcal{T}_i} \left[\rho_+(x) + \delta_{\mathrm{lb}}^+(x)\right] \\
&\geqslant \min_{0\neq x\in\mathcal{T}_i} \rho_+(x) + \min_{0\neq x\in\mathcal{T}_i} \delta_{\mathrm{lb}}^+(x) \\
&= \lambda_i^+ + \min_{0\neq x\in\mathcal{T}_i} \delta_{\mathrm{lb}}^+(x).
\end{aligned}$$

They give (A.19a) below, and (A.19b) as well upon switching the roles of $\boldsymbol{Q}$ and $\widetilde{\boldsymbol{Q}}$:

$$\min_{0\neq x\in\mathcal{T}_i} \delta_{\mathrm{lb}}^+(x) \leqslant \tilde{\lambda}_i^+ - \lambda_i^+ \leqslant \max_{0\neq x\in\mathcal{S}_i} \delta_{\mathrm{ub}}^+(x), \tag{A.19a}$$

$$\min_{0\neq x\in\widetilde{\mathcal{T}}_i} \tilde{\delta}_{\mathrm{lb}}^+(x) \leqslant \lambda_i^+ - \tilde{\lambda}_i^+ \leqslant \max_{0\neq x\in\widetilde{\mathcal{S}}_i} \tilde{\delta}_{\mathrm{ub}}^+(x). \tag{A.19b}$$

It follows from (A.19) that

$$\begin{aligned}
|\Delta\lambda_i^+| &\leqslant \max\left\{\max_{0\neq x\in\mathcal{S}_i} \delta_{\mathrm{ub}}^+(x), \max_{0\neq x\in\widetilde{\mathcal{S}}_i} \tilde{\delta}_{\mathrm{ub}}^+(x)\right\} \\
&\leqslant \max_{x\neq 0}\{\delta_{\mathrm{ub}}^+(x), \tilde{\delta}_{\mathrm{ub}}^+(x)\} = \gamma_{\mathrm{uu}}^+, \\
|\Delta\lambda_i^+| &\leqslant \max\left\{-\min_{0\neq x\in\mathcal{T}_i} \delta_{\mathrm{lb}}^+(x), -\min_{0\neq x\in\widetilde{\mathcal{T}}_i} \tilde{\delta}_{\mathrm{lb}}^+(x)\right\} \\
&\leqslant \max_{x\neq 0}\{-\delta_{\mathrm{lb}}^+(x), -\tilde{\delta}_{\mathrm{lb}}^+(x)\} = \gamma_{\mathrm{ll}}^+, \\
|\Delta\lambda_i^+| &\leqslant \max\left\{-\min_{0\neq x\in\mathcal{T}_i} \delta_{\mathrm{lb}}^+(x), \max_{0\neq x\in\mathcal{S}_i} \delta_{\mathrm{ub}}^+(x)\right\} \\
&\leqslant \max_{x\neq 0}\{-\delta_{\mathrm{lb}}^+(x), \delta_{\mathrm{ub}}^+(x)\} = \gamma_{\mathrm{lu}}^+, \\
|\Delta\lambda_i^+| &\leqslant \max\left\{-\min_{0\neq x\in\widetilde{\mathcal{T}}_i} \tilde{\delta}_{\mathrm{lb}}^+(x), \max_{0\neq x\in\widetilde{\mathcal{S}}_i} \tilde{\delta}_{\mathrm{ub}}^+(x)\right\} \\
&\leqslant \max_{x\neq 0}\{-\tilde{\delta}_{\mathrm{lb}}^+(x), \tilde{\delta}_{\mathrm{ub}}^+(x)\} = \tilde{\gamma}_{\mathrm{lu}}^+.
\end{aligned}$$

This completes the proof of (A.18) for the + case. □

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

Proof of Theorem 6.1. We only prove the perturbation results for Λ_+. The case for Λ_- can be turned into one for Λ_+ by considering the pos-type eigenvalues of $\boldsymbol{Q}(-\lambda)$ and $\widetilde{\boldsymbol{Q}}(-\lambda)$.

For any $\alpha > 0$, $x \neq 0$, we have

$$\epsilon_a < \alpha \quad \Rightarrow \quad |x^{\mathrm{H}}\Delta A x| < \alpha x^{\mathrm{H}} A x, \tag{A.20a}$$

$$\epsilon_a < \alpha \frac{\chi_\varsigma^2}{4\|A\|_2\|C\|_2} \quad \Rightarrow \quad |x^{\mathrm{H}}\Delta A x| < \alpha \frac{\varsigma(x)^2}{4|x^{\mathrm{H}}Cx|}, \tag{A.20b}$$

$$\epsilon_c < \alpha \frac{\chi_\varsigma^2}{4\|A\|_2\|C\|_2} \quad \Rightarrow \quad |x^{\mathrm{H}}\Delta C x| < \alpha \frac{\varsigma(x)^2}{4x^{\mathrm{H}}Ax}, \tag{A.20c}$$

$$\epsilon_b < \alpha \frac{\chi_\varsigma^2}{\|B\|_2(\|B\|_2 + 2\sqrt{\|A\|_2\|C\|_2})} \quad \Rightarrow \quad |x^{\mathrm{H}}\Delta B x| < \alpha |x^{\mathrm{H}} B x|, \tag{A.20d}$$

where (A.20a) and (A.20b) hold because

$$\left|\frac{x^{\mathrm{H}}\Delta A x}{x^{\mathrm{H}}Ax}\right| = \left|\frac{x^{\mathrm{H}}A^{1/2}(A^{-1/2}\Delta A A^{-1/2})A^{1/2}x}{x^{\mathrm{H}}A^{1/2}A^{1/2}x}\right| \leqslant \|A^{-1/2}\Delta A A^{-1/2}\|_2 = \epsilon_a,$$

and (A.20d) holds because its left inequality implies that

$$|x^{\mathrm{H}}\Delta B x| < \alpha \frac{\varsigma(x)^2}{|x^{\mathrm{H}}Bx| + \sqrt{4(x^{\mathrm{H}}Ax)|x^{\mathrm{H}}Cx|}} = \alpha\left(|x^{\mathrm{H}}Bx| - \sqrt{4(x^{\mathrm{H}}Ax)|x^{\mathrm{H}}Cx|}\right). \tag{A.21}$$

For item (1), we have $\Delta A = \Delta B = 0$, $\phi(x) = \tilde{\phi}(x) = 0$, $\psi(x) = -2(x^{\mathrm{H}}Ax)(x^{\mathrm{H}}\Delta C x)$, and (6.7). Under assumption (6.12), (A.20c) holds with $\alpha = 1$. Thus $g(1) = \varsigma(x)^2 + 2\psi(x) + \phi(x) > 0$, or equivalently the perturbed quadratic polynomial is still hyperbolic. Note that (A.5) holds for $\phi(x) = 0$. Thus $\delta_3(x,\xi) \leqslant 0$ and $\tilde{\delta}_3(x,\xi) \leqslant 0$. We can take, in (A.17),

$$\delta_{\mathrm{ub}}^+(x) = -\delta_2^+(x) = -\frac{x^{\mathrm{H}}\Delta C x}{\varsigma(x)}, \quad \tilde{\delta}_{\mathrm{ub}}^+(x) = -\tilde{\delta}_2^+(x) = \frac{x^{\mathrm{H}}\Delta C x}{\tilde{\varsigma}(x)} \tag{A.22}$$

to give

$$|\delta_{\mathrm{ub}}^+(x)| \leqslant \frac{\|\Delta C\|_2}{\min_{x\neq 0}\varsigma_0(x)}, \quad |\tilde{\delta}_{\mathrm{ub}}^+(x)| \leqslant \frac{\|\Delta C\|_2}{\min_{x\neq 0}\tilde{\varsigma}_0(x)}.$$

Using (A.18), we have $\|\Delta\Lambda_+\|_2 \leqslant \gamma_{\mathrm{uu}}^+$, and thus (6.13).

For item (2), we have $\Delta B = \Delta C = 0$, $\phi(x) = \tilde{\phi}(x) = 0$, and $\psi(x) = -2(x^{\mathrm{H}}Cx)(x^{\mathrm{H}}\Delta A x)$. Under assumption (6.14), (6.7) holds, and (A.20a) and (A.20b) hold with $\alpha = 1$. Thus $g(1) = \varsigma(x)^2 + 2\psi(x) + \phi(x) > 0$, or equivalently

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

the perturbed quadratic polynomial is still hyperbolic. Note that (A.5) holds for $\phi(x) = 0$. Thus $\delta_3(x, \xi) \leqslant 0$ and $\tilde{\delta}_3(x, \xi) \leqslant 0$. We can take, in (A.17),

$$\delta_{\mathrm{ub}}^{+}(x) = -\frac{x^{\mathrm{H}}Ax}{x^{\mathrm{H}}\widetilde{A}x}\,\delta_2^{+}(x) = -\frac{x^{\mathrm{H}}Ax}{x^{\mathrm{H}}\widetilde{A}x}\,\frac{\rho_+(x)^2(x^{\mathrm{H}}\Delta Ax)}{\varsigma(x)},$$
$$\tilde{\delta}_{\mathrm{ub}}^{+}(x) = -\frac{x^{\mathrm{H}}\widetilde{A}x}{x^{\mathrm{H}}Ax}\,\tilde{\delta}_2^{+}(x) = \frac{x^{\mathrm{H}}\widetilde{A}x}{x^{\mathrm{H}}Ax}\,\frac{\tilde{\rho}_+(x)^2(x^{\mathrm{H}}\Delta Ax)}{\tilde{\varsigma}(x)},$$

along with (A.12), to give

$$|\delta_{\mathrm{ub}}^{+}(x)| \leqslant \frac{1}{1-\epsilon_a}\frac{(\lambda_{\max}^{+})^2\|\Delta A\|_2}{\min_{x\neq 0}\varsigma_0(x)}, \quad |\tilde{\delta}_{\mathrm{ub}}^{+}(x)| \leqslant (1+\epsilon_a)\frac{(\tilde{\lambda}_{\max}^{+})^2\|\Delta A\|_2}{\min_{x\neq 0}\tilde{\varsigma}_0(x)}.$$

Using (A.18), we have $\|\Delta\Lambda_+\|_2 \leqslant \gamma_{\mathrm{uu}}^{+}$, and thus (6.15).

For item (3), we have $\Delta A = \Delta C = 0$, $\phi(x) = \tilde{\phi}(x) = (x^{\mathrm{H}}Bx)(x^{\mathrm{H}}\Delta Bx)$, $\psi(x) = (x^{\mathrm{H}}\Delta Bx)^2$, and (6.7). Under assumption (6.16), (A.20d) and (A.21) hold with $\alpha = 1$. By (A.21), we see that

$$\sqrt{4(x^{\mathrm{H}}Ax)|x^{\mathrm{H}}Cx|} < |x^{\mathrm{H}}Bx| - |x^{\mathrm{H}}\Delta Bx| \leqslant |x^{\mathrm{H}}Bx + x^{\mathrm{H}}\Delta Bx|.$$

Thus

$$\begin{aligned} g(1) &= \varsigma(x)^2 + 2\psi(x) + \phi(x) \\ &= (x^{\mathrm{H}}\Delta Bx)^2 + 2(x^{\mathrm{H}}\Delta Bx)(x^{\mathrm{H}}Bx) + (x^{\mathrm{H}}Bx)^2 - 4(x^{\mathrm{H}}Ax)(x^{\mathrm{H}}Cx) \\ &\geqslant \left[x^{\mathrm{H}}\Delta Bx + x^{\mathrm{H}}Bx - \sqrt{4(x^{\mathrm{H}}Ax)|x^{\mathrm{H}}Cx|}\right] \\ &\quad \times \left[x^{\mathrm{H}}\Delta Bx + x^{\mathrm{H}}Bx + \sqrt{4(x^{\mathrm{H}}Ax)|x^{\mathrm{H}}Cx|}\right] \\ &> 0, \end{aligned}$$

or equivalently the perturbed quadratic polynomial is still hyperbolic. By (A.20d), we have $|\psi(x)| = |x^{\mathrm{H}}Bx| > |x^{\mathrm{H}}\Delta Bx| = \phi(x)$. Thus (A.5) holds. Notice that

$$\begin{aligned} \varsigma(x)^2\phi(x) - \psi(x)^2 &= \varsigma(x)^2(x^{\mathrm{H}}\Delta Bx)^2 - [(x^{\mathrm{H}}Bx)(x^{\mathrm{H}}\Delta Bx)]^2 \\ &= -4(x^{\mathrm{H}}Ax)(x^{\mathrm{H}}Cx)(x^{\mathrm{H}}\Delta Bx)^2 \end{aligned}$$

to get

$$\delta_3(x, \xi) = -\frac{(x^{\mathrm{H}}Cx)(x^{\mathrm{H}}\Delta Bx)^2}{[f(\xi; x)]^3},$$

where $f(\xi; x) = [\varsigma(x)^2 + 2\psi(x)\xi + \phi(x)\xi^2]^{1/2}$. Since

$$\min_{0\leqslant\xi\leqslant 1} f(\xi; x) = \min\{f(0), f(1)\} = \min\{\varsigma(x), \tilde{\varsigma}(x)\}, \tag{A.23}$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

we can take, in (A.17),

$$\delta^+_{\mathrm{ub}}(x) = -\delta^+_2(x) + \frac{|x^{\mathrm{H}}Cx||x^{\mathrm{H}}\Delta Bx|^2}{\min\{\varsigma(x), \tilde{\varsigma}(x)\}^3} = -\frac{\rho_+(x)(x^{\mathrm{H}}\Delta Bx)}{\varsigma(x)} + \frac{|x^{\mathrm{H}}Cx||x^{\mathrm{H}}\Delta Bx|^2}{\min\{\varsigma(x), \tilde{\varsigma}(x)\}^3},$$

$$\tilde{\delta}^+_{\mathrm{ub}}(x) = -\tilde{\delta}^+_2(x) + \frac{|x^{\mathrm{H}}\widetilde{C}x||x^{\mathrm{H}}\Delta Bx|^2}{\min\{\varsigma(x), \tilde{\varsigma}(x)\}^3} = \frac{\tilde{\rho}_+(x)(x^{\mathrm{H}}\Delta Bx)}{\tilde{\varsigma}(x)} + \frac{|x^{\mathrm{H}}\widetilde{C}x||x^{\mathrm{H}}\Delta Bx|^2}{\min\{\varsigma(x), \tilde{\varsigma}(x)\}^3}$$

to give

$$|\delta^+_{\mathrm{ub}}(x)| \leqslant \frac{\lambda^+_{\max}}{\min_{x\neq 0}\varsigma_0(x)}\|\Delta B\|_2 + \frac{\|C\|_2}{\chi^3_\varsigma}\|\Delta B\|_2^2,$$

$$|\tilde{\delta}^+_{\mathrm{ub}}(x)| \leqslant \frac{\tilde{\lambda}^+_{\max}}{\min_{x\neq 0}\tilde{\varsigma}_0(x)}\|\Delta B\|_2 + \frac{\|\widetilde{C}\|_2}{\chi^3_\varsigma}\|\Delta B\|_2^2.$$

(For the quadratic function $h(t) = a(t-c)^2 + b$ with $a > 0$, if $|c| \geqslant 1$, that is, c, the minimal point of $h(t)$ for $t \in \mathbb{R}$, is not in the interval $(0, 1)$, then the minimal point of $h(t)$ on $[0, 1]$ must be either 0 or 1. For the case here, $c = \psi(x)/\phi(x)$.) Using (A.18), we have $\|\Delta\Lambda_+\|_2 \leqslant \gamma^+_{\mathrm{uu}}$, and thus (6.17).

For item (4), we have $\Delta A = \Delta C = 0$. Consider the shifted $\boldsymbol{Q}_{\lambda_0}(\lambda)$ as defined in (6.2). By item (2) of Lemma A.2, $\boldsymbol{Q}_{\lambda_0}(\lambda)$ and $\widetilde{\boldsymbol{Q}}_{\lambda_0}(\lambda)$ are overdamped for

$$\lambda_0 \in (-\infty, \min\{\lambda_1^-, \tilde{\lambda}_1^-\}] \cup [\max\{\lambda_n^+, \tilde{\lambda}_n^+\}, +\infty).$$

In particular, $B_{\lambda_0} \succ 0$, $C_{\lambda_0} \succeq 0$, $\widetilde{B}_{\lambda_0} \succ 0$, $\widetilde{C}_{\lambda_0} \succeq 0$. Note that $\varsigma_{\lambda_0}(x) \equiv \varsigma(x)$, $\widetilde{\varsigma}_{\lambda_0}(x) \equiv \widetilde{\varsigma}(x)$. Under assumption (6.18) (we will use the same symbols as those for $\boldsymbol{Q}$ but with the subscript λ_0 to represent the corresponding quantities for $\boldsymbol{Q}_{\lambda_0}$), $|\psi_{\lambda_0}(x)| > \phi_{\lambda_0}(x)$. Thus (A.5) for $\boldsymbol{Q}_{\lambda_0}(\lambda)$ and $\widetilde{\boldsymbol{Q}}_{\lambda_0}(\lambda)$ holds. Just as in item (3) (note that $\Delta B_{\lambda_0} = \Delta B$ since $\Delta A = 0$),

$$\varsigma_{\lambda_0}(x)^2\phi_{\lambda_0}(x) - \psi_{\lambda_0}(x)^2 = -4(x^{\mathrm{H}}Ax)(x^{\mathrm{H}}C_{\lambda_0}x)(x^{\mathrm{H}}\Delta Bx)^2 < 0,$$

which yields $\delta_{3;\lambda_0}(x, \xi) \leqslant 0$, and thus we can take, in (A.17),

$$\delta^+_{\mathrm{ub};\lambda_0}(x) = -\delta^+_{2;\lambda_0}(x) = -\frac{\rho_{+;\lambda_0}(x)(x^{\mathrm{H}}\Delta Bx)}{\varsigma(x)},$$

$$\tilde{\delta}^+_{\mathrm{ub};\lambda_0}(x) = -\tilde{\delta}^+_{2;\lambda_0}(x) = -\frac{\tilde{\rho}_{+;\lambda_0}(x)(x^{\mathrm{H}}\Delta Bx)}{\tilde{\varsigma}(x)}$$

to give

$$|\delta^+_{\mathrm{ub};\lambda_0}(x)| \leqslant \frac{\lambda^+_{\max;\lambda_0}}{\min_{x\neq 0}\varsigma_0(x)}\|\Delta B\|_2, \quad |\tilde{\delta}^+_{\mathrm{ub};\lambda_0}(x)| \leqslant \frac{\tilde{\lambda}^+_{\max;\lambda_0}}{\min_{x\neq 0}\tilde{\varsigma}_0(x)}\|\Delta B\|_2.$$

Using (A.18), we have $\|\Delta\Lambda_{+;\lambda_0}\|_2 \leqslant \gamma^+_{\mathrm{uu};\lambda_0}$, and thus (6.19).

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

For item (5), under assumption (6.20), we have $\epsilon_a < \gamma < 1$, and (A.20) holds with $\alpha = \gamma$. Then (6.7) holds, and

$$\begin{aligned}
|\psi(x)| &\leqslant |x^{\mathrm{H}}Bx||x^{\mathrm{H}}\Delta Bx| + 2(x^{\mathrm{H}}Ax)|x^{\mathrm{H}}\Delta Cx| + 2|x^{\mathrm{H}}Cx||x^{\mathrm{H}}\Delta Ax| \\
&< |x^{\mathrm{H}}Bx|^2\gamma + \frac{\varsigma(x)^2}{2}\gamma + \frac{\varsigma(x)^2}{2}\gamma \\
&= [|x^{\mathrm{H}}Bx|^2 + \varsigma(x)^2]\gamma, \\
|\phi(x)| &\leqslant |x^{\mathrm{H}}\Delta Bx|^2 + 4|x^{\mathrm{H}}\Delta Ax||x^{\mathrm{H}}\Delta Cx| \\
&< |x^{\mathrm{H}}Bx|^2\gamma^2 + |x^{\mathrm{H}}\Delta Ax|\frac{\varsigma(x)^2\gamma}{x^{\mathrm{H}}Ax} \\
&< |x^{\mathrm{H}}Bx|^2\gamma^2 + \varsigma(x)^2\gamma^2 \\
&= [|x^{\mathrm{H}}Bx|^2 + \varsigma(x)^2]\gamma^2,
\end{aligned}$$

which gives

$$\begin{aligned}
g(1) &= \varsigma(x)^2 + 2\psi(x) + \phi(x) \\
&> \varsigma(x)^2(1 - 2\gamma - \gamma^2) - |x^{\mathrm{H}}Bx|^2(2\gamma + \gamma^2) \\
&\geqslant (x^{\mathrm{H}}x)^2\left[\chi_\varsigma^2(1 - 2\gamma - \gamma^2) - \|B\|_2^2(2\gamma + \gamma^2)\right] \\
&= (x^{\mathrm{H}}x)^2\left[\chi_\varsigma^2 - (\|B\|_2^2 + \chi_\varsigma^2)(2\gamma + \gamma^2)\right] \\
&= 0;
\end{aligned}$$

that is, the perturbed quadratic polynomial is still hyperbolic. By the same reasoning we had for items (1)–(3), (A.5) holds and, at the same time, we have (A.23). Note that

$$\begin{aligned}
\varsigma(x)^2\phi(x) - \psi(x)^2 = &-4\left[(x^{\mathrm{H}}Ax)(x^{\mathrm{H}}\Delta Cx) - (x^{\mathrm{H}}Cx)(x^{\mathrm{H}}\Delta Ax)\right]^2 \\
&- 4\left[(x^{\mathrm{H}}Ax)(x^{\mathrm{H}}\Delta Bx) - (x^{\mathrm{H}}Bx)(x^{\mathrm{H}}\Delta Ax)\right] \\
&\times \left[(x^{\mathrm{H}}Cx)(x^{\mathrm{H}}\Delta Bx) - (x^{\mathrm{H}}Bx)(x^{\mathrm{H}}\Delta Cx)\right],
\end{aligned}$$

and similarly

$$\begin{aligned}
\tilde{\varsigma}(x)^2\tilde{\phi}(x) - \tilde{\psi}(x)^2 = &-4\left[-(x^{\mathrm{H}}\widetilde{A}x)(x^{\mathrm{H}}\Delta Cx) + (x^{\mathrm{H}}\widetilde{C}x)(x^{\mathrm{H}}\Delta Ax)\right]^2 \\
&- 4\left[-(x^{\mathrm{H}}\widetilde{A}x)(x^{\mathrm{H}}\Delta Bx) + (x^{\mathrm{H}}\widetilde{B}x)(x^{\mathrm{H}}\Delta Ax)\right] \\
&\times \left[-(x^{\mathrm{H}}\widetilde{C}x)(x^{\mathrm{H}}\Delta Bx) + (x^{\mathrm{H}}\widetilde{B}x)(x^{\mathrm{H}}\Delta Cx)\right] \\
= &\ \varsigma(x)^2\phi(x) - \psi(x)^2.
\end{aligned}$$

Now take

$$\delta_{\mathrm{ub}}^+(x) = -\frac{x^{\mathrm{H}}Ax}{x^{\mathrm{H}}\widetilde{A}x}\delta_2^+(x) + \frac{|\varsigma(x)^2\phi(x) - \psi(x)^2|}{(x^{\mathrm{H}}\widetilde{A}x)\min\{\varsigma(x), \tilde{\varsigma}(x)\}^3},$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

$$\tilde{\delta}^{+}_{\mathrm{ub}}(x) = -\frac{x^{\mathrm{H}}\widetilde{A}x}{x^{\mathrm{H}}Ax}\tilde{\delta}^{+}_{2}(x) + \frac{|\varsigma(x)^2\phi(x) - \psi(x)^2|}{(x^{\mathrm{H}}Ax)\min\{\varsigma(x), \tilde{\varsigma}(x)\}^3}$$

in (A.17). Noting that $|x^{\mathrm{H}}\Delta Ax/x^{\mathrm{H}}Ax| \leqslant \epsilon_a$, we have

$$\begin{aligned}|\varsigma(x)^2\phi(x) - \psi(x)^2| \leqslant\ & 4(x^{\mathrm{H}}Ax)^2\|C\|_2^2[\epsilon_c + \epsilon_a]^2 \\ & + 4(x^{\mathrm{H}}Ax)\|B\|_2^2\|C\|_2[\epsilon_b + \epsilon_a][\epsilon_b + \epsilon_c].\end{aligned}$$

Using (A.18), we have $\|\Delta\Lambda_+\|_2 \leqslant \gamma^{+}_{\mathrm{uu}}$, and thus (6.22). □

The rest of this appendix is devoted to the proof of Theorem 6.2.

LEMMA A.5. *Suppose that $\Delta A = \Delta B = 0$ and that (6.12) holds. Let $\epsilon_1 \leqslant \epsilon_2 \leqslant \cdots \leqslant \epsilon_n$ be the eigenvalues of ΔC, and let γ and $\widetilde{\gamma}$ be given by (6.26).*

(1) *Given $X \in \mathbb{C}^{n\times k}$ with $\mathrm{rank}(X) = k$, denote the eigenvalues of $X^{\mathrm{H}}\boldsymbol{Q}(\lambda)X$ by*

$$\lambda^{-}_{1,X} \leqslant \cdots \leqslant \lambda^{-}_{k,X} \leqslant \lambda^{+}_{1,X} \leqslant \cdots \leqslant \lambda^{+}_{k,X},$$

and the eigenvalues of $X^{\mathrm{H}}\widetilde{\boldsymbol{Q}}(\lambda)X$ by $\tilde{\lambda}^{\pm}_{j,X}$ arranged in the same way. Then

$$-\sum_{i=1}^{k}\frac{\max\{0,-\epsilon_1\}+\epsilon_{n-1+i}}{\widetilde{\gamma}} \leqslant \sum_{i=1}^{k}\Delta\lambda^{+}_{i,X} \leqslant -\sum_{i=1}^{k}\frac{\min\{0,-\epsilon_n\}+\epsilon_i}{\gamma}, \tag{A.24a}$$

$$\sum_{i=1}^{k}\frac{\min\{0,-\epsilon_n\}+\epsilon_i}{\gamma} \leqslant \sum_{i=1}^{k}\Delta\lambda^{-}_{i,X} \leqslant \sum_{i=1}^{k}\frac{\max\{0,-\epsilon_1\}+\epsilon_{n-1+i}}{\widetilde{\gamma}}. \tag{A.24b}$$

(2) *For any $1 \leqslant i_1 < \cdots < i_k \leqslant n$,*

$$-\sum_{i=1}^{k}\frac{\max\{0,-\epsilon_1\}+\epsilon_{n+1-i}}{\widetilde{\gamma}} \leqslant \sum_{i=1}^{k}\Delta\lambda^{+}_{i_k} \leqslant -\sum_{i=1}^{k}\frac{\min\{0,-\epsilon_n\}+\epsilon_i}{\gamma}, \tag{A.25a}$$

$$\sum_{i=1}^{k}\frac{\min\{0,-\epsilon_n\}+\epsilon_i}{\gamma} \leqslant \sum_{i=1}^{k}\Delta\lambda^{-}_{i_k} \leqslant \sum_{i=1}^{k}\frac{\max\{0,-\epsilon_1\}+\epsilon_{n+1-i}}{\widetilde{\gamma}}. \tag{A.25b}$$

Proof. Assumption (6.12) guarantees that $\widetilde{\boldsymbol{Q}}(\lambda)$ is still hyperbolic. Without loss of generality, we may assume that X has orthonormal columns; otherwise, we consider $V^{\mathrm{H}}\boldsymbol{Q}(\lambda)V$ instead, where V is from a QR decomposition $X = VR$ of X, $V^{\mathrm{H}}V = I_k$, and $R \in \mathbb{C}^{k\times k}$. Evidently $X^{\mathrm{H}}\boldsymbol{Q}(\lambda)X$ and $V^{\mathrm{H}}\boldsymbol{Q}(\lambda)V$ have the same eigenvalues.

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

Recall the linearization (4.1) for $\boldsymbol{Q}(\lambda)$. We linearize

$$\boldsymbol{Q}_X(\lambda) := X^{\mathrm{H}}\boldsymbol{Q}(\lambda)X \equiv A_X\lambda^2 + B_X\lambda + C_X$$

in the same way to get

$$\mathscr{A}_X - \lambda\mathscr{B}_X \equiv \begin{bmatrix} -C_X & 0 \\ 0 & A_X \end{bmatrix} - \lambda \begin{bmatrix} B_X & A_X \\ A_X & 0 \end{bmatrix} = \mathscr{L}_{\boldsymbol{Q}_X}(\lambda).$$

Next we apply Theorem 4.2 to $\boldsymbol{Q}_X(\lambda)$ to obtain the associated eigendecomposition, and denote the corresponding quantities by the same symbols as those for $\boldsymbol{Q}(\lambda)$ but with the subscript X to indicate them being for $\boldsymbol{Q}_X(\lambda)$. In particular, we will have

$$U_X = [u^+_{1,X}, \ldots, u^+_{k,X}], \quad \Lambda_{+,X} = \operatorname{diag}(\lambda^+_{1,X}, \lambda^+_{2,X}, \ldots, \lambda^+_{k,X}),$$

where $u^+_{i,X}$ are the eigenvectors of $\boldsymbol{Q}_X(\lambda)$, $\varsigma_X(u^+_{i,X}) = 1$, and

$$S_X = \begin{bmatrix} U_X \\ U_X\Lambda_{+,X} \end{bmatrix}, \quad S_X^{\mathrm{H}}\mathscr{B}_X S_X = I_k.$$

Also $S_X^{\mathrm{H}}\widetilde{\mathscr{B}}_X S_X = I_k$ since $\widetilde{\mathscr{B}}_X = \mathscr{B}_X$. Note that $U_X \in \mathbb{C}^{k\times k}$ is nonsingular. By Theorems 4.1 and [**38**, Corollary 2.1],

$$\inf_{Z^{\mathrm{H}}\mathscr{B}_X Z = I_k} \operatorname{trace}(Z^{\mathrm{H}}\mathscr{A}_X Z) = \sum_{i=1}^{k} \lambda^+_{i,X} = \operatorname{trace}(S_X^{\mathrm{H}}\mathscr{A}_X S_X).$$

Let $\epsilon_{1,X} \leqslant \cdots \leqslant \epsilon_{k,X}$ be the eigenvalues of $\Delta C_X = X^{\mathrm{H}}\Delta C X$. Since X has orthonormal columns, we have $\epsilon_i \leqslant \epsilon_{i,X} \leqslant \epsilon_{n-k+i}$ by the Cauchy interlacing theorem, and thus

$$\sum_{i=1}^{k}\epsilon_i \leqslant \sum_{i=1}^{k}\epsilon_{i,X} \leqslant \sum_{i=1}^{k}\epsilon_{n+1-i}.$$

For the sake of presentation, we will drop the superscript $+$ in $u^+_{i,X}$ in the rest of this proof. We have

$$\begin{aligned}
\sum_{i=1}^{k}\tilde{\lambda}^+_{i,X} &= \inf_{Z^{\mathrm{H}}\widetilde{\mathscr{B}}_X Z = I_k} \operatorname{trace}(Z^{\mathrm{H}}\widetilde{\mathscr{A}_X} Z) \\
&\leqslant \operatorname{trace}(S_X^{\mathrm{H}}\widetilde{\mathscr{A}_X} S_X) \quad (\text{since } S_X^{\mathrm{H}}\widetilde{\mathscr{B}}_X S_X = I_k) \\
&= \operatorname{trace}(S_X^{\mathrm{H}}\mathscr{A}_X S_X) + \operatorname{trace}(S_X^{\mathrm{H}}\Delta\mathscr{A}_X S_X) \\
&= \sum_{i=1}^{k}\lambda^+_{i,X} - \operatorname{trace}(U_X^{\mathrm{H}}\Delta C_X U_X). \qquad \text{(A.26)}
\end{aligned}$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

Let $\mu = \min\{0, -\epsilon_n\} \leqslant 0$. For any scalar $\tau_0 \in (0,1)$, set $\tau^2 = \tau_0^2\gamma = \tau_0^2(\lambda_1^+ - \lambda_n^-)\lambda_{\min}(A)$, and

$$E_X = -\mu U_X^{\mathrm{H}} U_X, \qquad D_X = U_X^{\mathrm{H}}(U_X^{-\mathrm{H}}U_X^{-1} - \tau^2 I)U_X,$$
$$\mathscr{C}_X = \begin{bmatrix} \tau^{-2}(\Delta C_X + \mu I) & 0 \\ 0 & E_X \end{bmatrix} \in \mathbb{C}^{2k\times 2k}, \quad \mathscr{D}_X = \begin{bmatrix} I & 0 \\ 0 & D_X \end{bmatrix} \in \mathbb{C}^{2k\times 2k}.$$

Note that, by (4.15a), (4.15e), and (4.16),

$$U_X^{\mathrm{H}} A_X U_X \preceq (\lambda_{1,X}^+ - \lambda_{k,X}^-)^{-1} I \preceq (\lambda_1^+ - \lambda_n^-)^{-1} I,$$

which yields

$$U_X^{-\mathrm{H}}U_X^{-1} \succeq (\lambda_1^+ - \lambda_n^-)A_X \succeq (\lambda_1^+ - \lambda_n^-)\lambda_{\min}(A_X) I \succeq (\lambda_1^+ - \lambda_n^-)\lambda_{\min}(A) I = \gamma I \succ \tau^2 I.$$

Thus, $D_X \succ 0$, and so $\mathscr{D}_X \succ 0$. Hence the matrix pencil $\mathscr{C}_X - \lambda\mathscr{D}_X$ has $2k$ finite eigenvalues ν_i $(i = 1, \ldots, 2k)$. By the choice of μ, $\Delta C_X + \mu I \preceq 0$ and $E_X \succeq 0$. Therefore these ν_i can be ordered as

$$\nu_1 \leqslant \cdots \leqslant \nu_k \leqslant 0 \leqslant \nu_{k+1} \leqslant \cdots \leqslant \nu_{2k},$$

where ν_i for $i = 1, \ldots, k$ are the eigenvalues of $\tau^{-2}(\Delta C_X + \mu I)$ and ν_i for $i = k+1, \ldots, 2k$ are the generalized eigenvalues of $E_X - \lambda D_X$. By the Courant–Fischer min–max principle, we have, for $i = 1, \ldots, k$,

$$\begin{aligned} \nu_i &= \min_{\dim \mathcal{X} = i} \max_{0\neq x\in\mathcal{X}} \frac{x^{\mathrm{H}}(\Delta C_X + \mu I)x}{\tau^2 x^{\mathrm{H}}x} \\ &= \frac{1}{\tau^2}\left[\mu + \min_{\dim \mathcal{X} = i} \max_{0\neq x\in\mathcal{X}} \frac{x^{\mathrm{H}}\Delta C_X x}{x^{\mathrm{H}}x}\right] \\ &= \frac{1}{\tau^2}[\mu + \epsilon_{i,X}] \geqslant \frac{1}{\tau^2}[\mu + \epsilon_i] = \frac{1}{\tau_0^2\gamma}[\mu + \epsilon_i]. \end{aligned}$$

By the arbitrary choice of $\tau_0 \in (0,1)$, $\nu_i \geqslant (\mu + \epsilon_i)/\gamma$. For the matrix $T_X := \begin{bmatrix} \tau U_X \\ I \end{bmatrix}$, we have

$$T_X^{\mathrm{H}}\mathscr{D}_X T_X = \tau^2 U_X^{\mathrm{H}}U_X + D_X = I,$$
$$T_X^{\mathrm{H}}\mathscr{C}_X T_X = \tau^2\tau^{-2}U_X^{\mathrm{H}}(\Delta C_X + \mu I)U_X + E_X = U_X^{\mathrm{H}}\Delta C_X U_X.$$

Therefore

$$\mathrm{trace}(U_X^{\mathrm{H}}\Delta C_X U_X) = \mathrm{trace}(T_X^{\mathrm{H}}\mathscr{C}_X T_X) \geqslant \min_{Z^{\mathrm{H}}\mathscr{D}_X Z = I} \mathrm{trace}(Z^{\mathrm{H}}\mathscr{C}_X Z) = \sum_{i=1}^{k} \nu_i.$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

Thus, (A.26) becomes

$$\sum_{i=1}^{k} \Delta\lambda_{i,X}^{+} \leqslant -\sum_{i=1}^{k} \nu_i \leqslant -\sum_{i=1}^{k} \frac{\mu+\epsilon_i}{\gamma} = -\sum_{i=1}^{k} \frac{\min\{0,-\epsilon_n\}+\epsilon_i}{\gamma}. \tag{A.27}$$

Think of $\boldsymbol{Q}$ as obtained from perturbing $\widetilde{\boldsymbol{Q}}$, and apply (A.27) to get

$$-\sum_{i=1}^{k} \Delta\lambda_{i,X}^{+} \leqslant -\sum_{i=1}^{k} \frac{\min\{0,-(-\epsilon_1)\}+(-\epsilon_{n-1+i})}{\widetilde{\gamma}}, \tag{A.28}$$

which, combined with (A.27), leads to (A.24a). Apply (A.24a) to $\boldsymbol{Q}(-\lambda)$ and $\widetilde{\boldsymbol{Q}}(-\lambda)$ to get (A.24b).

Now we prove (A.25). With all sup being taken over $\mathcal{X}_1 \subset \cdots \subset \mathcal{X}_k$ and $\operatorname{codim} \mathcal{X}_j = i_j - 1$, and all inf over $x_j \in \mathcal{X}_j$, $X = [x_1, \ldots, x_k]$, and $\operatorname{rank}(X) = k$, we have, by Theorem 5.3,

$$\begin{aligned}
\sum_{j=1}^{k} \tilde{\lambda}_{i_k}^{+} &= \sup\inf \sum_{j=1}^{k} \tilde{\lambda}_{k,X}^{+} \\
&\leqslant \sup\inf \left[\sum_{j=1}^{k} \lambda_{k,X}^{+} - \sum_{i=1}^{k} \frac{\min\{0,-\epsilon_n\}+\epsilon_i}{\gamma} \right] \quad \text{(by (A.27))} \\
&= \sup\inf \sum_{j=1}^{k} \lambda_{k,X}^{+} - \sum_{i=1}^{k} \frac{\min\{0,-\epsilon_n\}+\epsilon_i}{\gamma} \\
&\leqslant \sum_{j=1}^{k} \lambda_{i_k}^{+} - \sum_{i=1}^{k} \frac{\min\{0,-\epsilon_n\}+\epsilon_i}{\gamma}.
\end{aligned} \tag{A.29}$$

Similarly,

$$\sum_{j=1}^{k} \lambda_{i_k}^{+} \leqslant \sum_{j=1}^{k} \tilde{\lambda}_{i_k}^{+} - \sum_{i=1}^{k} \frac{\min\{0,-(-\epsilon_1)\}+(-\epsilon_{n-1+i})}{\widetilde{\gamma}}. \tag{A.30}$$

The inequalities in (A.25a) are consequences of (A.29) and (A.30). Apply (A.25a) to $\boldsymbol{Q}(-\lambda)$ and $\widetilde{\boldsymbol{Q}}(-\lambda)$ to get (A.25b). □

LEMMA A.6. *Suppose that $\Delta A = \Delta B = 0$ and that (6.12) holds. We have, for $1 \leqslant j \leqslant n$,*

$$\tilde{\lambda}_j^{+} \leqslant \lambda_j^{+} \quad \text{and} \quad \tilde{\lambda}_j^{-} \geqslant \lambda_j^{-} \quad \text{if } \Delta C \succeq 0, \tag{A.31a}$$

$$\tilde{\lambda}_j^{+} \geqslant \lambda_j^{+} \quad \text{and} \quad \tilde{\lambda}_j^{-} \leqslant \lambda_j^{-} \quad \text{if } \Delta C \preceq 0. \tag{A.31b}$$

Consequently $\widetilde{\gamma} \leqslant \gamma$ if $\Delta C \succeq 0$, and $\widetilde{\gamma} \geqslant \gamma$ if $\Delta C \preceq 0$.

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

Proof. Assumption (6.12) guarantees that $\widetilde{\boldsymbol{Q}}(\lambda)$ is still hyperbolic. By (5.2), we see that

$$\begin{aligned} \tilde{\rho}_+(x) \leqslant \rho_+(x) \quad &\text{and} \quad \tilde{\rho}_-(x) \geqslant \tilde{\rho}_-(x) \quad \text{if } \Delta C \succeq 0, \\ \tilde{\rho}_+(x) \geqslant \rho_+(x) \quad &\text{and} \quad \tilde{\rho}_-(x) \leqslant \tilde{\rho}_-(x) \quad \text{if } \Delta C \preceq 0. \end{aligned}$$

Now use Theorem 5.2 to get (A.31). □

Proof of Theorem 6.2. Assumption (6.12) guarantees that $\widetilde{\boldsymbol{Q}}(\lambda)$ is still hyperbolic.

As in Lemma A.5, let $\epsilon_1 \leqslant \epsilon_2 \leqslant \cdots \leqslant \epsilon_n$ be the eigenvalues of ΔC.

Consider first the case when $\Delta C \succeq 0$. Then $0 \leqslant \epsilon_1$. Also $\Delta\lambda_i^+ \leqslant 0$ for all i by Lemma A.6. Therefore the leftmost inequality in (A.25a) gives

$$\sum_{i=1}^{k} |\Delta\lambda_{i_k}^+| \leqslant \sum_{i=1}^{k} \frac{\epsilon_{n+1-i}}{\widetilde{\gamma}}$$

for any $1 \leqslant i_1 < \cdots < i_k \leqslant n$. As a result of [**58**, Theorem II.3.6 and Theorem II.3.17], we have

$$\|\Delta\Lambda_+\|_{\mathrm{ui}} \leqslant \frac{\|\Delta C\|_{\mathrm{ui}}}{\widetilde{\gamma}}. \tag{A.32}$$

Similarly, use the rightmost inequality in (A.25b) to get

$$\|\Delta\Lambda_-\|_{\mathrm{ui}} \leqslant \frac{\|\Delta C\|_{\mathrm{ui}}}{\widetilde{\gamma}}. \tag{A.33}$$

Now we turn to the case when $\Delta C \preceq 0$. Then $\epsilon_n \leqslant 0$. Also $\Delta\lambda_i^+ \geqslant 0$ for all i by Lemma A.6. Therefore the rightmost inequality in (A.25a) gives

$$\sum_{i=1}^{k} |\Delta\lambda_{i_k}^+| \leqslant \sum_{i=1}^{k} \frac{|\epsilon_i|}{\gamma}$$

for any $1 \leqslant i_1 < \cdots < i_k \leqslant n$. Again as a result of [**58**, Theorem II.3.6 and Theorem II.3.17], we have

$$\|\Delta\Lambda_+\|_{\mathrm{ui}} \leqslant \frac{\|\Delta C\|_{\mathrm{ui}}}{\gamma}. \tag{A.34}$$

Similarly, use the leftmost inequality in (A.25b) to get

$$\|\Delta\Lambda_-\|_{\mathrm{ui}} \leqslant \frac{\|\Delta C\|_{\mathrm{ui}}}{\gamma}. \tag{A.35}$$

The inequalities (A.32)–(A.33) together give (6.27) for the case when ΔC is semidefinite.

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

For the general case when ΔC is indefinite, we can decompose into $\Delta C = \Delta C_+ - \Delta C_-$, where $\Delta C_\pm \succeq 0$ and

$$\mathrm{eig}(\Delta C_+) = \{\max\{0, \epsilon_i\},\ 1 \leqslant i \leqslant n\}, \quad \mathrm{eig}(\Delta C_-) = \{\max\{0, -\epsilon_i\},\ 1 \leqslant i \leqslant n\}.$$

In particular, $\|\Delta C_\pm\|_{\mathrm{ui}} \leqslant \|\Delta C\|_{\mathrm{ui}}$. Let $\widehat{C} = C - \Delta C_-$ and $\widehat{\boldsymbol{Q}}(\lambda) = \lambda^2 A + \lambda B + \widehat{C}$. We claim that $\widehat{\boldsymbol{Q}}(\lambda)$ is hyperbolic. This is because $\widetilde{C} = C + \Delta C_+ - \Delta C_- \succeq C - \Delta C_- = \widehat{C}$, and thus, for any $x \neq 0$,

$$0 < (x^{\mathrm{H}} B x)^2 - 4(x^{\mathrm{H}} A x)(x^{\mathrm{H}} \widetilde{C} x) \leqslant (x^{\mathrm{H}} B x)^2 - 4(x^{\mathrm{H}} A x)(x^{\mathrm{H}} \widehat{C} x),$$

where the first inequality holds because $\widetilde{\boldsymbol{Q}}(\lambda)$ is hyperbolic. Apply what we just proved to $\boldsymbol{Q}$ and $\widehat{\boldsymbol{Q}}$ to get

$$\|\widehat{\Lambda}_\pm - \Lambda_\pm\|_{\mathrm{ui}} \leqslant \frac{\|\Delta C_-\|_{\mathrm{ui}}}{\gamma} \leqslant \frac{\|\Delta C\|_{\mathrm{ui}}}{\gamma}, \tag{A.36}$$

where $\widehat{\Lambda}_\pm$ are similarly defined for $\widehat{\boldsymbol{Q}}$ to $\Lambda_\pm$ for $\boldsymbol{Q}$. Notice that $\widetilde{C} = \widehat{C} + \Delta C_+$, and apply what we just proved to $\boldsymbol{Q}$ and $\widehat{\boldsymbol{Q}}$ to get

$$\|\widetilde{\Lambda}_\pm - \widehat{\Lambda}_\pm\|_{\mathrm{ui}} \leqslant \frac{\|\Delta C_+\|_{\mathrm{ui}}}{\widetilde{\gamma}} \leqslant \frac{\|\Delta C\|_{\mathrm{ui}}}{\widetilde{\gamma}}. \tag{A.37}$$

Finally,

$$\|\widetilde{\Lambda}_\pm - \Lambda_\pm\|_{\mathrm{ui}} \leqslant \|\widetilde{\Lambda}_\pm - \widehat{\Lambda}_\pm\|_{\mathrm{ui}} + \|\widehat{\Lambda}_\pm - \Lambda_\pm\|_{\mathrm{ui}} \leqslant 2 \cdot \frac{\|\Delta C\|_{\mathrm{ui}}}{\min\{\gamma, \widetilde{\gamma}\}},$$

as was to be shown. □

Appendix B. Positive semidefinite matrix pencil

Let $\boldsymbol{A} - \lambda \boldsymbol{B}$ be a matrix pencil of order n; that is, $\boldsymbol{A},\ \boldsymbol{B} \in \mathbb{C}^{n\times n}$.

DEFINITION B.1 [40]. $\boldsymbol{A} - \lambda \boldsymbol{B}$ is said to be *Hermitian* if both $\boldsymbol{A}$, $\boldsymbol{B}$ are Hermitian, or *positive (semi)definite* if it is Hermitian and there exists $\lambda_0 \in \mathbb{R}$ such that $\boldsymbol{A} - \lambda_0 \boldsymbol{B} \succ 0$ ($\boldsymbol{A} - \lambda_0 \boldsymbol{B} \succeq 0$).

The concept of a positive semidefinite pencil is closely related to that of the so-called *definite pencil* in the past literature [56, 59, 60]. The latter only requires that some linear combination (with possibly complex coefficients) is positive definite and thus is necessarily a regular pencil; that is, $\det(\boldsymbol{A} - \lambda \boldsymbol{B}) \not\equiv 0$. Definition B.1

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

uses more restrictive linear combinations, and also a positive semidefinite pencil of this definition may possibly be singular; that is, possibly $\det(\boldsymbol{A}-\lambda\boldsymbol{B})\equiv 0$.

To include, possibly, the case in which $\boldsymbol{A}-\lambda\boldsymbol{B}$ is a singular pencil, we say that $\mu\neq\infty$ is a *finite eigenvalue* of $\boldsymbol{A}-\lambda\boldsymbol{B}$ if

$$\operatorname{rank}(\boldsymbol{A}-\mu\boldsymbol{B})<\max_{\lambda\in\mathbb{C}}\operatorname{rank}(\boldsymbol{A}-\lambda\boldsymbol{B}), \tag{B.1}$$

and that $x\in\mathbb{C}^n$ is a corresponding *eigenvector* if $0\neq x\notin\mathcal{N}(\boldsymbol{A})\cap\mathcal{N}(\boldsymbol{B})$ satisfies

$$\boldsymbol{A}x=\mu\boldsymbol{B}x, \tag{B.2}$$

or, equivalently, $0\neq x\in\mathcal{N}(\boldsymbol{A}-\mu\boldsymbol{B})\backslash(\mathcal{N}(\boldsymbol{A})\cap\mathcal{N}(\boldsymbol{B}))$, where $\mathcal{N}(\cdot)$ is the null space of a matrix.

In the rest of this subsection, $\boldsymbol{A}-\lambda\boldsymbol{B}$ is assumed to be a positive semidefinite pencil. Let the inertia of $\boldsymbol{B}$ be $(i_-(\boldsymbol{B}),i_0(\boldsymbol{B}),i_+(\boldsymbol{B}))$, meaning that $\boldsymbol{B}$ has $i_-(\boldsymbol{B})$ negative, $i_0(\boldsymbol{B})$ zero, and $i_+(\boldsymbol{B})$ positive eigenvalues, respectively, and set

$$n_-:=i_-(\boldsymbol{B}),\quad n_+:=i_+(\boldsymbol{B}),\quad r:=\operatorname{rank}(\boldsymbol{B})=n_++n_-.$$

Given $0\leqslant k_+\leqslant n_+$ and $0\leqslant k_-\leqslant n_-$, set

$$J_k=\begin{bmatrix}I_{k_+}&&\\&-I_{k_-}&\\&&\end{bmatrix}.$$

We proved the following theorem in [40, Lemma 3.8], but later found out that it had been obtained in [14, Theorem 4.1] for the regular pencil case and in [65, Theorem A1] for the positive definite Hermitian pencil case with nonsingular B.

THEOREM B.1 [14, 40, 65]. *Let $\boldsymbol{A}-\lambda\boldsymbol{B}$ be a positive semidefinite Hermitian pencil of order n, and suppose that $\lambda_0\in\mathbb{R}$ such that $\boldsymbol{A}-\lambda_0\boldsymbol{B}\succeq 0$.*

(1) *There exists a nonsingular $W\in\mathbb{C}^{n\times n}$ such that*

$$W^{\mathrm{H}}\boldsymbol{A}W=\begin{array}{c}\\ n_1\\ r-n_1\\ n-r\end{array}\!\!\begin{array}{c}\begin{array}{ccc}n_1 & r-n_1 & n-r\end{array}\\ \begin{bmatrix}\Lambda_1&&\\&\Lambda_0&\\&&\Lambda_\infty\end{bmatrix}\end{array},\quad W^{\mathrm{H}}\boldsymbol{B}W=\begin{array}{c}\\ n_1\\ r-n_1\\ n-r\end{array}\!\!\begin{array}{c}\begin{array}{ccc}n_1 & r-n_1 & n-r\end{array}\\ \begin{bmatrix}\Omega_1&&\\&\Omega_0&\\&&0\end{bmatrix}\end{array}, \tag{B.3}$$

where

(a) $\Lambda_1=\operatorname{diag}(s_1\alpha_1,\ldots,s_{n_1}\alpha_{n_1})$, $\Omega_1=\operatorname{diag}(s_1,\ldots,s_{n_1})$, $s_i=\pm1$, *and* $\Lambda_1-\lambda_0\Omega_1\succ 0$;

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

(b) $\Lambda_0 = \mathrm{diag}(\Lambda_{0,1}, \dots, \Lambda_{0,m+m_0})$ *and* $\Omega_0 = \mathrm{diag}(\Omega_{0,1}, \dots, \Omega_{0,m+m_0})$ *with*

$$\Lambda_{0,i} = t_i\lambda_0, \qquad \Omega_{0,i} = t_i = \pm 1 \quad \text{for } 1 \leqslant i \leqslant m,$$

$$\Lambda_{0,i} = \begin{bmatrix} 0 & \lambda_0 \\ \lambda_0 & 1 \end{bmatrix}, \quad \Omega_{0,i} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \text{for } m+1 \leqslant i \leqslant m+m_0.$$

There is no such pair (Λ_0, Ω_0) *if* $\boldsymbol{A} - \lambda_0\boldsymbol{B} \succ 0$. *Evidently* $m + 2m_0 = r - n_1$.

(c) $\Lambda_\infty = \mathrm{diag}(\alpha_{r+1}, \dots, \alpha_n) \succeq 0$ *with* $\alpha_i \in \{1, 0\}$ *for* $r+1 \leqslant i \leqslant n$.

The representations in (B.3) *are uniquely determined by* $\boldsymbol{A} - \lambda\boldsymbol{B}$, *up to a simultaneous permutation of the corresponding* 1×1 *and* 2×2 *diagonal block pairs* $(s_i\alpha_i, s_i)$ *for* $1 \leqslant i \leqslant n_1$, $(\Lambda_{0,i}, \Omega_{0,i})$ *for* $1 \leqslant i \leqslant m + m_0$, *and* $(\alpha_i, 0)$ *for* $r+1 \leqslant i \leqslant n$.

(2) $\boldsymbol{A} - \lambda\boldsymbol{B}$ *has* $n_+ + n_-$ *finite eigenvalues, all of which are real. Denote these finite eigenvalues by* $\lambda_i^{\pm}$, *and arrange them as (this ordering is different from the one we used in* [38, 40] *for the neg-type eigenvalues, in order to be consistent with what we are using in this paper for hyperbolic matrix polynomials; see Theorem* 3.1*)*

$$\lambda_1^- \leqslant \cdots \leqslant \lambda_{n_-}^- \leqslant \lambda_1^+ \leqslant \cdots \leqslant \lambda_{n_+}^+. \tag{B.4}$$

(3) $\{\gamma \in \mathbb{R} \mid \boldsymbol{A} - \gamma\boldsymbol{B} \succeq 0\} = [\lambda_{n_-}^-, \lambda_1^+]$. *Moreover, if* $\boldsymbol{A} - \lambda\boldsymbol{B}$ *is regular, then* $\boldsymbol{A} - \lambda\boldsymbol{B}$ *is a positive definite pencil if and only if* $\lambda_{n_-}^- < \lambda_1^+$, *in which case*

$$\{\gamma \in \mathbb{R} \mid \boldsymbol{A} - \gamma\boldsymbol{B} \succ 0\} = (\lambda_{n_-}^-, \lambda_1^+).$$

The next perturbation theorem for positive definite pencils seems to be new. It resembles the Wielandt–Lidskii–Mirsky inequality (6.25) and many others in [9, 33, 34, 56, 59].

THEOREM B.2. *Let* $\boldsymbol{A} - \lambda\boldsymbol{B}$ *and* $\widetilde{\boldsymbol{A}} - \lambda\widetilde{\boldsymbol{B}}$ *be two positive definite Hermitian pencils of order* n *with nonsingular* $\boldsymbol{B}$ *and* $\widetilde{\boldsymbol{B}}$, *admitting the following eigendecompositions (such decompositions are guaranteed by Theorem* B.1*):*

$$W^{\mathrm{H}}\boldsymbol{A}W = J\Lambda, \quad W^{\mathrm{H}}\boldsymbol{B}W = J, \tag{B.5a}$$

$$\widetilde{W}^{\mathrm{H}}\widetilde{\boldsymbol{A}}\widetilde{W} = \widetilde{J}\widetilde{\Lambda}, \quad \widetilde{W}^{\mathrm{H}}\widetilde{\boldsymbol{B}}\widetilde{W} = \widetilde{J}, \tag{B.5b}$$

where Λ *is diagonal with diagonal entries consisting of eigenvalues of* $\boldsymbol{A} - \lambda\boldsymbol{B}$ *in the ascending order,* $J = \mathrm{diag}(-I_{i_-(\boldsymbol{B})}, I_{i_+(\boldsymbol{B})})$, *and similarly for* $\widetilde{\Lambda}$ *and* $\widetilde{J}$. *Then, for any unitarily invariant norm* $\|\cdot\|_{\mathrm{ui}}$,

$$\|\widetilde{\Lambda} - \Lambda\|_{\mathrm{ui}} \leqslant \|W\|_2\|\widetilde{W}\|_2(\|\widetilde{\boldsymbol{A}} - \boldsymbol{A}\|_{\mathrm{ui}} + \xi\|\widetilde{\boldsymbol{B}} - \boldsymbol{B}\|_{\mathrm{ui}}), \tag{B.6}$$

where $\xi = \max\{\|\Lambda\|_2, \|\widetilde{\Lambda}\|_2\}$.

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

Proof. We have

$$\begin{aligned} &\boldsymbol{A}WW^{\mathrm{H}}\boldsymbol{B} - \boldsymbol{B}WW^{\mathrm{H}}\boldsymbol{A} = 0,\\ &\widetilde{\boldsymbol{A}}WW^{\mathrm{H}}\boldsymbol{B} - \widetilde{\boldsymbol{B}}WW^{\mathrm{H}}\boldsymbol{A} = \widetilde{\boldsymbol{A}}WW^{\mathrm{H}}\boldsymbol{B} - \widetilde{\boldsymbol{B}}WW^{\mathrm{H}}\boldsymbol{A} - (\boldsymbol{A}WW^{\mathrm{H}}\boldsymbol{B} - \boldsymbol{B}WW^{\mathrm{H}}\boldsymbol{A})\\ &\qquad = (\widetilde{\boldsymbol{A}} - \boldsymbol{A})WW^{\mathrm{H}}\boldsymbol{B} - (\widetilde{\boldsymbol{B}} - \boldsymbol{B})WW^{\mathrm{H}}\boldsymbol{A}. \end{aligned} \tag{B.7}$$

Premultiply and postmultiply (B.7) by $\widetilde{J}\widetilde{W}^{\mathrm{H}}$ and WJ, and plug the decompositions in (B.5) into (B.7) to get

$$\widetilde{\Lambda}\widetilde{W}^{-1}W - \widetilde{W}^{-1}W\Lambda = \widetilde{J}\widetilde{W}^{\mathrm{H}}(\widetilde{\boldsymbol{A}} - \boldsymbol{A})W - \widetilde{J}\widetilde{W}^{\mathrm{H}}(\widetilde{\boldsymbol{B}} - \boldsymbol{B})W\Lambda. \tag{B.8}$$

Switching the roles of $\boldsymbol{A} - \lambda\boldsymbol{B}$ and $\widetilde{\boldsymbol{A}} - \lambda\widetilde{\boldsymbol{B}}$, we conclude from (B.8) that

$$\Lambda W^{-1}\widetilde{W} - W^{-1}\widetilde{W}\widetilde{\Lambda} = JW^{\mathrm{H}}(\boldsymbol{A} - \widetilde{\boldsymbol{A}})\widetilde{W} - JW^{\mathrm{H}}(\boldsymbol{B} - \widetilde{\boldsymbol{B}})\widetilde{W}\widetilde{\Lambda}. \tag{B.9}$$

It follows from (B.8) and (B.9) that

$$\|\widetilde{\Lambda}\widetilde{W}^{-1}W - \widetilde{W}^{-1}W\Lambda\|_{\mathrm{ui}} \leqslant \|W\|_2\|\widetilde{W}\|_2(\|\widetilde{\boldsymbol{A}} - \boldsymbol{A}\|_{\mathrm{ui}} + \xi\|\widetilde{\boldsymbol{B}} - \boldsymbol{B}\|_{\mathrm{ui}}), \tag{B.10a}$$

$$\|\Lambda W^{-1}\widetilde{W} - W^{-1}\widetilde{W}\widetilde{\Lambda}\|_{\mathrm{ui}} \leqslant \|W\|_2\|\widetilde{W}\|_2(\|\widetilde{\boldsymbol{A}} - \boldsymbol{A}\|_{\mathrm{ui}} + \xi\|\widetilde{\boldsymbol{B}} - \boldsymbol{B}\|_{\mathrm{ui}}). \tag{B.10b}$$

Let $\widetilde{W}^{-1}W = U\Sigma V^{\mathrm{H}}$ be the SVD of $\widetilde{W}^{-1}W$, and set $\boldsymbol{C} = V^{\mathrm{H}}\Lambda V$ and $\widetilde{\boldsymbol{C}} = U^{\mathrm{H}}\widetilde{\Lambda}U$, both of which are Hermitian. It can be verified that

$$\begin{aligned} \widetilde{\Lambda}\widetilde{W}^{-1}W - \widetilde{W}^{-1}W\Lambda &= U(\widetilde{\boldsymbol{C}}\Sigma - \Sigma\boldsymbol{C})V^{\mathrm{H}},\\ \Lambda W^{-1}\widetilde{W} - W^{-1}\widetilde{W}\widetilde{\Lambda} &= V(\boldsymbol{C}\Sigma^{-1} - \Sigma^{-1}\widetilde{\boldsymbol{C}})U. \end{aligned}$$

Theorem 2.1 of [**8**] yields

$$\|\widetilde{\boldsymbol{C}} - \boldsymbol{C}\|_{\mathrm{ui}}^2 \leqslant \|\widetilde{\boldsymbol{C}}\Sigma - \Sigma\boldsymbol{C}\|_{\mathrm{ui}}\|\boldsymbol{C}\Sigma^{-1} - \Sigma^{-1}\widetilde{\boldsymbol{C}}\|_{\mathrm{ui}}. \tag{B.11}$$

Mirsky's theorem [**58**, page 204] says that

$$\|\widetilde{\Lambda} - \Lambda\|_{\mathrm{ui}} \leqslant \|\widetilde{\boldsymbol{C}} - \boldsymbol{C}\|_{\mathrm{ui}}. \tag{B.12}$$

The main result, (B.6), is now a consequence of (B.10)–(B.12). □

In Theorem B.2, the upper bound by (B.6) contains $\|W\|_2$ and $\|\widetilde{W}\|_2$. They can be bounded, too, in terms of extreme pos-type and neg-type eigenvalues.

THEOREM B.3. *Let $\boldsymbol{A} - \lambda\boldsymbol{B}$ be a positive definite Hermitian pencil of order n with nonsingular B and with eigenvalues given by and ordered as in (B.4), and let its eigen-decomposition be given by (B.5a). Then, for any $\lambda_0 \in (\lambda_{n_-}^-, \lambda_1^+)$,*

$$\|W\|_2 \leqslant \sqrt{\max\{\lambda_{n_+}^+ - \lambda_0, \lambda_0 - \lambda_1^-\}\|(\boldsymbol{A} - \lambda_0\boldsymbol{B})^{-1}\|_2}, \tag{B.13a}$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

$$\|W^{-1}\|_2 \leqslant \sqrt{\frac{1}{\min\{\lambda_1^+ - \lambda_0, \lambda_0 - \lambda_{n_-}^-\}}\|\boldsymbol{A} - \lambda_0\boldsymbol{B}\|_2}. \tag{B.13b}$$

In particular, taking $\lambda_0 = (\lambda_{n_-}^- + \lambda_1^+)/2$ *gives*

$$\|W\|_2 \leqslant \sqrt{(\lambda_{n_+}^+ - \lambda_1^-)\|(\boldsymbol{A} - \lambda_0\boldsymbol{B})^{-1}\|_2}, \tag{B.14a}$$

$$\|W^{-1}\|_2 \leqslant \sqrt{\frac{2}{\lambda_1^+ - \lambda_{n_-}^-}\|\boldsymbol{A} - \lambda_0\boldsymbol{B}\|_2}. \tag{B.14b}$$

Proof. For $\lambda_0 \in (\lambda_{n_-}^-, \lambda_1^+)$, $\boldsymbol{A} - \lambda_0\boldsymbol{B} \succ 0$. We have $\boldsymbol{A} - \lambda_0\boldsymbol{B} \succeq \lambda_{\min}(\boldsymbol{A} - \lambda_0\boldsymbol{B})I_n$, and thus

$$\lambda_{\min}(\boldsymbol{A}-\lambda_0\boldsymbol{B})\, W^{\mathrm{H}}W \preceq W^{\mathrm{H}}(\boldsymbol{A}-\lambda_0\boldsymbol{B})W = J(\Lambda-\lambda_0 I) \preceq \max\{\lambda_{n_+}^+-\lambda_0, \lambda_0-\lambda_1^-\}\, I,$$

which gives (B.13a). We also have

$$W^{\mathrm{H}}(\boldsymbol{A} - \lambda_0\boldsymbol{B})W = J(\Lambda - \lambda_0 I) \succeq \min\{\lambda_1^+ - \lambda_0, \lambda_0 - \lambda_{n_-}^-\}I$$

to give

$$W^{-\mathrm{H}}W^{-1} \preceq \frac{1}{\min\{\lambda_1^+ - \lambda_0, \lambda_0 - \lambda_{n_-}^-\}}(\boldsymbol{A} - \lambda_0\boldsymbol{B}),$$

which yields (B.13b). □

Appendix C. Proof of Theorem 8.2

We recall (5.4) to see that

$$\begin{aligned}\varsigma(x) &:= [(x^{\mathrm{H}}Bx)^2 - 4(x^{\mathrm{H}}Ax)(x^{\mathrm{H}}Cx)]^{1/2}\\ &= \pm x^{\mathrm{H}}[2\rho_\pm(x)A + B]x\\ &= \pm x^{\mathrm{H}}\boldsymbol{Q}'(\rho_\pm(x))x,\end{aligned} \tag{C.1}$$

and $\varsigma_0(x) = \varsigma(x)/\|x\|_2^2$. For a perturbation $E \in \mathbb{C}^{n\times n}$ which is assumed Hermitian, we define

$$\boldsymbol{Q}_E(\lambda) := \boldsymbol{Q}(\lambda) + E = \lambda^2 A + \lambda B + C + E. \tag{C.2}$$

When $\boldsymbol{Q}_E(\lambda)$ is also hyperbolic, the pos-type and neg-type Rayleigh quotients, denoted by $\rho_{E;\pm}$, can be defined for $\boldsymbol{Q}_E(\lambda)$. Accordingly, we will define ς_E and $\varsigma_{E;0}$, too. Specifically,

$$\rho_{E;\pm}(x) = \frac{-(x^{\mathrm{H}}Bx) \pm \{(x^{\mathrm{H}}Bx)^2 - 4(x^{\mathrm{H}}Ax)(x^{\mathrm{H}}[C+E]x)\}^{1/2}}{2(x^{\mathrm{H}}Ax)}, \tag{C.3}$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

and

$$\varsigma_E(x) := \{(x^{\mathrm{H}}Bx)^2 - 4(x^{\mathrm{H}}Ax)(x^{\mathrm{H}}[C+E]x)\}^{1/2} \tag{C.4a}$$
$$= \pm x^{\mathrm{H}}[2\rho_{E;\pm}(x)\,A + B]x,$$
$$\varsigma_{E;0}(x) := \frac{\varsigma_E(x)}{\|x\|_2^2}. \tag{C.4b}$$

LEMMA C.1. *Suppose that $\boldsymbol{Q}_E(\lambda)$ in (C.2) is also hyperbolic.*

(1) *Let (λ_1^+, u_1^+) and (μ_1^+, v_1^+) be the smallest eigenpair (by the smallest (largest) pos-type/neg-type eigenpair, we mean that the eigenvalue in question is the smallest (largest) of that given type. The same naming is used for the usual linear eigenpair, too) of the pos-type of $\boldsymbol{Q}(\lambda)$ and $\boldsymbol{Q}_E(\lambda)$, respectively. Then*

$$\frac{\lambda_{\min}(E)}{\varsigma_0(u_1^+)} \leqslant \lambda_1^+ - \mu_1^+ \leqslant \frac{\lambda_{\max}(E)}{\varsigma_{E;0}(v_1^+)}. \tag{C.5}$$

(2) *Let (λ_n^+, u_n^+) and (μ_n^+, v_n^+) be the largest eigenpair of the pos-type of $\boldsymbol{Q}(\lambda)$ and $\boldsymbol{Q}_E(\lambda)$, respectively. Then*

$$\frac{\lambda_{\min}(E)}{\varsigma_0(v_n^+)} \leqslant \lambda_n^+ - \mu_n^+ \leqslant \frac{\lambda_{\max}(E)}{\varsigma_{E;0}(u_n^+)}. \tag{C.6}$$

(3) *Let (λ_1^-, u_1^-) and (μ_1^-, v_1^-) be the smallest eigenpair of the neg-type of $\boldsymbol{Q}(\lambda)$ and $\boldsymbol{Q}_E(\lambda)$, respectively. Then*

$$\frac{\lambda_{\min}(E)}{\varsigma_0(v_1^-)} \leqslant \mu_1^- - \lambda_1^- \leqslant \frac{\lambda_{\max}(E)}{\varsigma_{E;0}(u_1^-)}. \tag{C.7}$$

(4) *Let (λ_n^-, u_n^-) and (μ_n^-, v_n^-) be the largest eigenpair of the neg-type of $\boldsymbol{Q}(\lambda)$ and $\boldsymbol{Q}_E(\lambda)$, respectively. Then*

$$\frac{\lambda_{\min}(E)}{\varsigma_0(u_n^-)} \leqslant \mu_n^- - \lambda_n^- \leqslant \frac{\lambda_{\max}(E)}{\varsigma_{E;0}(v_n^-)}. \tag{C.8}$$

Proof. As in the proof of Lemma A.4, we have

$$\mu_1^+ = \min_x \rho_{E;+}(x) \leqslant \rho_{E;+}(u_1^+) \leqslant \rho_+(u_1^+) + \delta_{\mathrm{ub}}^+(u_1^+) = \lambda_1^+ + \delta_{\mathrm{ub}}^+(u_1^+),$$

which gives

$$\mu_1^+ - \lambda_1^+ \leqslant \delta_{\mathrm{ub}}^+(u_1^+), \quad \lambda_1^+ - \mu_1^+ \leqslant \tilde{\delta}_{\mathrm{ub}}^+(v_1^+), \tag{C.9}$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

where the second inequality is actually obtained from the first one by switching the roles of $\boldsymbol{Q}(\lambda)$ and $\boldsymbol{Q}_E(\lambda)$. Now use (A.22) in the proof of Theorem 6.1 for $\Delta A = \Delta B = 0$ and $\Delta C = E$ to get item (1).

Similarly, we have

$$\lambda_n^+ = \max_x \rho_+(x) \geqslant \rho_+(v_n^+) \geqslant \rho_{E;+}(v_n^+) - \delta_{\mathrm{ub}}^+(v_n^+) = \mu_n^+ - \delta_{\mathrm{ub}}^+(v_n^+),$$

which gives

$$\mu_n^+ - \lambda_n^+ \leqslant \delta_{\mathrm{ub}}^+(v_n^+), \quad \lambda_n^+ - \mu_n^+ \leqslant \tilde{\delta}_{\mathrm{ub}}^+(u_n^+), \tag{C.10}$$

where the second inequality is also obtained from switching the roles of $\boldsymbol{Q}(\lambda)$ and $\boldsymbol{Q}_E(\lambda)$. Now use (A.22) in the proof of Theorem 6.1 for $\Delta A = \Delta B = 0$ and $\Delta C = E$ to get item (2).

Items (3) and (4) are corollaries of items (2) and (1) applied to $\boldsymbol{Q}(-\lambda)$ and $\boldsymbol{Q}_E(-\lambda)$. □

LEMMA C.2. *$\boldsymbol{Q}_E(\lambda)$ with $E = -\sigma I$ is hyperbolic if*

$$\sigma > -\frac{(\lambda_1^+ - \lambda_n^-)^2 \lambda_{\min}(A)}{4}. \tag{C.11}$$

Proof. For any vector $x \neq 0$, we have

$$\begin{aligned}
&(x^{\mathrm{H}}Bx)^2 - 4(x^{\mathrm{H}}Ax)(x^{\mathrm{H}}[C - \sigma I]x) \\
&\quad = (x^{\mathrm{H}}Bx)^2 - 4(x^{\mathrm{H}}Ax)(x^{\mathrm{H}}Cx) + 4\sigma(x^{\mathrm{H}}Ax)(x^{\mathrm{H}}x) \\
&\quad = [\rho_+(x) - \rho_-(x)]^2 (x^{\mathrm{H}}Ax)^2 + 4\sigma(x^{\mathrm{H}}Ax)(x^{\mathrm{H}}x) \\
&\quad \geqslant (x^{\mathrm{H}}Ax)(x^{\mathrm{H}}x) \left[(\lambda_1^+ - \lambda_n^-)^2 \frac{x^{\mathrm{H}}Ax}{x^{\mathrm{H}}x} + 4\sigma \right] \\
&\quad \geqslant (x^{\mathrm{H}}Ax)(x^{\mathrm{H}}x)[(\lambda_1^+ - \lambda_n^-)^2 \lambda_{\min}(A) + 4\sigma] \\
&\quad > 0,
\end{aligned}$$

where the last inequality holds because of (C.11). □

So ς_E and $\varsigma_{E;0}$ are well defined for any $E = -\sigma I$ satisfying (C.11). To emphasize such special $E = -\sigma I$, we introduce the notation

$$\varsigma_\sigma(x) := \varsigma_E(v), \quad \varsigma_{\sigma;0}(v) := \varsigma_{E;0}(v) \quad \text{for } E = -\sigma I. \tag{C.12}$$

For $\boldsymbol{\rho} \in (\lambda_1^{\mathrm{typ}}, \lambda_n^{\mathrm{typ}})$, it follows from Theorem 3.1 that the largest eigenvalue, denoted by ω_1, of the matrix $\boldsymbol{Q}(\boldsymbol{\rho})$ is nonnegative, and thus this $\sigma = \omega_1$ automatically satisfies (C.11). But the smallest eigenvalue, denoted also by ω_1, of $\boldsymbol{Q}(\boldsymbol{\rho})$ is nonpositive, and (C.11) may fail for $\sigma = \omega_1$ unless $|\omega_1|$ is sufficiently tiny.

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

LEMMA C.3. *Given* $\lambda_1^{\rm typ} \leqslant \rho \leqslant \lambda_n^{\rm typ}$, *let* (ω_1, v_1) *be the largest eigenpair of the matrix* $\boldsymbol{Q}(\rho)$ *if* $(\mathrm{typ}, \ell) \in \{(+,1),(-,n)\}$ *or the smallest eigenpair of the matrix* $\boldsymbol{Q}(\rho)$ *if* $(\mathrm{typ}, \ell) \in \{(+,n),(-,1)\}$. *If* (C.11) *holds with* $\sigma = \omega_1$, *then*

$$\frac{\varsigma_0(u_1^+)}{\varsigma_{\omega_1;0}(v_1)}(\rho - \lambda_1^+) \leqslant \frac{\omega_1}{\varsigma_{\omega_1;0}(v_1)} \leqslant \rho - \lambda_1^+ \quad \textit{for } (\mathrm{typ},\ell) = (+,1), \tag{C.13a}$$

$$\frac{\varsigma_{\omega_1;0}(u_n^+)}{\varsigma_0(v_1)}(\lambda_n^+ - \rho) \leqslant \frac{-\omega_1}{\varsigma_0(v_1)} \leqslant \lambda_n^+ - \rho \quad \textit{for } (\mathrm{typ},\ell) = (+,n), \tag{C.13b}$$

$$\frac{\varsigma_{\omega_1;0}(u_1^-)}{\varsigma_0(v_1)}(\rho - \lambda_1^-) \leqslant \frac{-\omega_1}{\varsigma_0(v_1)} \leqslant \rho - \lambda_1^- \quad \textit{for } (\mathrm{typ},\ell) = (-,1), \tag{C.13c}$$

$$\frac{\varsigma_0(u_n^-)}{\varsigma_{\omega_1;0}(v_1)}(\lambda_n^- - \rho) \leqslant \frac{\omega_1}{\varsigma_{\omega_1;0}(v_1)} \leqslant \lambda_n^- - \rho \quad \textit{for } (\mathrm{typ},\ell) = (-,n). \tag{C.13d}$$

Moreover, for ρ *sufficiently close to* $\lambda_\ell^{\rm typ}$,

$$\frac{\omega_1}{\varsigma_{\omega_1;0}(v_1)} = \rho - \lambda_1^+ + O([\rho - \lambda_1^+]^2) \quad \textit{for } (\mathrm{typ},\ell) = (+,1), \tag{C.14a}$$

$$\frac{-\omega_1}{\varsigma_0(v_1)} = \lambda_n^+ - \rho + O([\lambda_n^+ - \rho]^2) \quad \textit{for } (\mathrm{typ},\ell) = (+,n), \tag{C.14b}$$

$$\frac{-\omega_1}{\varsigma_0(v_1)} = \rho - \lambda_1^- + O([\rho - \lambda_1^-]^2) \quad \textit{for } (\mathrm{typ},\ell) = (-,1), \tag{C.14c}$$

$$\frac{\omega_1}{\varsigma_{\omega_1;0}(v_1)} = \lambda_n^- - \rho + O([\lambda_n^- - \rho]^2) \quad \textit{for } (\mathrm{typ},\ell) = (-,n). \tag{C.14d}$$

Proof. Consider the case when $(\mathrm{typ},\ell) = (+,1)$. We have $\omega_1 \geqslant 0$ and $[\boldsymbol{Q}(\rho) - \omega_1 I]\, v_1 = 0$. Since ω_1 is the largest eigenvalue of $\boldsymbol{Q}(\rho)$, $\boldsymbol{Q}(\rho) - \omega_1 I \preceq 0$. Thus, (ρ, v_1) is the smallest pos-type eigenpair of $\boldsymbol{Q}_E(\lambda)$ with $E = -\omega_1 I$. By Lemma C.1,

$$\frac{\omega_1}{\varsigma_{E;0}(v_1)} \leqslant \rho - \lambda_1^+ \leqslant \frac{\omega_1}{\varsigma_0(u_1)},$$

which gives (C.13a). To prove (C.14a), we denote by $\alpha(t)$ the largest eigenvalue of $\boldsymbol{Q}(t)$ near $t = \lambda_1^+$. Then $\alpha(\lambda_1^+) = 0$ and $\alpha(\rho) = \omega_1$. Note that

$$\boldsymbol{Q}(\rho)v_1 = \omega_1 v_1 \quad \Rightarrow \quad v_1^{\rm H}\boldsymbol{Q}(\rho)v_1 = \omega_1 v_1^{\rm H} v_1 \quad \Rightarrow \quad v_1^{\rm H}[\boldsymbol{Q}(\rho) - \omega_1 I]v_1 = 0;$$

that is, ρ is a Rayleigh quotient of $\boldsymbol{Q}_E(\lambda)$ with $E = -\omega_1 I$. Therefore

$$\alpha'(\rho) = \frac{v_1^{\rm H}\boldsymbol{Q}'(\rho)v_1}{v_1^{\rm H}v_1} = \frac{v_1^{\rm H}\boldsymbol{Q}_E'(\rho)v_1}{v_1^{\rm H}v_1} = \varsigma_{\omega_1;0}(v_1),$$

where the first equality is due to [**58**, page 183], and the third equality is due to (C.1). Finally, $\alpha(\lambda_1^+) = \alpha(\rho) + \varsigma_{\omega_1;0}(v_1)(\lambda_1^+ - \rho) + O(|\lambda_1^+ - \rho|^2)$, which leads to (C.14a). □

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

REMARK C.1. There is a different proof of Lemma C.3, without using Lemma C.1. For the case when (typ, ℓ) = (+, 1), ($\boldsymbol{\rho}$, v_1) is the smallest pos-type eigenpair of $\boldsymbol{Q}_E(\lambda) = \lambda^2 A + \lambda B + C - \omega_1 I$. By direct calculations

$$\omega_1 = \omega_1 - \frac{u_1^{\mathrm{H}} \boldsymbol{Q}(\boldsymbol{\rho}) u_1}{u_1^{\mathrm{H}} u_1} + \varsigma_0(u_1)(\boldsymbol{\rho} - \lambda_1^+) + \frac{u_1^{\mathrm{H}} A u_1}{u_1^{\mathrm{H}} u_1}(\boldsymbol{\rho} - \lambda_1^+)^2, \tag{C.15a}$$

$$\omega_1 = \frac{v_1^{\mathrm{H}} \boldsymbol{Q}(\lambda_1^+) v_1}{v_1^{\mathrm{H}} v_1} \qquad + \varsigma_{\omega_1;0}(v_1)(\boldsymbol{\rho} - \lambda_1^+) - \frac{v_1^{\mathrm{H}} A v_1}{v_1^{\mathrm{H}} v_1}(\boldsymbol{\rho} - \lambda_1^+)^2. \tag{C.15b}$$

(In fact,

$$\begin{aligned}
&u_1^{\mathrm{H}} A u_1 (\boldsymbol{\rho} - \lambda_1^+)^2 + \varsigma(u_1)(\boldsymbol{\rho} - \lambda_1^+) \\
&\quad = u_1^{\mathrm{H}} A u_1 \big[\boldsymbol{\rho}^2 - 2\boldsymbol{\rho}\lambda_1^+ + (\lambda_1^+)^2\big] + (2\lambda_1^+ u_1^{\mathrm{H}} A u_1 + u_1^{\mathrm{H}} B u_1)(\boldsymbol{\rho} - \lambda_1^+) \\
&\quad = \boldsymbol{\rho}^2 u_1^{\mathrm{H}} A u_1 + \boldsymbol{\rho} u_1^{\mathrm{H}} B u_1 - (\lambda_1^+)^2 u_1^{\mathrm{H}} A u_1 - \lambda_1^+ u_1^{\mathrm{H}} B u_1 \\
&\quad = u_1^{\mathrm{H}} \boldsymbol{Q}(\boldsymbol{\rho}) u_1 - u_1^{\mathrm{H}} \boldsymbol{Q}(\lambda_1^+) u_1 \\
&\quad = u_1^{\mathrm{H}} \boldsymbol{Q}(\boldsymbol{\rho}) u_1, \\
&v_1^{\mathrm{H}} A v_1 (\boldsymbol{\rho} - \lambda_1^+)^2 - \varsigma_{\omega_1}(v_1)(\boldsymbol{\rho} - \lambda_1^+) \\
&\quad = v_1^{\mathrm{H}} A v_1 \big[\boldsymbol{\rho}^2 - 2\boldsymbol{\rho}\lambda_1^+ + (\lambda_1^+)^2\big] - (2\boldsymbol{\rho} v_1^{\mathrm{H}} A v_1 + v_1^{\mathrm{H}} B v_1)(\boldsymbol{\rho} - \lambda_1^+) \\
&\quad = (\lambda_1^+)^2 v_1^{\mathrm{H}} A v_1 + \lambda_1^+ v_1^{\mathrm{H}} B v_1 - \boldsymbol{\rho}^2 v_1^{\mathrm{H}} A v_1 - \boldsymbol{\rho} v_1^{\mathrm{H}} B v_1 \\
&\quad = v_1^{\mathrm{H}} \boldsymbol{Q}(\lambda_1^+) v_1 - v_1^{\mathrm{H}} \boldsymbol{Q}(\boldsymbol{\rho}) v_1 \\
&\quad = v_1^{\mathrm{H}} \boldsymbol{Q}(\lambda_1^+) v_1 - \omega_1 v_1^{\mathrm{H}} v_1.
\end{aligned}$$

They lead to the equations in (C.15) right away.) Along with $\boldsymbol{Q}(\boldsymbol{\rho}) - \omega_1 I \preceq 0$, $\boldsymbol{Q}(\lambda_1^+) \preceq 0$, they yield

$$\frac{\omega_1}{\varsigma_{\omega_1;0}(v_1)} \leqslant \boldsymbol{\rho} - \lambda_1^+ \leqslant \frac{\omega_1}{\varsigma_0(u_1)},$$

and then

$$\frac{\varsigma_0(u_1)}{\varsigma_{\omega_1;0}(v_1)}(\boldsymbol{\rho} - \lambda_1^+) \leqslant \frac{\omega_1}{\varsigma_{\omega_1;0}(v_1)} \leqslant \boldsymbol{\rho} - \lambda_1^+,$$

which is (C.13a).

While Lemmas C.4 and C.5 below are stated for any $g \in \mathbb{P}_{m-1}$ with the specified conditions satisfied, in their eventual application, it will be taken to be the one that minimizes ε_g.

LEMMA C.4. *Given $x \in \mathbb{C}^n$, assign $\boldsymbol{\rho}_\pm = \rho_\pm(x)$ and $\boldsymbol{\rho}_{g;\pm} = \rho_\pm(g(\boldsymbol{Q}(\boldsymbol{\rho}_+))x)$ for any $g \in \mathbb{P}_{m-1}$. Suppose that $\lambda_1^{\mathrm{typ}} \leqslant \boldsymbol{\rho}_{\mathrm{typ}} < \lambda_2^{\mathrm{typ}}$ if $\ell = 1$ or $\lambda_{n-1}^{\mathrm{typ}} < \boldsymbol{\rho}_{\mathrm{typ}} \leqslant \lambda_n^{\mathrm{typ}}$ if $\ell = n$, and let the eigenvalues of the matrix $\boldsymbol{Q}(\boldsymbol{\rho}_{\mathrm{typ}})$ be ω_j for $1 \leqslant j \leqslant n$, which*

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

can be arranged as

$$\begin{aligned}&\omega_1 > 0 > \omega_2 \geqslant \cdots \geqslant \omega_n \quad \text{if } (\mathrm{typ}, \ell) \in \{(+,1),(-,n)\}, \quad \text{or},\\ &\omega_1 < 0 < \omega_2 \leqslant \cdots \leqslant \omega_n \quad \text{if } (\mathrm{typ}, \ell) \in \{(+,n),(-,1)\}.\end{aligned}$$

Denote by v_1 *the eigenvector of* $\boldsymbol{Q}(\boldsymbol{\rho}_{\mathrm{typ}})$ *associated with its eigenvalue* ω_1. *Then, for a* $g \in \mathbb{P}_{m-1}$ *such that* $g(\omega_1) \neq 0$ *and*

$$\varepsilon_g := \max_{i \neq 1} \frac{|g(\omega_i)|}{|g(\omega_1)|} < 1, \tag{C.16}$$

we have

$$|\boldsymbol{\rho}_{g;\mathrm{typ}} - \lambda_\ell^{\mathrm{typ}}| \leqslant |\boldsymbol{\rho}_{\mathrm{typ}} - \lambda_\ell^{\mathrm{typ}}| - \frac{|\omega_1|}{|\boldsymbol{\rho}_{\mathrm{typ}} - \boldsymbol{\rho}_{g;\mathrm{typ}'}|\, a(v_1)} + \frac{|\omega_1|}{|\boldsymbol{\rho}_{\mathrm{typ}} - \boldsymbol{\rho}_{g;\mathrm{typ}'}|\, a(v_1)} h(\varepsilon_g, \omega_1), \tag{C.17}$$

where typ′ *is the opposite type of* typ, *and*

$$h(\varepsilon_g, \omega_1) = 1 - \frac{1-\varepsilon_g^2}{\left(1 + \varepsilon_g |\omega_1|^{1/2} \tau_A^{1/2}\right)^2}, \quad \tau_A = \frac{1}{|\omega_2|} \frac{\|A\|_2}{a(v_1)}. \tag{C.18}$$

Proof. Consider the case when $(\mathrm{typ}, \ell) = (+, 1)$, and write $\boldsymbol{\rho} = \boldsymbol{\rho}_+$. Without loss of generality, we may assume that $\|v_1\|_2 = 1$. Let the eigenvalue decomposition of $\boldsymbol{Q}(\boldsymbol{\rho})$ be

$$\boldsymbol{Q}(\boldsymbol{\rho}) = V \Sigma V^{\mathrm{H}}, \quad V = [v_1, \ldots, v_n], \quad \Sigma = \mathrm{diag}(\omega_1, \ldots, \omega_n),$$

where $\omega_1 > 0 > \omega_2 \geqslant \cdots \geqslant \omega_n$ and $V^{\mathrm{H}} V = I_n$. Set

$$\hat{x} = V^{\mathrm{H}} x = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix}, \quad \hat{x}_2 = \hat{x} - \xi_1 e_1 = \begin{bmatrix} 0 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix}.$$

Then

$$0 = x^{\mathrm{H}} \boldsymbol{Q}(\boldsymbol{\rho}) x = \hat{x}^{\mathrm{H}} \Sigma \hat{x} = \omega_1 |\xi_1|^2 + \sum_{i \neq 1} \omega_i |\xi_i|^2. \tag{C.19}$$

Note that, for any vector z, $z^{\mathrm{H}} \boldsymbol{Q}(\lambda) z = z^{\mathrm{H}} A z\, [\lambda - \rho_+(z)][\lambda - \rho_-(z)]$. Substitute $\lambda = \boldsymbol{\rho}$ and $z = g(\boldsymbol{Q}(\boldsymbol{\rho}))x$ to get

$$\begin{aligned}\boldsymbol{\rho}_g - \lambda_1^+ &= \boldsymbol{\rho} - \lambda_1^+ - \frac{1}{\boldsymbol{\rho} - \boldsymbol{\rho}_{g;-}} \cdot \frac{x^{\mathrm{H}} g(\boldsymbol{Q}(\boldsymbol{\rho}))^{\mathrm{H}} \boldsymbol{Q}(\boldsymbol{\rho}) g(\boldsymbol{Q}(\boldsymbol{\rho})) x}{x^{\mathrm{H}} g(\boldsymbol{Q}(\boldsymbol{\rho}))^{\mathrm{H}} A g(\boldsymbol{Q}(\boldsymbol{\rho})) x} \\ &= \boldsymbol{\rho} - \lambda_1^+ - \frac{1}{\boldsymbol{\rho} - \boldsymbol{\rho}_{g;-}} \cdot \frac{\hat{x}^{\mathrm{H}} g(\Sigma)^{\mathrm{H}} \Sigma g(\Sigma) \hat{x}}{\hat{x}^{\mathrm{H}} g(\Sigma)^{\mathrm{H}} \widehat{A} g(\Sigma) \hat{x}},\end{aligned} \tag{C.20}$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

where $\widehat{A} = V^{\mathrm{H}}AV$ and $\boldsymbol{\rho}_g = \boldsymbol{\rho}_{g;+}$. We need to estimate the right-hand side of (C.20). For that, we have

$$\begin{aligned}
\hat{x}^{\mathrm{H}}g(\Sigma)^{\mathrm{H}}\Sigma g(\Sigma)\hat{x} &= \omega_1|g(\omega_1)|^2|\xi_1|^2 + \sum_{i\neq 1}\omega_i|g(\omega_i)|^2|\xi_i|^2 \\
&\geqslant \omega_1|g(\omega_1)|^2|\xi_1|^2 + \varepsilon_g^2|g(\omega_1)|^2\sum_{i\neq 1}\omega_i|\xi_i|^2 \\
&= \omega_1|g(\omega_1)|^2|\xi_1|^2 - \varepsilon_g^2|g(\omega_1)|^2\omega_1|\xi_1|^2 \quad \text{(by (C.19))} \\
&= (1-\varepsilon_g^2)\omega_1|g(\omega_1)|^2|\xi_1|^2, \qquad \text{(C.21)} \\
\hat{x}^{\mathrm{H}}g(\Sigma)^{\mathrm{H}}\widehat{A}g(\Sigma)\hat{x} &= \|g(\Sigma)\hat{x}\|_{\widehat{A}}^2 \\
&= \|g(\omega_1)\xi_1 e_1 + g(\Sigma)\hat{x}_2\|_{\widehat{A}}^2 \\
&\leqslant [|g(\omega_1)|\,|\xi_1|\,\|e_1\|_{\widehat{A}} + \|g(\Sigma)\hat{x}_2\|_{\widehat{A}}]^2 \\
&\leqslant [|g(\omega_1)|\,|\xi_1|\,\|e_1\|_{\widehat{A}} + \varepsilon_g|g(\omega_1)|\|\hat{x}_2\|_{\widehat{A}}]^2 \\
&\leqslant \left[|g(\omega_1)|\,|\xi_1|\,\|e_1\|_{\widehat{A}} + \varepsilon_g|g(\omega_1)|\left(\|A\|_2\frac{\omega_1}{-\omega_2}|\xi_1|^2\right)^{1/2}\right]^2 \qquad \text{(C.22)} \\
&= |g(\omega_1)|^2|\xi_1|^2 v_1^{\mathrm{H}}Av_1\left[1+\varepsilon_g\left(\frac{\omega_1}{-\omega_2}\frac{\|A\|_2}{v_1^{\mathrm{H}}Av_1}\right)^{1/2}\right]^2, \qquad \text{(C.23)}
\end{aligned}$$

where the inequality sign at (C.22) holds because

$$\|\hat{x}_2\|_{\widehat{A}}^2 \leqslant \|\widehat{A}\|_2\|\hat{x}_2\|_2^2 = \|V^{\mathrm{H}}AV\|_2\sum_{i\neq 1}|\xi_i|^2 \leqslant \|A\|_2\frac{\sum_{i\neq 1}\omega_i|\xi_i|^2}{\omega_2} = \|A\|_2\frac{\omega_1}{-\omega_2}|\xi_1|^2$$

by (C.19). Thus, from (C.20), (C.21), and (C.23),

$$\boldsymbol{\rho}_g - \lambda_1^+ \leqslant \boldsymbol{\rho} - \lambda_1^+ - \frac{\omega_1}{(\boldsymbol{\rho}-\boldsymbol{\rho}_{g;-})v_1^{\mathrm{H}}Av_1}\frac{1-\varepsilon_g^2}{\left[1+\varepsilon_g\left(\frac{\omega_1}{-\omega_2}\frac{\|A\|_2}{v_1^{\mathrm{H}}Av_1}\right)^{1/2}\right]^2}, \qquad \text{(C.24)}$$

which gives (C.17) for the case when $(\mathrm{typ}, \ell) = (+, 1)$. □

LEMMA C.5. *Under the conditions of Lemma C.4, we have*

$$|\boldsymbol{\rho}_{g;\mathrm{typ}} - \lambda_\ell^{\mathrm{typ}}| \leqslant \frac{|\omega_1|}{\varsigma_0(v_1)}\varepsilon_g^2 + \frac{1-\varepsilon_g^2}{\varsigma_0(v_1)}(3\tau_A^{1/2} + 2\chi_1)\varepsilon_g|\omega_1|^{3/2} + O(\omega_1^2), \qquad \text{(C.25)}$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

provided that

$$\varepsilon_g|\omega_1|^{1/2}\max\{\tau_A^{1/2},\zeta\chi_1\}<1,\quad 4a(v_1)|\omega_1|<\varsigma_0(v_1)^2, \tag{C.26}$$

where τ_A, τ_B, *and* τ_C *are defined in* (8.25),

$$\chi_1=\frac{b_0(v_1)^2\tau_B^{1/2}+2a(v_1)c_0(v_1)(\tau_A^{1/2}+\tau_C^{1/2})}{\varsigma_0(v_1)^2}, \tag{C.27}$$

$$\zeta=4+6\varepsilon_g\omega_1^{1/2}\tau_B^{1/2}+4\varepsilon_g^2\omega_1\tau_B+\varepsilon_g^3\omega_1^{3/2}\tau_B^{3/2}, \tag{C.28}$$

and the shift $\lambda_0\geqslant\lambda_n^+$ *in defining* $b_0(\,\cdot\,)$ *and* $c_0(\,\cdot\,)$ *in* (8.21). *Alternatively,*

$$\begin{aligned}|\boldsymbol{\rho}_{g;\mathrm{typ}}-\lambda_\ell^{\mathrm{typ}}|\leqslant{}&\varepsilon_g^2|\boldsymbol{\rho}_{\mathrm{typ}}-\lambda_\ell^{\mathrm{typ}}|+(1-\varepsilon_g^2)(3\tau_A^{1/2}+2\chi_1)\varepsilon_g|\boldsymbol{\rho}_{\mathrm{typ}}-\lambda_\ell^{\mathrm{typ}}|^{3/2}\\&+O(|\boldsymbol{\rho}_{\mathrm{typ}}-\lambda_\ell^{\mathrm{typ}}|^2),\end{aligned} \tag{C.29}$$

provided that

$$|\boldsymbol{\rho}_{\mathrm{typ}}-\lambda_\ell^{\mathrm{typ}}|<\max\left\{\frac{\varsigma_0(v_1)}{4a(v_1)},\frac{1}{\varsigma_0(v_1)\varepsilon_g^2\max\{\tau_A,\zeta^2\chi_1^2\}}\right\}. \tag{C.30}$$

Proof. Consider the case when $(\mathrm{typ},\ell)=(+,1)$, and write $\boldsymbol{\rho}=\boldsymbol{\rho}_+$. Without loss of generality, we may assume that $\|v_1\|_2=1$. Write $x_g=g(\boldsymbol{Q}(\boldsymbol{\rho}))x$, and

$$\begin{gathered}t_M=\omega_1^{1/2}\tau_M^{1/2}\quad\text{for } M=A,B,C,\\ \boldsymbol{a}=a(v_1),\quad \boldsymbol{b}=b(v_1),\quad \boldsymbol{c}=c(v_1),\\ \boldsymbol{b}_0=b_0(v_1),\quad \boldsymbol{c}_0=c_0(v_1).\end{gathered}$$

By Lemma C.4, $\boldsymbol{\rho}_g\leqslant\boldsymbol{\rho}$ (see (C.24)) and

$$\boldsymbol{\rho}_g-\lambda_1^+\leqslant\delta_0+\delta_1+\delta_2+\delta_3, \tag{C.31}$$

where

$$\begin{aligned}0\leqslant\delta_0&=\boldsymbol{\rho}-\lambda_1^+-\frac{\omega_1}{\varsigma_{\omega_1;0}(v_1)}=O(|\boldsymbol{\rho}-\lambda_1^+|^2)=O(\omega_1^2),\\ \delta_1&=\frac{\omega_1}{\varsigma_{\omega_1;0}(v_1)}-\frac{\omega_1}{(\boldsymbol{\rho}_g-\boldsymbol{\rho}_{g;-})\boldsymbol{a}},\\ \delta_2&=\frac{\omega_1}{(\boldsymbol{\rho}_g-\boldsymbol{\rho}_{g;-})\boldsymbol{a}}-\frac{\omega_1}{(\boldsymbol{\rho}-\boldsymbol{\rho}_{g;-})\boldsymbol{a}},\\ \delta_3&=\frac{\omega_1}{(\boldsymbol{\rho}-\boldsymbol{\rho}_{g;-})\boldsymbol{a}}h(\varepsilon_g,\omega_1).\end{aligned} \tag{C.32}$$

The rest of the proof is mainly to estimate δ_1, δ_2, and δ_3.

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

For δ_2, we have

$$0 \leqslant \delta_2 = \frac{\omega_1}{\boldsymbol{a}} \frac{\boldsymbol{\rho} - \boldsymbol{\rho}_g}{(\boldsymbol{\rho}_g - \boldsymbol{\rho}_{g;-})(\boldsymbol{\rho} - \boldsymbol{\rho}_{g;-})} \leqslant \frac{\omega_1}{\boldsymbol{a}} \frac{\boldsymbol{\rho} - \lambda_1^+}{(\boldsymbol{\rho}_g - \boldsymbol{\rho}_{g;-})(\boldsymbol{\rho} - \boldsymbol{\rho}_{g;-})} = O(\omega_1^2), \tag{C.33}$$

where we have used (C.14a).

Consider δ_1. If $4\boldsymbol{a}\omega_1 < \boldsymbol{b}^2 - 4\boldsymbol{a}\boldsymbol{c}$, which holds for sufficiently tiny ω_1, then

$$\frac{1}{\varsigma_{\omega_1}(v_1)} = \frac{1}{\sqrt{\boldsymbol{b}^2 - 4\boldsymbol{a}(\boldsymbol{c} - \omega_1)}} = \frac{1}{\sqrt{\boldsymbol{b}^2 - 4\boldsymbol{a}\boldsymbol{c}}} \left[1 - \frac{2\boldsymbol{a}}{\boldsymbol{b}^2 - 4\boldsymbol{a}\boldsymbol{c}}\omega_1 + O(\omega_1^2)\right]. \tag{C.34}$$

By item (2) of Lemma A.2, any shift $\lambda_0 \geqslant \lambda_n^+$ makes $\boldsymbol{Q}_{\lambda_0}(\lambda)$ overdamped; that is, $B_{\lambda_0} \succ 0$ and $C_{\lambda_0} \succeq 0$. It can be verified that

$$\boldsymbol{b}_0^2 - 4\boldsymbol{a}\boldsymbol{c}_0 = \boldsymbol{b}^2 - 4\boldsymbol{a}\boldsymbol{c} = [\varsigma(v_1)]^2.$$

We get, similarly to (C.23),

$$\begin{aligned}
\boldsymbol{a}\,|g(\omega_1)|^2|\xi_1|^2(1 - 2\varepsilon_g t_A) &\leqslant x_g^{\mathrm{H}} A x_g \quad \leqslant \boldsymbol{a}\,|g(\omega_1)|^2|\xi_1|^2(1 + \varepsilon_g t_A)^2,\\
\boldsymbol{b}_0|g(\omega_1)|^2|\xi_1|^2(1 - 2\varepsilon_g t_B) &\leqslant x_g^{\mathrm{H}} B_{\lambda_0} x_g \leqslant \boldsymbol{b}_0|g(\omega_1)|^2|\xi_1|^2(1 + \varepsilon_g t_B)^2,\\
\boldsymbol{c}_0|g(\omega_1)|^2|\xi_1|^2(1 - 2\varepsilon_g t_C) &\leqslant x_g^{\mathrm{H}} C_{\lambda_0} x_g \leqslant \boldsymbol{c}_0|g(\omega_1)|^2|\xi_1|^2(1 + \varepsilon_g t_C)^2.
\end{aligned}$$

Note that $\boldsymbol{\rho}_g - \lambda_0$ (recalling that $\boldsymbol{\rho}_g$ is the shorthand for $\boldsymbol{\rho}_{g;+}$) and $\boldsymbol{\rho}_{g;-} - \lambda_0$ are two distinct roots of $x_g^{\mathrm{H}} A x_g \lambda^2 + x_g^{\mathrm{H}} B_{\lambda_0} x_g \lambda + x_g^{\mathrm{H}} C_{\lambda_0} x_g = 0$ in λ. So

$$\begin{aligned}
\frac{1}{(\boldsymbol{\rho}_g - \boldsymbol{\rho}_{g;-})\boldsymbol{a}} &= \frac{x_g^{\mathrm{H}} A x_g}{\boldsymbol{a}\sqrt{(x_g^{\mathrm{H}} B_{\lambda_0} x_g)^2 - 4(x_g^{\mathrm{H}} A x_g)(x_g^{\mathrm{H}} C_{\lambda_0} x_g)}}\\
&\geqslant \frac{1 - 2\varepsilon_g t_A}{\sqrt{\boldsymbol{b}_0^2(1 + \varepsilon_g t_B)^4 - 4\boldsymbol{a}\boldsymbol{c}_0(1 - 2\varepsilon_g t_A)(1 - 2\varepsilon_g t_C)}}\\
&= \frac{1 - 2\varepsilon_g t_A}{\sqrt{\boldsymbol{b}_0^2 - 4\boldsymbol{a}\boldsymbol{c}_0 + 4\varepsilon_g(\boldsymbol{b}_0^2 t_B + 2\boldsymbol{a}\boldsymbol{c}_0 t_A + 2\boldsymbol{a}\boldsymbol{c}_0 t_C) + 2\varepsilon_g^2(3\boldsymbol{b}_0^2 t_B^2 - 8\boldsymbol{a}\boldsymbol{c}_0 t_A t_C) + 4\varepsilon_g^3 \boldsymbol{b}_0^2 t_B^3 + \varepsilon_g^4 \boldsymbol{b}_0^2 t_B^4}}\\
&= \frac{1 - 2\varepsilon_g t_A}{\sqrt{(\boldsymbol{b}_0^2 - 4\boldsymbol{a}\boldsymbol{c}_0)(1 + 4\varepsilon_g \chi_1 \omega_1^{1/2} + 2\varepsilon_g^2 \chi_2 \omega_1) + 4\varepsilon_g^3 \boldsymbol{b}_0^2 t_B^3 + \varepsilon_g^4 \boldsymbol{b}_0^2 t_B^4}}\\
&= \frac{1}{\sqrt{\boldsymbol{b}_0^2 - 4\boldsymbol{a}\boldsymbol{c}_0}}(1 - 2\varepsilon_g \omega_1^{1/2} \tau_A^{1/2})[1 - 2\varepsilon_g \chi_1 \omega_1^{1/2} + \varepsilon_g^2(6\chi_1^2 - \chi_2)\omega_1 + \cdots] \qquad \text{(C.35)}\\
&= \frac{1}{\sqrt{\boldsymbol{b}^2 - 4\boldsymbol{a}\boldsymbol{c}}}[1 - 2\varepsilon_g(\tau_A^{1/2} + \chi_1)\omega_1^{1/2} + \varepsilon_g^2(6\chi_1^2 - \chi_2 + 4\tau_A^{1/2}\chi_1)\omega_1 + O(\omega_1^{3/2})], \qquad \text{(C.36)}
\end{aligned}$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

where

$$\chi_1 = \frac{\boldsymbol{b}_0^2\tau_B^{1/2} + 2\boldsymbol{a}\boldsymbol{c}_0(\tau_A^{1/2} + \tau_C^{1/2})}{\boldsymbol{b}^2 - 4\boldsymbol{a}\boldsymbol{c}}, \quad \chi_2 = \frac{3\boldsymbol{b}_0^2\tau_B - 8\boldsymbol{a}\boldsymbol{c}_0\tau_A^{1/2}\tau_C^{1/2}}{\boldsymbol{b}^2 - 4\boldsymbol{a}\boldsymbol{c}}.$$

In obtaining (C.35), we need $\zeta\varepsilon_g\chi_1\omega_1^{1/2} < 1$, where $\zeta = 4 + 6\varepsilon_g t_B + 4\varepsilon_g^2 t_B^2 + \varepsilon_g^3 t_B^3$. For the expansion in (C.35), it is needed that

$$4\varepsilon_g\chi_1\omega_1^{1/2} + 2\varepsilon_g^2\chi_2\omega_1 + \frac{4\varepsilon_g^3\boldsymbol{b}_0^2 t_B^3}{\boldsymbol{b}^2 - 4\boldsymbol{a}\boldsymbol{c}} + \frac{\varepsilon_g^4\boldsymbol{b}_0^2 t_B^4}{\boldsymbol{b}^2 - 4\boldsymbol{a}\boldsymbol{c}} < 1.$$

However,

$$\frac{2\varepsilon_g^2\chi_2\omega_1 + 4\varepsilon_g^3\boldsymbol{b}_0^2 t_B^3/(\boldsymbol{b}^2 - 4\boldsymbol{a}\boldsymbol{c}) + \varepsilon_g^4\boldsymbol{b}_0^2 t_B^4/(\boldsymbol{b}^2 - 4\boldsymbol{a}\boldsymbol{c})}{4\varepsilon_g\chi_1\omega_1^{1/2}}$$

$$\leqslant \frac{2\varepsilon_g^2 3\boldsymbol{b}_0^2 t_B^2 + 4\varepsilon_g^3\boldsymbol{b}_0^2 t_B^3 + \varepsilon_g^4\boldsymbol{b}_0^2 t_B^4}{4\varepsilon_g\boldsymbol{b}_0^2 t_B} = \frac{\varepsilon_g t_B}{4}(6 + 4\varepsilon_g t_B + \varepsilon_g^2 t_B^2).$$

Using (C.36), we have, for δ_1,

$$\begin{aligned}
\delta_1 &= \frac{\omega_1}{\varsigma_{\omega_1;0}(v_1)} - \frac{\omega_1}{(\boldsymbol{\rho}_g - \boldsymbol{\rho}_{g;-})\boldsymbol{a}} \\
&= \frac{\omega_1}{\sqrt{\boldsymbol{b}^2 - 4\boldsymbol{a}\boldsymbol{c}}}\left[1 - \frac{2\boldsymbol{a}}{\boldsymbol{b}^2 - 4\boldsymbol{a}\boldsymbol{c}}\omega_1 + O(\omega_1^2)\right] \\
&\quad - \frac{\omega_1}{\sqrt{\boldsymbol{b}^2 - 4\boldsymbol{a}\boldsymbol{c}}}\big[1 - 2\varepsilon_g(\tau_A^{1/2} + \chi_1)\omega_1^{1/2} + \varepsilon_g^2(6\chi_1^2 - \chi_2 + 4\tau_A^{1/2}\chi_1)\omega_1 \\
&\quad + O(\omega_1^{3/2})\big] \\
&= \frac{2\varepsilon_g(\tau_A^{1/2} + \chi_1)\omega_1^{3/2}}{\sqrt{\boldsymbol{b}^2 - 4\boldsymbol{a}\boldsymbol{c}}} + O(\omega_1^2).
\end{aligned} \tag{C.37}$$

Now we turn to δ_3. If $\varepsilon_g t_A < 1$, then

$$\begin{aligned}
h(\varepsilon_g, \omega_1) &= 1 - (1 - \varepsilon_g^2)(1 + \varepsilon_g t_A)^{-2} \\
&= 1 - (1 - \varepsilon_g^2)(1 - \varepsilon_g t_A + 2\varepsilon_g^2 t_A^2 - 3\varepsilon_g^3 t_A^3 + \cdots) \\
&= \varepsilon_g^2 + (1 - \varepsilon_g^2)(\varepsilon_g t - 2\varepsilon_g^2 t_A^2 + \cdots) \\
&= \varepsilon_g^2 + \varepsilon_g(1 - \varepsilon_g^2)t_A + O(t_A^2) \\
&= \varepsilon_g^2 + \varepsilon_g(1 - \varepsilon_g^2)\omega_1^{1/2}\tau_A^{1/2} + O(\omega_1),
\end{aligned}$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

$$\begin{aligned} h(\varepsilon_g, \omega_1) &= 1 - (1-\varepsilon_g^2)(1+t_A\varepsilon_g)^{-2} \\ &\geqslant 1 - (1-\varepsilon_g^2) \\ &= \varepsilon_g^2 \geqslant 0. \end{aligned}$$

Therefore

$$\begin{aligned} \delta_3 &= \frac{\omega_1}{(\boldsymbol{\rho} - \boldsymbol{\rho}_{g;-})a} h(\varepsilon_g, \omega_1) \\ &= \frac{\omega_1\varepsilon_g^2 + \varepsilon_g(1-\varepsilon_g^2)\omega_1^{3/2}\tau_A^{1/2}}{(\boldsymbol{\rho} - \boldsymbol{\rho}_{g;-})a} + O(\omega_1^2). \end{aligned} \tag{C.38}$$

We have finished estimating δ_i for $i = 0, 1, 2, 3$. Now, combine (C.31), (C.32), (C.33), (C.37), and (C.38) to get

$$\begin{aligned} \boldsymbol{\rho}_g - \lambda_1^+ &\leqslant \frac{2\varepsilon_g(\tau_A^{1/2} + \chi_1)\omega_1^{3/2}}{\sqrt{\boldsymbol{b}^2 - 4\boldsymbol{ac}}} + \frac{\omega_1\varepsilon_g^2 + \varepsilon_g(1-\varepsilon_g^2)\omega_1^{3/2}\tau_A^{1/2}}{(\boldsymbol{\rho} - \boldsymbol{\rho}_{g;-})\boldsymbol{a}} + O(\omega_1^2) \\ &= \frac{\varepsilon_g^2}{(\boldsymbol{\rho} - \boldsymbol{\rho}_{g;-})\boldsymbol{a}}\omega_1 + \left(\frac{2(\tau_A^{1/2} + \chi_1)}{\sqrt{\boldsymbol{b}^2 - 4\boldsymbol{ac}}} + \frac{(1-\varepsilon_g^2)\tau_A^{1/2}}{(\boldsymbol{\rho} - \boldsymbol{\rho}_{g;-})\boldsymbol{a}}\right)\varepsilon_g\omega_1^{3/2} + O(\omega_1^2), \end{aligned}$$

which, along with

$$\frac{1}{(\boldsymbol{\rho} - \boldsymbol{\rho}_{g;-})\boldsymbol{a}} = \frac{1}{(\boldsymbol{\rho}_g - \boldsymbol{\rho}_{g;-})\boldsymbol{a}} - \frac{\delta_2}{\omega_1} = \frac{1}{\sqrt{\boldsymbol{b}^2 - 4\boldsymbol{ac}}}[1 - 2\varepsilon_g(\tau_A^{1/2} + \chi_1)\omega_1^{1/2}] + O(\omega_1),$$

yields (C.25). Use (C.34) to see that

$$\frac{1}{\varsigma_0(v_1)} = \frac{1}{\varsigma_{\omega_1;0}(v_1)}\left[1 + \frac{2\boldsymbol{a}}{\boldsymbol{b}^2 - 4\boldsymbol{ac}}\omega_1 + O(\omega_1^2)\right],$$

substituting it and (C.14a) into (C.25) to get (C.29). □

We are now ready to prove Theorem 8.2.

Proof of Theorem 8.2. Item (1) is a direct consequence of item (4) of Theorem 8.1.

Item (2) is a consequence of Lemma C.5 upon letting g be the minimizer that gives the minimal ε_{m-1} and using $|\boldsymbol{\rho}_{i+1} - \lambda_\ell^{\text{typ}}| \leqslant |\boldsymbol{\rho}_g - \lambda_\ell^{\text{typ}}|$.

For item (3), again let g be the minimizer that gives the minimal ε_{m-1}. As $i \to \infty$ in item (2), we have $\omega_1 \to 0$, $\omega_2 \to \gamma$, and $v_1 \to u_\ell^{\text{typ}}$ in direction, and thus

$$\lim_{i\to\infty} \eta(v_1) = \lim_{i\to\infty} 3\tau_A^{1/2} + 2\frac{(b_0(v_1))^2\tau_{B_{\lambda_0}}^{1/2} + 2a(v_1)c_0(v_1)(\tau_A^{1/2} + \tau_{C_{\lambda_0}}^{1/2})}{\varsigma_0(v_1)^2} = \eta,$$

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

as given by (8.29). Now let

$$\hat{g}(t) = \mathscr{T}_{m-1}\left(\frac{2t-(\omega_n+\omega_2)}{\omega_n-\omega_2}\right) \Big/ \mathscr{T}_{m-1}\left(-\frac{1+\hat{\kappa}}{1-\hat{\kappa}}\right), \quad \hat{\kappa} = \frac{\omega_2-\omega_1}{\omega_n-\omega_1},$$

where $\mathscr{T}_{m-1}(t)$ is the $(m-1)$th Chebyshev polynomial of the first kind. Then [35, Section 2]

$$\varepsilon_{m-1} \leqslant \varepsilon_{\hat{g}} \leqslant \max_{\omega_2 \leqslant t \leqslant \omega_n} |\hat{g}(t)| = 2\left[\left(\frac{1+\sqrt{\hat{\kappa}}}{1-\sqrt{\hat{\kappa}}}\right)^{m-1} + \left(\frac{1+\sqrt{\hat{\kappa}}}{1-\sqrt{\hat{\kappa}}}\right)^{-(m-1)}\right]^{-1},$$

which goes to ε as $i \to \infty$ because $\hat{\kappa} \to \kappa$. □

References

[1] M. Al-Ammari and F. Tisseur, 'Hermitian matrix polymomials with real eigenvalues of definite type. Part I. Classification', *Linear Algebra Appl.* **436**(10) (2012), 3954–3973.

[2] A. R. Amir-Moéz, 'Extreme properties of eigenvalues of a Hermitian transformation and singular values of the sum and product of linear transformations', *Duke Math. J.* **23** (1956), 463–476.

[3] 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*, 3rd edn (SIAM, Philadelphia, 1999).

[4] Z. Bai, R.-C. Li and Y. Su, Lecture notes on matrix eigenvalue computations. *Prepared for 2009 Summer School on Numerical Linear Algebra*, Chinese Academy of Science, July 2009.

[5] L. Barkwell and P. Lancaster, 'Overdamped and gyroscopic vibrating systems', *J. Appl. Mech.* **59**(1) (1992), 176–181.

[6] T. Betcke, N. J. Higham, V. Mehrmann, C. Schröder and F. Tisseur, 'NLEVP: a collection of nonlinear eigenvalue problems', *ACM Trans. Math. Software* **39**(2) (2013), 7:1–7:28.

[7] R. Bhatia, *Matrix Analysis*, Graduate Texts in Mathematics, 169 (Springer, New York, 1996).

[8] R. Bhatia, F. Kittaneh and R.-C. Li, 'Some inequalities for commutators and an application to spectral variation. II', *Linear Multilinear Algebra* **43**(1–3) (1997), 207–220.

[9] R. Bhatia and R.-C. Li, 'On perturbations of matrix pencils with real spectra. II', *Math. Comp.* **65**(214) (1996), 637–645.

[10] M. T. Chu and S. Xu, 'Spectral decomposition of real symmetric quadratic λ-matrices and its applications', *Math. Comp.* **78** (2009), 293–313.

[11] G. Davis, 'Numerical solution of a quadratic matrix equation', *SIAM J. Sci. Statist. Comput.* **2**(2) (1981), 164–175.

[12] J. Demmel, *Applied Numerical Linear Algebra* (SIAM, Philadelphia, PA, 1997).

[13] R. Duffin, 'A minimax theory for overdamped networks', *Indiana Univ. Math. J.* **4** (1955), 221–233.

[14] D. C. Dzeng and W. W. Lin, 'Homotopy continuation method for the numerical solutions of generalised symmetric eigenvalue problems', *J. Aust. Math. Soc. B* **32** (1991), 437–456, 4.

[15] D. K. Faddeev and V. N. Faddeeva, *Computational Methods of Linear Algebra*, Undergraduate Mathematics Books (W.H. Freeman & Co. Ltd, San Francisco, 1963), Translated by R. C. Williams.

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

[16] K. Fan, 'On a theorem of Weyl concerning eigenvalues of linear transformations. I', *Proc. Natl. Acad. Sci. USA* **35**(11) (1949), 652–655.
[17] I. Gohberg, P. Lancaster and L. Rodman, *Matrix Polynomials Classics in Applied Mathematics,* 58, (SIAM, Philadelphia, 2009).
[18] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd edn (Johns Hopkins University Press, Baltimore, MD, 1996).
[19] G. H. Golub and Q. Ye, 'An inverse free preconditioned Krylov subspace method for symmetric generalized eigenvalue problems', *SIAM J. Sci. Comput.* **24**(1) (2002), 312–334.
[20] A. Greenbaum, *Iterative Methods for Solving Linear Systems* (SIAM, Philadelphia, 1997).
[21] C.-H. Guo, 'Numerical solution of a quadratic eigenvalue problem', *Linear Algebra Appl.* **385**(0) (2004), 391–406.
[22] C.-H. Guo, N. J. Higham and F. Tisseur, 'Detecting and solving hyperbolic quadratic eigenvalue problems', *SIAM J. Matrix Anal. Appl.* **30**(4) (2009), 1593–1613.
[23] C.-H. Guo and P. Lancaster, 'Algorithms for hyperbolic quadratic eigenvalue problems', *Math. Comp.* **74** (2005), 1777–1791.
[24] N. J. Higham and H.-M. Kim, 'Numerical analysis of a quadratic matrix equation', *IMA J. Numer. Anal.* **20**(4) (2000), 499–519.
[25] N. J. Higham, R.-C. Li and F. Tisseur, 'Backward error of polynomial eigenproblems solved by linearization', *SIAM J. Matrix Anal. Appl.* **29**(4) (2007), 1218–1241.
[26] N. J. Higham, D. Mackey and F. Tisseur, 'Definite matrix polynomials and their linearization by definite pencils', *SIAM J. Matrix Anal. Appl.* **31**(2) (2009), 478–502.
[27] N. J. Higham, F. Tisseur and P. M. Van Dooren, 'Detecting a definite Hermitian pair and a hyperbolic or elliptic quadratic eigenvalue problem, and associated nearness problems', *Linear Algebra Appl.* **351–352** (2002), 455–474.
[28] A. V. Knyazev, 'Toward the optimal preconditioned eigensolver: locally optimal block preconditioned conjugate gradient method', *SIAM J. Sci. Comput.* **23**(2) (2001), 517–541.
[29] A. V. Knyazev and K. Neymeyr, 'A geometric theory for preconditioned inverse iteration III: a short and sharp convergence estimate for generalized eigenvalue problems', *Linear Algebra Appl.* **358**(1–3) (2003), 95–114.
[30] J. Kovač-Striko and K. Veselić, 'Trace minimization and definiteness of symmetric pencils', *Linear Algebra Appl.* **216** (1995), 139–158.
[31] P. Lancaster, 'Inverse spectral problems for semisimple damped vibrating systems', *SIAM J. Matrix Anal. Appl.* **29**(1) (2007), 279–301.
[32] P. Lancaster and F. Tisseur, 'Hermitian quadratic matrix polynomials: solvents and inverse problems', *Linear Algebra Appl.* **436**(10) (2012), 4017–4026.
[33] R.-C. Li, 'On perturbations of matrix pencils with real spectra', *Math. Comp.* **62** (1994), 231–265.
[34] R.-C. Li, 'On perturbations of matrix pencils with real spectra, a revisit', *Math. Comp.* **72** (2003), 715–728.
[35] R.-C. Li, 'On Meinardus' examples for the conjugate gradient method', *Math. Comp.* **77**(261) (2008), 335–352. Electronically published on September 17, 2007.
[36] R.-C. Li, Rayleigh quotient based optimization methods for eigenvalue problems. *Technical Report* 2014-04, Department of Mathematics, University of Texas at Arlington, January 2014. Lecture summary for 2013 G. Golub SIAM Summer School; to appear in *Ser. Contemp. Appl. Math*.
[37] R.-C. Li, W.-W. Lin and C.-S. Wang, 'Structured backward error for palindromic polynomial eigenvalue problems', *Numer. Math.* **116**(1) (2010), 95–122.

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

[38] X. Liang and R.-C. Li, 'Extensions of Wielandt's min–max principles for positive semi-definite pencils', *Linear Multilinear Algebra* **62**(8) (2014), 1032–1048.
[39] X. Liang and R.-C. Li, The hyperbolic quadratic eigenvalue problem. *Technical Report* 2014-01, Department of Mathematics, University of Texas at Arlington, January 2014. Available at http://www.uta.edu/math/preprint/.
[40] X. Liang, R.-C. Li and Z. Bai, 'Trace minimization principles for positive semi-definite pencils', *Linear Algebra Appl.* **438** (2013), 3085–3106.
[41] V. B. Lidskii, 'The proper values of the sum and product of symmetric matrices', *Dokl. Akad. Nauk SSSR* **75** (1950), 769–772. (in Russian). Translation by C. Benster available from the National Translation Center of the Library of Congress.
[42] D. E. Longsine and S. F. McCormick, 'Simultaneous Rayleigh-quotient minimization methods for $Ax = \lambda Bx$', *Linear Algebra Appl.* **34** (1980), 195–234.
[43] D. S. Mackey, N. Mackey, C. Mehl and V. Mehrmann, 'Structured polynomial eigenvalue problems: good vibrations from good linearizations', *SIAM J. Matrix Anal. Appl.* **28**(4) (2006), 1029–1051.
[44] D. S. Mackey, N. Mackey, C. Mehl and V. Mehrmann, 'Vector spaces of linearizations for matrix polynomials', *SIAM J. Matrix Anal. Appl.* **28**(4) (2006), 971–1004.
[45] A. S. Markus, *Introduction to the Spectral Theory of Polynomial Operator Pencils*, Translations of Mathematical Monographs, 71 (American Mathematical Society, Providence, RI, 1988).
[46] L. Mirsky, 'Symmetric gauge functions and unitarily invariant norms', *Quart. J. Math.* **11** (1960), 50–59.
[47] C. B. Moler and G. W. Stewart, 'An algorithm for generalized matrix eigenvalue problems', *SIAM J. Numer. Anal.* **10**(2) (1973), 241–256.
[48] J. Nocedal and S. Wright, *Numerical Optimization*, 2nd edn (Springer, New York, 2006).
[49] E. E. Ovtchinnikov, 'Sharp convergence estimates for the preconditioned steepest descent method for Hermitian eigenvalue problems', *SIAM J. Numer. Anal.* **43**(6) (2006), 2668–2689.
[50] B. N. Parlett, *The Symmetric Eigenvalue Problem* (SIAM, Philadelphia, 1998).
[51] B. T. Polyak, *Introduction to Optimization* (Optimization Software, New York, 1987).
[52] E. H. Rogers, 'A mimmax theory for overdamped systems', *Arch. Ration. Mech. Anal.* **16** (1964), 89–96.
[53] E. H. Rogers, 'Variational properties of nonlinear spectra', *Indiana Univ. Math. J.* **18** (1969), 479–490.
[54] Y. Saad, *Iterative Methods for Sparse Linear Systems*, 2nd edn (SIAM, Philadelphia, 2003).
[55] B. Samokish, 'The steepest descent method for an eigenvalue problem with semi-bounded operators', *Izv. Vyssh. Uchebn. Zaved. Mat.* **5** (1958), 105–114. (in Russian).
[56] G. W. Stewart, 'Perturbation bounds for the definite generalized eigenvalue problem', *Linear Algebra Appl.* **23** (1979), 69–86.
[57] G. W. Stewart, *Matrix Algorithms, Vol II: Eigensystems* (SIAM, Philadelphia, 2001).
[58] G. W. Stewart and J.-G. Sun, *Matrix Perturbation Theory* (Academic Press, Boston, 1990).
[59] J.-G. Sun, 'A note on Stewart's theorem for definite matrix pairs', *Linear Algebra Appl.* **48** (1982), 331–339.
[60] J.-G. Sun, 'Perturbation bounds for eigenspaces of a definite matrix pair', *Numer. Math.* **41** (1983), 321–343.
[61] W. Sun and Y.-X. Yuan, *Optimization Theory and Methods—Nonlinear Programming* (Springer, New York, 2006).

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press

[62] I. Takahashi, 'A note on the conjugate gradient method', *Inform. Process. Japan* **5** (1965), 45–49.
[63] F. Tisseur, 'Backward error and condition of polynomial eigenvalue problems', *Linear Algebra Appl.* **309**(1–3) (2000), 339–361.
[64] F. Tisseur and K. Meerbergen, 'The quadratic eigenvalue problem', *SIAM Rev.* **43**(2) (2001), 235–386.
[65] K. Veselić, 'A Jacobi eigenreduction algorithm for definite matrix pairs', *Numer. Math.* **64** (1993), 241–269.
[66] K. Veselić, 'Note on interlacing for hyperbolic quadratic pencils', in *Recent Advances in Operator Theory in Hilbert and Krein Spaces*, (eds. J. Behrndt, K.-H. Förster and C. Trunk) Operator Theory: Advances and Applications, 198 (Birkhäuser, Boston, 2010), 305–307.
[67] K. Veselić, *Damped Oscillations of Linear Systems*, Lecture Notes in Mathematics, 2023 (Springer, Berlin, 2011).
[68] H. Voss, 'A minmax principle for nonlinear eigenproblems depending continuously on the eigenparameter', *Numer. Linear Algebra Appl.* **16**(11–12) (2009), 899–913.
[69] H. Voss and B. Werner, 'A minimax principle for nonlinear eigenvalue problems with applications to nonoverdamped systems', *Math. Methods Appl. Sci.* **4** (1982), 415–424.
[70] S. Wei and I. Kao, 'Vibration analysis of wire and frequency response in the modern wiresaw manufacturing process', *J. Sound Vib.* **231**(5) (2000), 1383–1395.
[71] H. Weyl, 'Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung)', *Math. Ann.* **71** (1912), 441–479.
[72] H. Wielandt, 'An extremum property of sums of eigenvalues', *Proc. Amer. Math. Soc.* **6** (1955), 106–110.
[73] H. Yang, 'Conjugate gradient methods for the Rayleigh quotient minimization of generalized eigenvalue problems', *Computing* **51** (1993), 79–94.

https://doi.org/10.1017/fms.2015.14 Published online by Cambridge University Press