Structured backward error and condition of generalized eigenvalue problems

Higham, Desmond J. and Higham, Nicholas J.

1998

MIMS EPrint: **2006.164**

Manchester Institute for Mathematical Sciences
School of Mathematics

The University of Manchester

ISSN 1749-9097

SIAM J. MATRIX ANAL. APPL.
Vol. 20, No. 2, pp. 493–512

STRUCTURED BACKWARD ERROR AND CONDITION OF GENERALIZED EIGENVALUE PROBLEMS*

DESMOND J. HIGHAM† AND NICHOLAS J. HIGHAM‡

Abstract. Backward errors and condition numbers are defined and evaluated for eigenvalues and eigenvectors of generalized eigenvalue problems. Both normwise and componentwise measures are used. Unstructured problems are considered first, and then the basic definitions are extended so that linear structure in the coefficient matrices (for example, Hermitian, Toeplitz, Hamiltonian, or band structure) is preserved by the perturbations.

Key words. generalized eigenvalue problem, quadratic eigenvalue problem, backward error, condition number, structured matrices

AMS subject classifications. 65F15, 65F35

PII. S0895479896313188

1. Introduction.

1.1. Backward error and condition. Backward errors and condition numbers play an important role in modern numerical linear algebra. Backward errors reveal the stability of a numerical method. Condition numbers explain the sensitivity of the solution of a problem to perturbations in the data, and, in the case where this perturbation is a backward error, the product of condition number times backward error provides a first order error bound for the computed solution. The theory of backward error and conditioning is now well developed for linear systems and least squares problems. For eigenproblems, although there is a large literature on perturbation theory (see [30] and the references therein), a number of aspects of backward errors and condition numbers have not been considered, particularly for the generalized eigenvalue problem. Our aim here is to give a thorough development of backward error and condition for the generalized eigenvalue problem, for both normwise and componentwise measures, with particular emphasis on respecting structure in the coefficient matrices.

We consider the generalized eigenvalue problem $Ax = \lambda Bx$, where $A, B \in \mathbb{C}^{n\times n}$. If $x \neq 0$ then we say that λ is an eigenvalue and x the corresponding eigenvector of the pair (A, B). We develop backward errors for approximate eigenpairs and condition numbers for eigenvalues and eigenvectors. We do not treat deflating subspaces (the appropriate generalization of invariant subspaces), leaving this important topic to future work; for existing results on deflating subspaces see Stewart and Sun [30, Section VI.2.4] and Kågström and Poromaa [21].

In section 2 we develop normwise backward errors and condition numbers for a general class of normwise measures of the perturbations to A and B. In particular, we show that, for the 2-norm, requiring the perturbations to respect Hermitian structure in A and B has no effect on the backward error of an approximate eigenpair corresponding to a real approximate eigenvalue and has no effect on the condition number

*Received by the editors December 3, 1996; accepted for publication (in revised form) by J. Varah February 19, 1998; published electronically November 23, 1998.
http://www.siam.org/journals/simax/20-2/31318.html
†Department of Mathematics, University of Strathclyde, Glasgow G1 1XH, Scotland (na.dhigham@na-net.ornl.gov, http://www.strath.ac.uk/~aas96106/).
‡Department of Mathematics, University of Manchester, Manchester M13 9PL, England (higham@ma.man.ac.uk, http://www.ma.man.ac.uk/~higham/).

of an eigenvalue of a definite pair. It is widely appreciated that for problems with badly scaled or sparse data componentwise analysis can yield much stronger results than normwise analysis [18]. In section 3 we give a treatment analogous to that of section 2 but for componentwise measures.

Many applications lead to eigenproblems containing matrices with linear structure, and only perturbations that preserve the structure may be physically meaningful [31]. For example, the quadratic eigenvalue problem has the form

$$(\lambda^2 C + \lambda D + E)v = 0, \qquad C, D, E \in \mathbb{C}^{n\times n}. \tag{1.1}$$

By writing $u = \lambda v$ we can express the problem as a generalized eigenvalue problem:

$$\begin{bmatrix} D & E \\ E & 0 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \lambda \begin{bmatrix} -C & 0 \\ 0 & E \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}. \tag{1.2}$$

The coefficient matrices are clearly highly structured, and arbitrary perturbations to these matrices do not correspond to perturbations of the original quadratic eigenvalue problem—we must respect the structure in order to obtain meaningful results. Another highly structured eigenproblem is that of a Hamiltonian matrix

$$\mathcal{H} = \begin{bmatrix} F & G \\ H & -F^T \end{bmatrix}, \qquad F \in \mathbb{R}^{n\times n}, \quad G = G^T \in \mathbb{R}^{n\times n}, \quad H = H^T \in \mathbb{R}^{n\times n}.$$

It has been a long-standing open problem to develop numerical methods for the Hamiltonian eigenproblem that require only $O(n^3)$ operations and for which the computed eigenvalues or invariant subspaces are the exact ones of a nearby Hamiltonian matrix; that is, the backward error preserves the Hamiltonian structure. Benner, Mehrmann, and Xu [4] have developed a numerical method that comes close to satisfying these requirements, having a Hamiltonian backward error with respect to a Hamiltonian matrix related to $\mathcal{H}$. For testing this method and deriving error bounds it is therefore useful to have backward errors and condition numbers that respect Hamiltonian structure.

In section 4 we define backward errors and condition numbers that respect arbitrary linear structure in the matrices A and B. The backward error is characterized as the minimum norm solution to a rectangular system, while explicit formulas are obtained for the condition numbers. We have previously carried out similar analysis for linear systems [17]; there we used a Kronecker product-based approach, but here we use a different technique based on "pattern matrices."

Brief numerical experiments are reported in section 5 in order to illustrate the analysis.

1.2. Preliminaries. For the backward error analysis we make no assumptions on A and B, but for the definition and derivation of condition numbers we assume that the pair (A, B) is *regular*, that is, that $\det(A - \lambda B)$ is not identically zero in λ.

Significant advantages accrue from treating the generalized eigenvalue problem in the form $\beta Ax = \alpha Bx$, where an eigenvalue is now defined by a pair of scalars (α, β) [30, p. 272]. For example, when B is singular a nonzero null vector of B is an eigenvector with $(\alpha, \beta) = (\alpha, 0)$, whereas in the original formulation we have an eigenvalue $\lambda = \infty$. The (α, β) framework thus elegantly handles infinite eigenvalues and treats A and B symmetrically. Moreover, perturbation expansions of α and β individually provide complete information about eigenvalue sensitivity [29], [30, p. 293], although

some of this information is inevitably lost in the definition of a single condition number. We have chosen to derive and state all our results in terms of λ for three main reasons. First, λ is the desired quantity in most practical applications of the generalized eigenvalue problem. Second, analysis with the (α, β) form is naturally carried out using the chordal metric, but this metric is scale dependent (see section 2.2). Finally, the standard eigenproblem ($B = I$) is an important special case and we would like our results to be directly applicable to it.

For parts of the analysis we will consider a Hermitian pair (A, B) (that is, a pair in which A and B are Hermitian) and assume that it is *definite*, that is, that

$$\min\left\{ \left((x^*Ax)^2 + (x^*Bx)^2\right)^{1/2} : x \in \mathbb{C}^n,\ \|x\|_2 = 1 \right\}$$

is positive, which is certainly true if A or B is positive definite. A definite pair (A, B) has the property that there is a nonsingular matrix X such that X^*AX and X^*BX are diagonal. For details of the theory of the generalized eigenvalue problem see [16, Sections 7.7, 8.7], [25, Chap. 15], or [30, Chap. 6].

We require some definitions involving norms. The norm $\|\cdot\|^D$ dual to a given vector norm $\|\cdot\|$ on $\mathbb{C}^n$ is defined by

$$\|x\|^D = \max_{w \neq 0} \frac{|w^*x|}{\|w\|},$$

and we say that z is a vector dual to y if

$$z^*y = \|z\|^D \|y\| = 1.$$

The mixed subordinate matrix norm $\|\cdot\|_{\alpha,\beta}$ on $\mathbb{C}^{n\times n}$ is defined by

$$\|A\|_{\alpha,\beta} = \max_{x \neq 0} \frac{\|Ax\|_\beta}{\|x\|_\alpha}.$$

The generality obtained by allowing $\alpha \neq \beta$ causes no complications in the statements and proofs of our results and permits coverage of the potentially useful norms

$$\|A\|_{1,\beta} = \max_j \|A(:,j)\|_\beta, \qquad \|A\|_{\alpha,\infty} = \max_i \|A(i,:)^*\|_\alpha^D,$$

which include the special case $\|A\|_{1,\infty} = \max_{i,j} |a_{ij}|$. We note for later reference that

$$\|xy^*\|_{\alpha,\beta} = \|x\|_\beta \|y\|_\alpha^D.$$

For complex α we define

$$\operatorname{sign}(\alpha) = \begin{cases} \dfrac{\overline{\alpha}}{|\alpha|}, & \alpha \neq 0, \\ 0, & \alpha = 0. \end{cases}$$

The sign of a vector z is defined componentwise as $\operatorname{sign}(z) = (\operatorname{sign}(z_i))$.

2. Normwise analysis.

2.1. Backward errors. We begin by considering backward errors for the generalized eigenvalue problem. The formulas obtained are useful for testing the stability of practical eigensolvers [1], [8], [21].

The normwise backward error of an approximate eigenpair $(\widetilde{x}, \widetilde{\lambda})$ is defined by

$$\eta(\widetilde{x}, \widetilde{\lambda}) := \min\{\, \epsilon : (A+\Delta A)\widetilde{x} = \widetilde{\lambda}(B+\Delta B)\widetilde{x}, \quad \|\Delta A\|_{\alpha,\beta} \le \epsilon \|E\|_{\alpha,\beta}, \quad \|\Delta B\|_{\alpha,\beta} \le \epsilon \|F\|_{\alpha,\beta} \,\}, \tag{2.1}$$

where E and F are given matrices of tolerances. Note that if A and B are real and $\widetilde{\lambda}$ is nonreal, the optimal perturbations in (2.1) may be nonreal. How to restrict to real perturbations is considered at the end of section 4.

The following theorem, which is a straightforward modification of a result of Rigal and Gaches on the normwise backward error for a linear system [26], gives an explicit expression for $\eta(\widetilde{x}, \widetilde{\lambda})$. For the case $\alpha = \beta$, this theorem and the following lemma are given by Frayssé and Toumazou [14].

THEOREM 2.1. *The normwise backward error $\eta(\widetilde{x}, \widetilde{\lambda})$ is given by*

$$\eta(\widetilde{x}, \widetilde{\lambda}) = \frac{\|r\|_\beta}{(\|E\|_{\alpha,\beta} + |\widetilde{\lambda}|\|F\|_{\alpha,\beta})\|\widetilde{x}\|_\alpha}, \tag{2.2}$$

where $r = \widetilde{\lambda}B\widetilde{x} - A\widetilde{x}$.

Proof. It is straightforward to show that the right-hand side of (2.2) is a lower bound for $\eta(\widetilde{x}, \widetilde{\lambda})$. This lower bound is easily seen to be attained for the feasible perturbations

$$\Delta A = \frac{\|E\|_{\alpha,\beta}}{\|E\|_{\alpha,\beta} + |\widetilde{\lambda}|\|F\|_{\alpha,\beta}} r z^*, \qquad \Delta B = -\operatorname{sign}(\widetilde{\lambda}) \frac{\|F\|_{\alpha,\beta}}{\|E\|_{\alpha,\beta} + |\widetilde{\lambda}|\|F\|_{\alpha,\beta}} r z^*, \tag{2.3}$$

where z is a vector dual to $\widetilde{x}$ with respect to the α-norm. ∎

If we are interested only in the approximate eigenvalue $\widetilde{\lambda}$ and are not concerned about $\widetilde{x}$, then a more appropriate measure of backward error may be

$$\eta(\widetilde{\lambda}) := \min_{\widetilde{x} \ne 0} \eta(\widetilde{x}, \widetilde{\lambda}).$$

This quantity has a closed-form expression, as shown by the next result.

LEMMA 2.2. *If $\widetilde{\lambda}$ is not an eigenvalue of the pair (A, B) then*

$$\eta(\widetilde{\lambda}) = \frac{1}{\|(\widetilde{\lambda}B - A)^{-1}\|_{\beta,\alpha} (\|E\|_{\alpha,\beta} + |\widetilde{\lambda}|\|F\|_{\alpha,\beta})}.$$

Proof. The result follows from Theorem 2.1 on using the equality, for nonsingular $C \in \mathbb{C}^{n\times n}$, $\min_{x\ne 0} \|Cx\|_\beta / \|x\|_\alpha = \|C^{-1}\|_{\beta,\alpha}^{-1}$. ∎

Similarly, for a given $\widetilde{x}$ we can consider minimizing the backward error over all $\widetilde{\lambda}$:

$$\eta(\widetilde{x}) = \min_{\widetilde{\lambda}} \eta(\widetilde{x}, \widetilde{\lambda}) = \min_{\widetilde{\lambda}} \frac{\|\widetilde{\lambda}B\widetilde{x} - A\widetilde{x}\|_\beta}{(\|E\|_{\alpha,\beta} + |\widetilde{\lambda}|\|F\|_{\alpha,\beta})\|\widetilde{x}\|_\alpha}. \tag{2.4}$$

This optimization problem appears to be analytically intractable, even for the 2-norm ($\alpha = \beta = 2$). However, we can obtain an upper bound for $\eta(\widetilde{x})$ by choosing $\widetilde{\lambda}$ to minimize the numerator in (2.4) only; for the 2-norm this value of $\widetilde{\lambda}$ is $\widetilde{x}^*B^*A\widetilde{x}/(\widetilde{x}^*B^*B\widetilde{x})$ if $B\widetilde{x} \neq 0$.

If A and B are Hermitian then it is desirable that the perturbations ΔA and ΔB in the definition of backward error preserve this property. Therefore the following backward error is of interest:

$$\eta^H(\widetilde{x},\widetilde{\lambda}) := \min\{\,\epsilon : (A+\Delta A)\widetilde{x} = \widetilde{\lambda}(B+\Delta B)\widetilde{x},\ \ \Delta A = \Delta A^*,\ \ \Delta B = \Delta B^*, \\ \|\Delta A\|_{\alpha,\beta} \le \epsilon\|E\|_{\alpha,\beta},\ \ \|\Delta B\|_{\alpha,\beta} \le \epsilon\|F\|_{\alpha,\beta}\,\}. \tag{2.5}$$

Clearly, $\eta^H(\widetilde{x},\widetilde{\lambda}) \ge \eta(\widetilde{x},\widetilde{\lambda})$, and the optimal perturbations for $\eta(\widetilde{x},\widetilde{\lambda})$ in (2.3) are not Hermitian in general. We wish to determine by how much $\eta^H(\widetilde{x},\widetilde{\lambda})$ can exceed $\eta(\widetilde{x},\widetilde{\lambda})$. The following result, which is an adaptation of a result of Bunch, Demmel, and Van Loan [5] for linear systems, shows that requiring the backward error perturbations to be Hermitian has no effect on the backward error in the 2-norm, provided that $\widetilde{\lambda}$ is real. The same result has also been obtained by Smoktunowicz [28], where it is stated only for definite pairs.

THEOREM 2.3. *If A and B are Hermitian and $\widetilde{\lambda}$ is real, then, for the 2-norm ($\alpha = \beta = 2$), we have $\eta_2^H(\widetilde{x},\widetilde{\lambda}) = \eta_2(\widetilde{x},\widetilde{\lambda})$.*

Proof. Let $r = \widetilde{\lambda}B\widetilde{x} - A\widetilde{x}$. We first find a Hermitian H that satisfies the constraint $H\widetilde{x} := (\Delta A - \widetilde{\lambda}\Delta B)\widetilde{x} = r$ in (2.5). We take $H := (\|r\|_2/\|\widetilde{x}\|_2)P$, where P is a suitably chosen Householder matrix—such a P exists since $\widetilde{\lambda}$ real implies $\widetilde{x}^*r$ real (if $r = \widetilde{x}$, we have to take $H = I$ instead). To satisfy $H = \Delta A - \widetilde{\lambda}\Delta B$ with Hermitian ΔA and ΔB, we define

$$\Delta A = \frac{\|E\|_2}{\|E\|_2 + |\widetilde{\lambda}|\|F\|_2}H, \qquad \Delta B = -\operatorname{sign}(\widetilde{\lambda})\frac{\|F\|_2}{\|E\|_2 + |\widetilde{\lambda}|\|F\|_2}H. \tag{2.6}$$

Now

$$\|H\|_2 = \|r\|_2/\|\widetilde{x}\|_2 = \eta_2(\widetilde{x},\widetilde{\lambda})(\|E\|_2 + |\widetilde{\lambda}|\|F\|_2),$$

using (2.2). From (2.6) it follows that $\eta_2^H(\widetilde{x},\widetilde{\lambda}) \le \eta_2(\widetilde{x},\widetilde{\lambda})$. But $\eta_2^H(\widetilde{x},\widetilde{\lambda}) \ge \eta_2(\widetilde{x},\widetilde{\lambda})$ by definition, so equality must hold. ∎

Note that if B is Hermitian positive definite, the perturbation ΔB that achieves $\eta_2^H(\widetilde{x},\widetilde{\lambda})$ in (2.5) does not necessarily keep $B+\Delta B$ positive definite. However, $B+\Delta B$ certainly will be positive definite if

$$\eta_2^H(\widetilde{x},\widetilde{\lambda})\|F\|_2 = \|\Delta B\|_2 < \lambda_{\min}(B), \tag{2.7}$$

where $\lambda_{\min}$ denotes the smallest eigenvalue.

It is worth pausing to discuss the significance of Theorem 2.3. One way to solve the generalized eigenvalue problem is by the QZ algorithm, which computes the generalized Schur decomposition and is normwise backward stable. If the QZ algorithm is applied to a Hermitian pair (A, B) then it does not preserve Hermitian structure. Theorem 2.3 implies that, nevertheless, each computed eigenpair containing a real eigenvalue is exact for a Hermitian pair that is a slight perturbation of (A, B), with the perturbation being different, in general, for each eigenpair.

Definitions and results analogous to those above hold for the backward error corresponding to an approximate eigenvalue $\widetilde{\lambda}$ and corresponding left eigenvector $\widetilde{y}$ ($\widetilde{y}^*A \approx \widetilde{\lambda}\widetilde{y}^*B$).

Also of interest is the normwise backward error of a triple $(\widetilde{x}, \widetilde{y}, \widetilde{\lambda})$, where $\widetilde{y}$ is an approximate left eigenvector. For the 2-norm, this backward error is defined by

$$\eta(\widetilde{x}, \widetilde{y}, \widetilde{\lambda}) := \min\{\, \epsilon : (A+\Delta A)\widetilde{x} = \widetilde{\lambda}(B+\Delta B)\widetilde{x},\ \ \widetilde{y}^*(A+\Delta A) = \widetilde{\lambda}\widetilde{y}^*(B+\Delta B),$$
$$\|\Delta A\|_2 \le \epsilon\|E\|_2,\ \ \|\Delta B\|_2 \le \epsilon\|F\|_2\,\}.$$

This backward error is evaluated explicitly in the following result.

THEOREM 2.4. *For the 2-norm ($\alpha = \beta = 2$), we have*

$$\eta_2(\widetilde{x}, \widetilde{y}, \widetilde{\lambda}) = \frac{1}{\|E\|_2 + |\widetilde{\lambda}|\|F\|_2} \max\left\{ \frac{\|r\|_2}{\|\widetilde{x}\|_2}, \frac{\|s\|_2}{\|\widetilde{y}\|_2} \right\},$$

where $r = \widetilde{\lambda}B\widetilde{x} - A\widetilde{x}$ and $s^* = \widetilde{\lambda}\widetilde{y}^*B - \widetilde{y}^*A$.

Proof. By taking 2-norms in the equations $r = \Delta A\widetilde{x} - \widetilde{\lambda}\Delta B\widetilde{x}$ and $s^* = \widetilde{y}^*\Delta A - \widetilde{\lambda}\widetilde{y}^*\Delta B$, we find that the claimed expression for $\eta_2(\widetilde{x}, \widetilde{y}, \widetilde{\lambda})$ is certainly a lower bound for it. We must show that this lower bound is attained.

Let $G = \Delta A - \widetilde{\lambda}\Delta B$. Then G satisfies the constraints

$$G\widetilde{x} = r, \quad \widetilde{y}^*G = s^*, \quad \text{and} \quad \widetilde{y}^*r = s^*\widetilde{x}. \tag{2.8}$$

A result of Kahan, Parlett and Jiang [22, Thm. 2′] (see also Saad [27, Thm. 3.10]) shows that the minimum value of $\|G\|_2$ subject to G satisfying the constraints (2.8) is

$$\max\left\{ \frac{\|r\|_2}{\|\widetilde{x}\|_2}, \frac{\|s\|_2}{\|\widetilde{y}\|_2} \right\}.$$

Let G_{opt} be a matrix that achieves this minimum, and define

$$\Delta A = \frac{\|E\|_2}{\|E\|_2 + |\widetilde{\lambda}|\|F\|_2} G_{\mathrm{opt}}, \qquad \Delta B = -\operatorname{sign}(\widetilde{\lambda}) \frac{\|F\|_2}{\|E\|_2 + |\widetilde{\lambda}|\|F\|_2} G_{\mathrm{opt}}.$$

Then $\Delta A - \widetilde{\lambda}\Delta B = G$,

$$\|\Delta A\|_2 = \frac{\|E\|_2}{\|E\|_2 + |\widetilde{\lambda}|\|F\|_2} \max\left\{ \frac{\|r\|_2}{\|\widetilde{x}\|_2}, \frac{\|s\|_2}{\|\widetilde{y}\|_2} \right\},$$

and $\|\Delta B\|_2$ satisfies the analogous equality. We have therefore shown that the lower bound for $\eta_2(\widetilde{x}, \widetilde{y}, \widetilde{\lambda})$ is attained. ∎

We remark that the formula for $\eta_2(\widetilde{x}, \widetilde{y}, \widetilde{\lambda})$ in Theorem 2.4 is the maximum of $\eta_2(\widetilde{x}, \widetilde{\lambda})$ and the analogous backward error for $(\widetilde{y}, \widetilde{\lambda})$.

2.2. Condition numbers. Let λ be a simple, finite, nonzero eigenvalue of the pair (A, B), with corresponding right eigenvector x and left eigenvector y, so that $Ax = \lambda Bx$ and $y^*A = \lambda y^*B$. A normwise condition number of λ can be defined as follows:

$$\kappa(\lambda) := \lim_{\epsilon \to 0} \sup \left\{ \frac{|\Delta\lambda|}{\epsilon|\lambda|} : (A+\Delta A)(x+\Delta x) = (\lambda+\Delta\lambda)(B+\Delta B)(x+\Delta x), \right.$$
$$\left. \|\Delta A\|_{\alpha,\beta} \le \epsilon\|E\|_{\alpha,\beta},\ \ \|\Delta B\|_{\alpha,\beta} \le \epsilon\|F\|_{\alpha,\beta} \right\}. \tag{2.9}$$

This definition is a little loose, because if λ' is an eigenvalue distinct from λ with corresponding eigenvector x', then we can take $\Delta A = \Delta B = 0$, $x + \Delta x \equiv x'$, and $\lambda + \Delta\lambda \equiv \lambda'$ to obtain $\kappa(\lambda) = \infty$. The definition therefore needs to be augmented with the requirement that $\Delta x \to 0$ as $\epsilon \to 0$. For simplicity of presentation we omit this requirement from the definitions of condition numbers.

From the definition of $\kappa(\lambda)$ we have, for the perturbed system in (2.9),

$$\frac{|\Delta\lambda|}{|\lambda|} \le \kappa(\lambda)\epsilon + O(\epsilon^2). \tag{2.10}$$

Expanding the constraint in (2.9) and premultiplying by y^* lead to

$$\begin{aligned}\Delta\lambda &= \frac{y^*\Delta A x - \lambda y^*\Delta B x + y^*\Delta A\Delta x - \lambda y^*\Delta B\Delta x}{y^*Bx + y^*B\Delta x + y^*\Delta B x + y^*\Delta B\Delta x}\\ &= \frac{y^*\Delta A x - \lambda y^*\Delta B x}{y^*Bx} + O(\epsilon^2). \qquad (2.11)\end{aligned}$$

To evaluate $\kappa(\lambda)$ we need to obtain a sharp bound for the first order term in this expansion. The following result is given in [14] for the case $\alpha = \beta$.

THEOREM 2.5. *The normwise condition number $\kappa(\lambda)$ is given by*

$$\kappa(\lambda) = \frac{\|y\|_\beta^D \|x\|_\alpha (\|E\|_{\alpha,\beta} + |\lambda|\|F\|_{\alpha,\beta})}{|\lambda||y^*Bx|}.$$

Proof. The given expression is clearly an upper bound for $\kappa(\lambda)$. We now show that the bound is attained. Let $G = \|x\|_\alpha uv^*$, where u is of unit β-norm and satisfies $u^*y = \|y\|_\beta^D$ and v is dual to x with respect to the α-norm. Then $y^*Gx = \|x\|_\alpha\|y\|_\beta^D$ and $\|G\|_{\alpha,\beta} = 1$. Let $\Delta A = \epsilon\|E\|_{\alpha,\beta}G$ and $\Delta B = -\operatorname{sign}(\lambda)\epsilon\|F\|_{\alpha,\beta}G$. Then $\|\Delta A\|_{\alpha,\beta} \le \epsilon\|E\|_{\alpha,\beta}$ and $\|\Delta B\|_{\alpha,\beta} \le \epsilon\|F\|_{\alpha,\beta}$ and the modulus of the first order term in (2.11) is $\epsilon\|y\|_\beta^D\|x\|_\alpha(\|E\|_{\alpha,\beta} + |\lambda|\|F\|_{\alpha,\beta})/|y^*Bx|$; dividing (2.11) by $\epsilon|\lambda|$ and taking the limit as $\epsilon \to 0$ then gives the desired equality. ∎

If λ is infinite then $\kappa(\lambda)$ is not defined, but one can consider the problem $\lambda^{-1}Ax = Bx$, which has a corresponding zero eigenvalue. For zero eigenvalues $\kappa(\lambda)$ is also undefined, and the absolute condition number (defined as in (2.9) but with $|\Delta\lambda|/\epsilon$ as the quantity to be maximized) is then the appropriate one to consider.

As for the backward error, if A and B are Hermitian it is natural to restrict the perturbations ΔA and ΔB in (2.9) to be Hermitian. The next lemma shows that for a definite pair this has no effect on the condition number in the 2-norm.

LEMMA 2.6. *Let the Hermitian pair (A, B) be definite. Let $\kappa^H(\lambda)$ denote the condition number defined as in* (2.9) *but with the additional requirement that ΔA and ΔB are Hermitian. Then, for the* 2*-norm* ($\alpha = \beta = 2$), $\kappa_2^H(\lambda) = \kappa_2(\lambda)$.

Proof. Since (A, B) is a definite pair, we can take $y = x$. For the 2-norm, the matrix G constructed in the proof of Theorem 2.5 is therefore $G = \|x\|_2^{-2}xx^*$, which is Hermitian. The result is immediate. ∎

It is instructive to compare the condition number $\kappa(\lambda)$ with one of Stewart and Sun [30, pp. 293–294] (cf. [14]). In our notation, they derive the approximate bound (correct to first order)

$$\chi(\lambda, \widetilde{\lambda}) := \frac{|\lambda - \widetilde{\lambda}|}{\sqrt{|\lambda|^2+1}\sqrt{|\widetilde{\lambda}|^2+1}} \lesssim \frac{\|x\|_2\|y\|_2}{\sqrt{|y^*Ax|^2 + |y^*Bx|^2}} \|\,[\,\Delta A \quad \Delta B\,]\,\|_2, \tag{2.12}$$

where $\widetilde{\lambda} = \lambda + \Delta\lambda$, and they regard the first factor in the upper bound as a condition number. The function χ is the chordal metric and has the property that $\chi(\lambda,\mu) = \chi(\lambda^{-1},\mu^{-1})$, which is appropriate for the generalized eigenvalue problem since $Ax = \lambda Bx$ and $\lambda^{-1}Ax = Bx$ are equally valid representations. Note also that the bound (2.12) is symmetric in A and B. Unfortunately, the chordal metric is scale dependent: $\chi(\alpha\lambda,\alpha\mu) \neq \chi(\lambda,\mu)$ in general. Hence $\chi(\lambda,\widetilde{\lambda})$ tends to differ greatly from the relative error when $|\lambda|$ and $|\widetilde{\lambda}|$ are both large or both small.

A normwise condition number for the eigenvector x corresponding to the simple eigenvalue λ can be defined by

$$\kappa(x) := \lim_{\epsilon\to 0}\sup\left\{ \frac{\|\Delta x\|_\alpha}{\epsilon\|x\|_\alpha} : (A+\Delta A)(x+\Delta x) = (\lambda+\Delta\lambda)(B+\Delta B)(x+\Delta x), \right.$$
$$g^*Bx = g^*B(x+\Delta x) \equiv 1,$$
$$\left. \|\Delta A\|_{\alpha,\beta} \le \epsilon\|E\|_{\alpha,\beta},\ \ \|\Delta B\|_{\alpha,\beta} \le \epsilon\|F\|_{\alpha,\beta} \right\}. \tag{2.13}$$

Because an eigenvector corresponding to a simple eigenvalue is unique only up to scalar multiples, it is important to normalize the eigenvectors for the perturbation theory. We use a linear normalization in (2.13) based on a constant vector g, which could, for example, be x or the left eigenvector y. The matrix B is included in the normalization equation because it simplifies the subsequent analysis; if B is nonsingular then we can set $g^* := g^*B^{-1}$ in order to remove B. See Chaitin-Chatelin and Frayssé [6, Section 4.4.2] and Stewart and Sun [30, pp. 240–241] for discussions of the normalization issue.

The next result gives an expression for the condition number; it generalizes a result in [6, Section 4.4.2], [7, Section 4.2.1] that applies to the standard eigenproblem.

THEOREM 2.7. *The normwise condition number $\kappa(x)$ is given by*

$$\kappa(x) = \|V(W^*(A-\lambda B)V)^{-1}W^*\|_{\beta,\alpha}\big(\|E\|_{\alpha,\beta} + |\lambda|\|F\|_{\alpha,\beta}\big), \tag{2.14}$$

where the full rank matrices $V, W \in \mathbb{C}^{n\times(n-1)}$ are chosen so that $g^*BV = 0$ and $W^*Bx = 0$.

Proof. From Theorem A.1 in the appendix we see that we have to find a sharp bound for $\|V(W^*(A-\lambda B)V)^{-1}W^*(\Delta A - \lambda\Delta B)x\|_\alpha/(\epsilon\|x\|_\alpha)$. This quantity is clearly bounded by the claimed expression for $\kappa(x)$. Writing $Z = V(W^*AV - \lambda I)^{-1}W^*$, equality is attained for

$$\Delta A = \epsilon\|E\|_{\alpha,\beta}\|x\|_\alpha ph^*, \qquad \Delta B = -\operatorname{sign}(\lambda)\epsilon\|F\|_{\alpha,\beta}\|x\|_\alpha ph^*,$$

where p satisfies $\|p\|_\beta = 1$ and $\|Zp\|_\alpha = \|Z\|_{\beta,\alpha}$, and h is dual to x with respect to the α-norm. ∎

Note that the expression for $\kappa(x)$ in (2.14) is finite and does not depend on the particular choice of V and W, as shown in the appendix.

As a comparison, we recall that Golub and Van Loan [16, p. 346] and Wilkinson [32, pp. 68–70] give perturbation expansions for the eigenvector corresponding to a simple eigenvalue, in the case $B = I$. These expansions do not readily lead to the identification of a condition number. Indeed, as Wilkinson notes [32, pp. 70, 85], examination of individual terms can give a misleading impression of the sensitivity because of the possibility of cancellation in the overall sum.

We show how Theorem 2.7 leads to an informative bound for $\kappa(x)$ when A is Hermitian, B is Hermitian positive definite, and we take $g = x$. In this case there is a nonsingular X so that

$$X^*BX = I, \qquad X^*AX = D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n).$$

It is easy to verify that we can set $X =: [\, x \quad V \,]$ and $W = V$. Putting $E = A$ and $F = B$ and using the 2-norm, we have

$$\begin{aligned}\kappa_2(x) &= \|V(V^*X^{-*}(D - \lambda I)X^{-1}V)^{-1}V^*\|_2(\|A\|_2 + |\lambda|\|B\|_2) \\ &= \|V \operatorname{diag}(\lambda_2 - \lambda, \ldots, \lambda_n - \lambda)^{-1}V^*\|_2(\|A\|_2 + |\lambda|\|B\|_2) \\ &\le \frac{\|V\|_2^2}{\min_{\lambda_i \neq \lambda} |\lambda - \lambda_i|}(\|A\|_2 + |\lambda|\|B\|_2).\end{aligned}$$

In the case of the standard eigenvalue problem, where $B = I$, we can take $\|V\|_2 = 1$ (and we would take $F = 0$, so that there would be no $|\lambda|\|B\|_2$ term).

3. Componentwise analysis.

3.1. Backward errors. The componentwise backward error of an approximate eigenpair $(\widetilde{x}, \widetilde{\lambda})$ is defined by

$$\omega(\widetilde{x}, \widetilde{\lambda}) := \min\{\, \epsilon : (A + \Delta A)\widetilde{x} = \widetilde{\lambda}(B + \Delta B)\widetilde{x}, \;\; |\Delta A| \le \epsilon E, \;\; |\Delta B| \le \epsilon F \,\}, \tag{3.1}$$

where E and F are now assumed to have nonnegative entries and inequalities hold componentwise.

The Oettli–Prager theorem [24] gives an explicit expression for the componentwise backward error of an approximate solution to a linear system. The next result is an analogue of that result for the eigenvalue problem and is given for the special case $B = I$, $F = 0$ in [10].

THEOREM 3.1. *The componentwise backward error is given by*

$$\omega(\widetilde{x}, \widetilde{\lambda}) = \max_i \frac{|r_i|}{\big((E + |\widetilde{\lambda}|F)|\widetilde{x}|\big)_i}, \tag{3.2}$$

where $r = \widetilde{\lambda}B\widetilde{x} - A\widetilde{x}$, and $\xi/0$ is interpreted as zero if $\xi = 0$ and infinity otherwise.

Proof. It is easy to show that the right-hand side of (3.2) is a lower bound for $\omega(\widetilde{x}, \widetilde{\lambda})$ and that this bound is attained for the feasible perturbations

$$\Delta A = D_1 E D_2, \qquad \Delta B = -\operatorname{sign}(\widetilde{\lambda}) D_1 F D_2, \tag{3.3}$$

where $D_1 = \operatorname{diag}\big(r_i/((E + |\widetilde{\lambda}|F)|\widetilde{x}|)_i\big)$ and $D_2 = \operatorname{diag}\big(\operatorname{sign}(\widetilde{x})\big)$. ∎

As for the normwise backward error, it is also of interest to consider the minima of $\omega(\widetilde{x}, \widetilde{\lambda})$ over all $\widetilde{x}$ and over all $\widetilde{\lambda}$:

$$\omega(\widetilde{\lambda}) = \min_{\widetilde{x} \neq 0} \omega(\widetilde{x}, \widetilde{\lambda}), \qquad \omega(\widetilde{x}) = \min_{\widetilde{\lambda} \neq 0} \omega(\widetilde{x}, \widetilde{\lambda}).$$

We have been unable to derive any useful bounds for $\omega(\widetilde{\lambda})$ and $\omega(\widetilde{x})$.

Drmač [11] derives an algorithm for solving the generalized eigenvalue problem when $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times n}$ are both symmetric positive definite. He shows that the computed eigenvalues are the exact ones of a pair $(A + \Delta A, B + \Delta B)$ satisfying

$$|\Delta a_{ij}| \le f(n)u\sqrt{a_{ii}a_{jj}}, \qquad |\Delta b_{ij}| \le f(n)u\sqrt{b_{ii}b_{jj}},$$

where u is the unit roundoff. To test an implementation of this method it is therefore appropriate, for each computed eigenpair $(\widehat{x}, \widehat{\lambda})$, to compare $\omega(\widehat{x}, \widehat{\lambda})$ with a suitable multiple of u, taking $e_{ij} = \sqrt{a_{ii}a_{jj}}$, $f_{ij} = \sqrt{b_{ii}b_{jj}}$.

3.2. Condition numbers. A componentwise condition number for a simple eigenvalue λ analogous to the normwise condition number (2.9) is defined by

$$\mathrm{cond}(\lambda) := \lim_{\epsilon\to 0}\sup\left\{ \frac{|\Delta\lambda|}{\epsilon|\lambda|} : (A+\Delta A)(x+\Delta x) = (\lambda+\Delta\lambda)(B+\Delta B)(x+\Delta x), \right.$$

$$\left. |\Delta A| \le \epsilon E, \ \ |\Delta B| \le \epsilon F \right\}. \tag{3.4}$$

From this definition it follows that, for the perturbed system in (3.4),

$$\frac{|\Delta\lambda|}{|\lambda|} \le \mathrm{cond}(\lambda)\epsilon + O(\epsilon^2). \tag{3.5}$$

A special case of the following result with $B = I$, $E = |A|$, and $F = 0$ is given by Geurts [15].

THEOREM 3.2. *The componentwise condition number* $\mathrm{cond}(\lambda)$ *is given by*

$$\mathrm{cond}(\lambda) = \frac{|y^*|E|x| + |\lambda||y^*|F|x|}{|\lambda||y^*Bx|}.$$

Proof. The proof is analogous to the proof of Theorem 2.5. The perturbations that are used to show that the expression for $\mathrm{cond}(\lambda)$ is attained are $\Delta A = \epsilon D_1 E D_2$ and $\Delta B = -\mathrm{sign}(\lambda)\epsilon D_1 F D_2$, where $D_1 = \mathrm{diag}(\mathrm{sign}(y^*))$ and $D_2 = \mathrm{diag}(\mathrm{sign}(x))$. ∎

We consider briefly the special case where $A \ge 0$ is irreducible and B is diagonal with positive diagonal entries. Here, the generalized eigenvalue problem is equivalent to the standard eigenvalue problem for $B^{-1}A$. Since $B^{-1}A$ is nonnegative and irreducible, the Perron–Frobenius theory can be applied [20, Thm. 8.4.4] to show that the spectral radius $\rho(B^{-1}A)$ is a positive eigenvalue with corresponding positive left and right eigenvectors. The following special result holds for $\lambda = \rho(B^{-1}A)$; it generalizes a result of Elsner et al. [13, Thm. 1] for the standard eigenproblem.

THEOREM 3.3. *Suppose* $A \ge 0$ *is irreducible and* $B = \mathrm{diag}(b_{ii})$, *with* $b_{ii} > 0$ *for all* i. *Let* $\lambda > 0$ *be the Perron root of* $B^{-1}A$, *assumed to be simple, and let* $E = A$ *and* $F = B$. *Then* $\mathrm{cond}(\lambda) = 2$. *Moreover, if* $\lambda + \Delta\lambda$ *is the Perron root of the pair* $(A + \Delta A, B + \Delta B)$ *defined in* (3.4) *then, for* $0 \le \epsilon < 1$,

$$\frac{|\Delta\lambda|}{|\lambda|} \le \frac{2\epsilon}{1-\epsilon} \tag{3.6}$$

(*which improves on* (3.5) *by quantifying the second order term*).

Proof. That $\mathrm{cond}(\lambda) = 2$ is trivial to verify. For the second part, note that since $|\Delta B| \le \epsilon B$, with B diagonal, and $|\Delta A| \le \epsilon A$,

$$\left(\frac{1-\epsilon}{1+\epsilon}\right) B^{-1}A \le (B+\Delta B)^{-1}(A+\Delta A) \le \left(\frac{1+\epsilon}{1-\epsilon}\right) B^{-1}A.$$

Since $\rho(\cdot)$ is monotone on the nonnegative matrices [20, Cor. 8.1.19],

$$\left(\frac{1-\epsilon}{1+\epsilon}\right)\rho(B^{-1}A) \le \rho\big((B+\Delta B)^{-1}(A+\Delta A)\big) \le \left(\frac{1+\epsilon}{1-\epsilon}\right)\rho(B^{-1}A),$$

which rearranges to give (3.6). □

As in the normwise case, forcing the backward error perturbations to respect Hermitian structure has no effect on the condition number for a definite pair.

LEMMA 3.4. *Let the Hermitian pair (A,B) be definite. Let* $\mathrm{cond}^H(\lambda)$ *denote the condition number defined as in* (3.4) *but with the additional requirement that ΔA and ΔB are Hermitian, and assume that E and F are Hermitian. Then* $\mathrm{cond}^H(\lambda) = \mathrm{cond}(\lambda)$.

Proof. Since (A,B) is a definite pair, we can take $y = x$, so the perturbations ΔA and ΔB constructed in the proof of Theorem 3.2 are Hermitian. The result is immediate. □

We define a componentwise condition number for the eigenvector x corresponding to the simple eigenvalue λ by

$$\mathrm{cond}(x) := \lim_{\epsilon\to 0}\sup\left\{ \frac{\|\Delta x\|_\infty}{\epsilon\|x\|_\infty} : (A+\Delta A)(x+\Delta x) = (\lambda+\Delta\lambda)(B+\Delta B)(x+\Delta x), \right.$$

$$\left. g^*Bx = g^*B(x+\Delta x) \equiv 1,\ \ |\Delta A| \le \epsilon E,\ \ |\Delta B| \le \epsilon F \right\}. \tag{3.7}$$

THEOREM 3.5. *The componentwise condition number* $\mathrm{cond}(x)$ *is given by*

$$\mathrm{cond}(x) = \frac{\| \, |V(W^*(A-\lambda B)V)^{-1}W^*|(E+|\lambda|F)|x| \, \|_\infty}{\|x\|_\infty}, \tag{3.8}$$

where the full rank matrices $V, W \in \mathbb{C}^{n\times(n-1)}$ are chosen so that $g^*BV = 0$ and $W^*Bx = 0$.

Proof. The proof is similar to the proof of Theorem 2.7. The perturbations that give equality are

$$\Delta A = \epsilon D_1 E D_2, \qquad \Delta B = -\,\mathrm{sign}(\lambda)\epsilon D_1 F D_2,$$

where $D_1 = \mathrm{diag}(\xi_j)$, $\xi_j = \mathrm{sign}((V(W^*(A-\lambda B)V)^{-1}W^*)_{kj})$, $D_2 = \mathrm{diag}(\mathrm{sign}(x))$, where the kth component of the vector in the numerator of (3.8) has the largest absolute value. □

In the special case $B = I$, $F = 0$, Theorem 3.5 reduces to a result in [6, Section 4.4.2].

4. Structured backward error and condition number. In sections 2 and 3 we considered a normwise backward error and normwise and componentwise condition numbers that respect Hermitian structure in the generalized eigenvalue problem. There are other structures of interest, such as Toeplitz [3], circulant and Hankel structure (the first of which can be general, symmetric, or Hermitian and all of which can be defined in the point or the block sense), augmented system structure [12], Hamiltonian structure [4], and general banded and sparse patterns. We begin with a simple illustration of the effect of taking account of structure. We consider a standard eigenproblem $Ax = \lambda x$ with a circulant A:

$$A = \begin{bmatrix} a & b \\ b & a \end{bmatrix}.$$

We take $a = b = 1$ and

$$\widetilde{x} = \begin{bmatrix} 1+\epsilon \\ -1 \end{bmatrix}, \quad \widetilde{\lambda} = \epsilon \quad (0 \le \epsilon \le 1),$$

which form an exact eigenpair for $\epsilon = 0$. We have

$$r = \widetilde{\lambda}\widetilde{x} - A\widetilde{x} = \begin{bmatrix} \epsilon^2 \\ -2\epsilon \end{bmatrix}.$$

The normwise backward error for the ∞-norm, with $E = A$, $B = I$, $F = 0$, is $\eta(\widetilde{x}, \widetilde{\lambda}) = \epsilon/(1+\epsilon)$. However, since A is a circulant, we may wish to preserve the circulant structure when we perturb A in the definition of the backward error; indeed, only circulant perturbations may be physically meaningful. Then we seek Δa, Δb such that

$$\begin{bmatrix} 1+\Delta a & 1+\Delta b \\ 1+\Delta b & 1+\Delta a \end{bmatrix} \begin{bmatrix} 1+\epsilon \\ -1 \end{bmatrix} = \epsilon \begin{bmatrix} 1+\epsilon \\ -1 \end{bmatrix}.$$

These two equations in two unknowns have a unique solution, for which $\Delta a = -1 + O(\epsilon)$, $\Delta b = -1$. Hence, if ϵ is small, then while $(\widetilde{x}, \widetilde{\lambda})$ is an exact eigenpair for a (non-circulant) matrix close to A, the nearest circulant of which it is an exact eigenpair is relatively far from A. Thus the permitted structure of backward perturbations can greatly affect the backward error.

In this section we assume that $A \in \mathbb{C}^{n\times n}$ and $B \in \mathbb{C}^{n\times n}$ can be parametrized[1]

$$A = \sum_{i=1}^{t} a_i U_i, \qquad B = \sum_{i=1}^{t} a_i V_i, \tag{4.1}$$

where $U_i \in \mathbb{C}^{n\times n}$ and $V_i \in \mathbb{C}^{n\times n}$ are matrices of constants, typically 0s and 1s, and the complex numbers a_i are independent parameters. Thus we are assuming that A and B are linear functions of a vector of parameters $a = (a_i)$. This formulation allows A and B to share parameters, as is necessary for the quadratic eigenvalue problem application (1.2). If B, for example, does not depend on a_j, then we set $V_j = 0$. The case of no structure corresponds to

$$t = 2n^2, \quad \{U_i\}_{i=1}^{n^2} = \{V_i\}_{i=n^2+1}^{2n^2} = \{e_i e_j^T\}_{i,j=1}^{n}, \quad \begin{cases} U_i = 0, & i > n^2, \\ V_i = 0, & i \le n^2, \end{cases} \tag{4.2}$$

where e_i is the ith unit vector.

Nonlinear structure is also of interest, such as Cauchy and Vandermonde structure. Condition numbers can be derived by linearizing and applying the techniques given here, but evaluating backward errors is a nonlinear optimization problem in general. See Bartels and Higham [2] for results for Vandermonde-like linear systems.

We consider perturbations $\Delta A = \sum_{i=1}^{t} \Delta a_i U_i$ and $\Delta B = \sum_{i=1}^{t} \Delta a_i V_i$ and measure them by

$$\psi_p(\Delta a) = \|D^{-1}\Delta a\|_p, \quad D = \operatorname{diag}(g),$$

[1] In fact, it is only the perturbations ΔA and ΔB that need to have such structure in our development, but it is natural to assume that A and B are structured too.

where g is a vector of nonnegative tolerances and $\|\cdot\|_p$ is the Hölder p-norm on $\mathbb{C}^n$:

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}, \qquad p \geq 1.$$

Note that $p = \infty$ corresponds to the componentwise measure used in (3.1), albeit now applied to the vector of parameters rather than the matrices as a whole.

We define the structured componentwise backward error of an approximate eigenpair $(\widetilde{x}, \widetilde{\lambda})$ by

$$\begin{aligned}\overline{\omega}_p(\widetilde{x}, \widetilde{\lambda}) := \min \Bigg\{ & \psi_p(\Delta a) : (A + \Delta A)\widetilde{x} = \widetilde{\lambda}(B + \Delta B)\widetilde{x}, \\ & \Delta A = \sum_{i=1}^{t} \Delta a_i U_i, \quad \Delta B = \sum_{i=1}^{t} \Delta a_i V_i \Bigg\}. \end{aligned} \tag{4.3}$$

Using the residual $r = \widetilde{\lambda} B\widetilde{x} - A\widetilde{x}$, the constraint in (4.3) can be rewritten

$$\begin{aligned} r &= \Delta A\widetilde{x} - \widetilde{\lambda}\Delta B\widetilde{x} \\ &= \left(\sum_{i=1}^{t} \Delta a_i U_i\right)\widetilde{x} - \widetilde{\lambda}\left(\sum_{i=1}^{t} \Delta a_i V_i\right)\widetilde{x} \\ &= \sum_{i=1}^{t} \Delta a_i (U_i - \widetilde{\lambda} V_i)\widetilde{x} \\ &= \widetilde{C}\Delta a, \end{aligned}$$

where

$$\widetilde{C} = \left[(U_1 - \widetilde{\lambda}V_1)\widetilde{x}, \ldots, (U_t - \widetilde{\lambda}V_t)\widetilde{x}\right] \in \mathbb{C}^{n\times t}.$$

Hence, defining

$$\widetilde{\Delta a} = D^{-1}\Delta a,$$

we have

$$r = \widetilde{C}\Delta a = \widetilde{C}D\widetilde{\Delta a} =: M\widetilde{\Delta a}, \qquad M \in \mathbb{C}^{n\times t}.$$

The linear system

$$M\widetilde{\Delta a} = r \tag{4.4}$$

can be under- or overdetermined, depending on the value of t. The backward error $\overline{\omega}_p(\widetilde{x}, \widetilde{\lambda})$ is the norm of a solution of minimal p-norm to (4.4). If $n > t$ or M is rank-deficient there may be no solution to (4.4), in which case we regard the structured componentwise backward error $\overline{\omega}_p(\widetilde{x}, \widetilde{\lambda})$ as infinite. Assuming that the system is consistent, for the 2-norm we have $\overline{\omega}_2(\widetilde{x}, \widetilde{\lambda}) = \|M^+ r\|_2$, where M^+ is the pseudo-inverse of M.

The structure of the matrix M depends on that of A and B in (4.1), and it may be possible to exploit this structure when solving (4.4). For the unstructured case with

$p = \infty$, we recover the expression (3.2) on using (4.2) and identifying D with E and F in (3.1), because (4.4) then reduces to n independent minimal ∞-norm problems that can be solved explicitly.

A structured componentwise condition number for the simple eigenvalue λ can be defined by

$$\overline{\mathrm{cond}}_p(\lambda) := \lim_{\epsilon\to 0}\sup\left\{ \frac{|\Delta\lambda|}{\epsilon|\lambda|} : (A+\Delta A)(x+\Delta x) = (\lambda+\Delta\lambda)(B+\Delta B)(x+\Delta x), \right.$$

$$\left. \Delta A = \sum_{i=1}^{t} \Delta a_i U_i, \quad \Delta B = \sum_{i=1}^{t} \Delta a_i V_i, \quad \psi_p(\Delta a) \le \epsilon \right\}. \tag{4.5}$$

THEOREM 4.1. *The structured componentwise condition number of λ is given by*

$$\overline{\mathrm{cond}}_p(\lambda) = \frac{\|y^*CD\|_q}{|\lambda||y^*Bx|},$$

where $p^{-1} + q^{-1} = 1$ and

$$C = [(U_1 - \lambda V_1)x, \ldots, (U_t - \lambda V_t)x] \in \mathbb{C}^{n\times t}. \tag{4.6}$$

Proof. The expansion (2.11) shows that we have to find a sharp bound for $y^*\Delta Ax - \lambda y^*\Delta Bx$. For the perturbations ΔA and ΔB in (4.5) we have

$$\begin{aligned} |y^*\Delta Ax - \lambda y^*\Delta Bx| &= \left| y^* \left(\sum_{i=1}^{t} \Delta a_i U_i \right) x - \lambda y^* \left(\sum_{i=1}^{t} \Delta a_i V_i \right) x \right| \\ &= |y^*C\Delta a| \\ &= |y^*CD \cdot D^{-1}\Delta a| \\ &\le \|y^*CD\|_q \psi_p(\Delta a). \end{aligned}$$

Equality is obtained for suitable Δa because equality is always possible in the Hölder inequality. □

It is not hard to see that for $p = \infty$ we recover the formula for the unstructured condition number $\mathrm{cond}(\lambda)$ in Theorem 3.2 when we use (4.2) and identify D with E and F.

A structured componentwise condition number for the eigenvector x corresponding to the simple eigenvalue λ can be defined by

$$\overline{\mathrm{cond}}_p(x) := \lim_{\epsilon\to 0}\sup\left\{ \frac{\|\Sigma^{-1}\Delta x\|_p}{\epsilon} : (A+\Delta A)(x+\Delta x) = (\lambda+\Delta\lambda)(B+\Delta B)(x+\Delta x), \right.$$

$$g^*Bx = g^*B(x+\Delta x) \equiv 1,$$

$$\left. \Delta A = \sum_{i=1}^{t} \Delta a_i U_i, \quad \Delta B = \sum_{i=1}^{t} \Delta a_i V_i, \quad \psi_p(\Delta a) \le \epsilon \right\}, \tag{4.7}$$

where $\Sigma = \mathrm{diag}(\sigma_i)$, with the σ_i positive tolerances. In the particular case $\sigma_i \equiv \|x\|_p$, we are measuring Δx in the usual normwise relative fashion.

THEOREM 4.2. *Assume that B is nonsingular. The structured componentwise condition number of x is given by*

$$\overline{\mathrm{cond}}_p(x) = \left\| \Sigma^{-1}V(W^*(A-\lambda B)V)^{-1}W^*CD \right\|_p,$$

where C is defined in (4.6) *and the full rank matrices $V, W \in \mathbb{C}^{n\times(n-1)}$ are chosen so that $g^*V = 0$ and $W^*x = 0$.*

Proof. From Theorem A.1 we see that we must find a sharp bound for $\|Z(\Delta A - \lambda\Delta B)x\|_p/\epsilon$, where $Z = \Sigma^{-1}V(W^*(A-\lambda B)V)^{-1}W^*$. As in the proof of Theorem 4.1, we have

$$\|Z(\Delta A - \lambda\Delta B)x\|_p = \|ZCD \cdot D^{-1}\Delta a\|_p \le \|ZCD\|_p \psi_p \Delta a,$$

and the latter inequality is attained for suitable Δa. ∎

In the above analysis, the parameter vector Δa is complex in general, although it can be taken to be real when all the data are real. In certain circumstances it is appropriate to restrict Δa to be real, even though the data are complex. For example, a 2×2 Hermitian perturbation ΔA can be parametrized

$$\Delta A = \begin{bmatrix} \Delta a_1 & \Delta a_2 - \Delta a_3 i \\ \Delta a_2 + \Delta a_3 i & \Delta a_4 \end{bmatrix}, \qquad \Delta a_k \in \mathbb{R}, \qquad k = 1\colon 4.$$

The backward error derivation must then be modified by taking real and imaginary parts in (4.4) to obtain a real system, of which a minimal norm solution is required. The quantities derived in Theorems 4.1 and 4.2 are now upper bounds for the condition numbers; the bounds are attained for $p = 1$ and therefore are within a factor t of being attained for other values of p.

5. Numerical experiments. To illustrate our results we present some numerical examples. All computations were carried out in MATLAB, which has unit roundoff $u = 2^{-53} \approx 1.1 \times 10^{-16}$. In each example we used the QZ algorithm [23] to compute the eigensystem. Since MATLAB's implementation of the QZ algorithm does not provide left eigenvectors, we computed left eigenvectors using inverse iteration. Condition numbers were evaluated using the computed eigenvalues and eigenvectors in place of the exact ones.

The first example is the problem with

$$A = \begin{bmatrix} 0.1 & 0.2 \\ 0.3 & 0.4 \end{bmatrix}, \qquad B = \begin{bmatrix} 0.1 & 0.1 \\ 0 & \sqrt{u} \end{bmatrix},$$

which is attributed to Wilkinson by Moler and Stewart [23]. One eigenvalue is close to -2 and the other is of order 10^7. The eigenpair corresponding to the small eigenvalue is found to be well conditioned. For the large eigenvalue and its eigenvector we find that, with $E = A$ and $F = B$ and using the 2-norm,

$$\begin{array}{ll} & \kappa(\lambda) = 1.9 \times 10^7, \quad \operatorname{cond}(\lambda) = 1.0 \times 10^1, \\ \text{for } g = x\colon & \kappa(x) = 8.5 \times 10^{13}, \quad \operatorname{cond}(x) = 6.0 \times 10^{13}, \\ \text{for } g = y\colon & \kappa(x) = 1.0 \times 10^1, \quad \operatorname{cond}(x) = 8.0 \times 10^0. \end{array}$$

This example shows that an eigenvalue may have different sensitivity with respect to normwise and componentwise perturbations and that the sensitivity of an eigenvector can depend strongly on how it is normalized.

The second example concerns symmetric structure in A and B. The matrix A is the `ipjfact` matrix ($a_{ij} = (i + j)!$) from The Test Matrix Toolbox [19] but with its rows and columns in reverse order, and B is the Pascal matrix from the same source; both matrices are positive definite.

TABLE 5.1
Results for symmetric 8×8 A *and* B.

	Backward error	Condition λ	Condition x
Unstructured normwise (η, κ)	2.1e-17	4.2e15	3.5e15
Structured normwise $(\overline{\omega}, \overline{\text{cond}})$	2.1e-17	5.2e15	4.2e15
Unstructured componentwise (ω, cond)	1.5e-8	1.6e6	2.6e6
Structured componentwise $(\overline{\omega}, \overline{\text{cond}})$	3.8e-8	4.4e5	7.7e5

TABLE 5.2
Results for Toeplitz 16×16 A *and* B.

	Backward error	Condition λ	Condition x
Unstructured normwise (η, κ)	2.5e-17	5.2e5	4.6e6
Structured normwise $(\overline{\omega}, \overline{\text{cond}})$	8.3e-17	3.7e5	5.7e6
Unstructured componentwise (ω, cond)	5.4e-14	6.6e1	4.3e2
Structured componentwise $(\overline{\omega}, \overline{\text{cond}})$	4.5e-12	5.6e0	7.0e1

In the third example we have Toeplitz A and B. A Toeplitz matrix $T \in \mathbb{C}^{n\times n}$ can be defined by the sequence $t_{n1}, t_{n-1,1}, \ldots, t_{11}, t_{12}, \ldots, t_{1n}$. With this notation, A is defined by $1, 2^2, \ldots, (2n-1)^2$, and B is defined by $10^{-8}(1, \alpha, \alpha^2, \ldots, \alpha^{2n-1})$, where $\alpha = 10^{8/(2n-1)}$, except that the diagonal of B is multiplied by -1.

In both examples we computed the eigenvalue with smallest real part (which happens to be real in both cases) and the corresponding left and right eigenvectors. The choice of the various parameters is summarized as follows:

- Normalization for eigenvector condition numbers: $g = y$ (left eigenvector).
- Normwise backward error η and condition numbers κ: 2-norm with $E = A$ and $F = B$.
- Componentwise backward error ω and condition numbers cond: $E = |A|$ and $F = |B|$.
- Structured backward error $\overline{\omega}$ and condition number $\overline{\text{cond}}$: $g = \|a\|_2$, giving "normwise" measure, and $g = |a|$, giving "componentwise" measure, using $\Sigma = \text{diag}(\|x\|_p)$ and $p = 2$.

The results are shown in Tables 5.1 and 5.2. Several features are worth noting. For the symmetric problem we see that the componentwise backward error is much larger than the unit roundoff and that requiring symmetry of the backward perturbations has little effect on this backward error (we know the same is true of the normwise backward error by Theorem 2.3). On the other hand, the eigenvalue and eigenvector are much less sensitive to componentwise perturbations than to normwise ones, and again a restriction to symmetric perturbations makes little difference. For the Toeplitz problem, requiring Toeplitz perturbations increases the componentwise backward error by two orders of magnitude and reduces the componentwise condition numbers by an order of magnitude.

In the examples above, the matrix M in (4.4) was always of full rank. In similar experiments with A and B both having symmetric Toeplitz structure we found that M was usually numerically rank-deficient and that the system (4.4) did not always have a solution (more precisely, the minimal-norm least-squares solution did not always have a small residual, so that the structured backward error was sometimes infinite). This behavior appears to be due to symmetries in the eigenvectors. It is known that any eigenvector x of a symmetric Toeplitz matrix with distinct eigenvalues satisfies

TABLE 5.3
Condition numbers for quadratic eigenvalue problem with data (5.1).

	Unstructured	Structured
Normwise	2.0e12	1.0e8
Componentwise	4.0e0	1.0e0

either $Jx = x$ or $Jx = -x$, where J is the identity matrix with its columns in reverse order [9]. It is easy to verify that, for such an x, the matrix $[U_1x, \ldots, U_nx]$ has some repeated rows (up to sign) and so is rank-deficient, where $U_1, \ldots, U_n$ are the pattern matrices corresponding to symmetric Toeplitz structure in an $n \times n$ real matrix. Similar symmetries appear to hold for the generalized eigenvalue problem, causing numerical rank-deficiency of M for "good" approximate eigenvectors $\widetilde{x}$.

Finally, we consider the generalized eigenvalue problem (1.2) corresponding to the quadratic eigenvalue problem (1.1), with

$$C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad E = \begin{bmatrix} 10^{-4} & 1 \\ 0 & 10^{-8} \end{bmatrix}. \tag{5.1}$$

We report in Table 5.3 condition numbers for the eigenvalue $\lambda = -1.0001 \times 10^{-4}$ (to five significant figures). The parameter vector is $a = [\mathrm{vec}(C)^T \, \mathrm{vec}(D)^T \, \mathrm{vec}(E)^T]^T$, where vec stacks the columns of a matrix into one long vector. For the normwise condition number we take $g = [\|C\|_2 e^T, \|D\|_2 e^T, \|E\|_2 e^T]^T$, where $e \in \mathbb{R}^2$ is the vector of 1s, and, for the componentwise condition number, $g = |a|$ and $p = 1$. We see that the normwise condition number for the quadratic eigenvalue problem is four orders of magnitude smaller than the condition number we obtain by ignoring the structure in the corresponding generalized eigenvalue problem. The componentwise condition numbers are much smaller than the normwise ones, and the structured condition number reveals perfect conditioning.

There are many interesting open problems in the areas of establishing the existence of finite structured backward errors and bounding the difference between structured and unstructured backward errors and condition numbers.

Appendix. The following result generalizes analysis of Geurts [15], Chaitin-Chatelin and Frayssé [6, Section 4.4.2], and Chatelin [7, Section 4.2.1] for the standard eigenproblem.

THEOREM A.1. *Let $A, B \in \mathbb{C}^{n\times n}$. Let λ be a simple eigenvalue and x the corresponding eigenvector of the pair (A, B) and let*

$$(A + \Delta A)(x + \Delta x) = (\lambda + \Delta\lambda)(B + \Delta B)(x + \Delta x).$$

Normalize x and $x + \Delta x$ by

$$g^*Bx = g^*B(x + \Delta x) = 1, \tag{A.1}$$

where g is a given vector (it is assumed that $g^*Bx \neq 0$). For sufficiently small ΔA and ΔB there is a unique Δx which can be expressed, to first order, as

$$\Delta x = -V(W^*(A - \lambda B)V)^{-1}W^*(\Delta A - \lambda\Delta B)x, \tag{A.2}$$

where the full rank matrices $V, W \in \mathbb{C}^{n\times(n-1)}$ are chosen so that $g^*BV = 0$ and $W^*Bx = 0$.

Proof. For sufficiently small ΔA and ΔB, $\lambda + \Delta\lambda$ is a simple eigenvalue of the pair $(A + \Delta A, B + \Delta B)$ [30, Thm. VI.2.1], making Δx unique when normalized as in (A.1).

We have

$$\Delta A x + A\Delta x = \lambda \Delta B x + \lambda B \Delta x + \Delta\lambda B x, \tag{A.3}$$

where this and all subsequent equations include first order terms only. Premultiplying by W^* gives

$$W^*(A - \lambda B)\Delta x = -W^*(\Delta A - \lambda \Delta B)x. \tag{A.4}$$

Define the matrices $M = [\, x \quad V \,]$, $N = [\, g \quad W \,]$. We show that M is nonsingular. Note first that

$$g^* B M = [\, g^* B x \quad g^* B V \,] = [\, 1 \quad 0 \,] = e_1^T. \tag{A.5}$$

Hence if $Mt = 0$ then $0 = g^* B M t = t_1$, and then $0 = Mt = Vt(2\colon n)$, which implies $t(2\colon n) = 0$. A similar proof shows that N is nonsingular.

Write $\Delta x = Mz$. Then

$$\begin{aligned} W^*(A - \lambda B)\Delta x &= W^*(A - \lambda B)\,[\, x \quad V \,]\, z \\ &= W^*\,[\, 0 \quad (A - \lambda B)V \,] \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \\ &= W^*(A - \lambda B) V z_2. \end{aligned} \tag{A.6}$$

Now we show that the matrix $W^*(A - \lambda B)V \in \mathbb{C}^{(n-1)\times(n-1)}$ is nonsingular.

We have

$$\begin{aligned} N^*(A - \mu B)M &= \begin{bmatrix} g^* \\ W^* \end{bmatrix} [\, (\lambda - \mu) B x \quad (A - \mu B)V \,] \\ &= \begin{bmatrix} \lambda - \mu & g^*(A - \mu B)V \\ 0 & W^*(A - \mu B)V \end{bmatrix}. \end{aligned}$$

Hence $\det(N^*)\det(A - \mu B)\det(M) = (\lambda - \mu)\det(W^*(A - \mu B)V)$. Since N and M are nonsingular and λ is a simple eigenvalue of (A, B) it follows that $\det(W^*(A - \mu B)V)$ is nonzero for $\mu = \lambda$, as required.

From (A.6) and (A.4) we obtain $z_2 = -(W^*(A - \lambda B)V)^{-1} W^*(\Delta A - \lambda \Delta B)x$. To determine the scalar z_1 we note that, from the normalization condition (A.1) and (A.5),

$$0 = g^* B \Delta x = g^* B M z = e_1^T z = z_1.$$

Hence $\Delta x = V z_2$, as required. ∎

The matrices V and W appearing in (A.2) can be explicitly constructed as follows. Let $Q_v^T B^* g = R_v = \pm \|B^* g\|_2 e_1$ and $Q_w^T B x = R_w = \pm \|Bx\|_2 e_1$ be QR factorizations. Then we can take $V = Q_v(:, 2\colon n)$ and $W = Q_w(:, 2\colon n)$.

Note that the matrix $V(W^*(A - \lambda B)V)^{-1} W^*$ is independent of the particular choice of V and W, since Δx in (A.2) is unique for sufficiently small ΔA and ΔB.

REFERENCES

[1] M. F. Anjos, S. J. Hammarling, and C. C. Paige, *Solving the Generalized Symmetric Eigenvalue Problem*, manuscript, 1992.

[2] S. G. Bartels and D. J. Higham, *The structured sensitivity of Vandermonde-like systems*, Numer. Math., 62 (1992), pp. 17–33.

[3] R. M. Beam and R. F. Warming, *The asymptotic spectra of banded Toeplitz and quasi-Toeplitz matrices*, SIAM J. Sci. Comput., 14 (1993), pp. 971–1006.

[4] P. Benner, V. Mehrmann, and H. Xu, *A New Method for Computing the Stable Invariant Subspace of a Real Hamiltonian Matrix*, Preprint SFB393/97-01, Fak. f. Mathematik, Technische Universität Chemnitz-Zwickau, Chemnitz, FRG, 1997; J. Comp. Math. Appl., to appear.

[5] J. R. Bunch, J. W. Demmel, and C. F. Van Loan, *The strong stability of algorithms for solving symmetric linear systems*, SIAM J. Matrix Anal. Appl., 10 (1989), pp. 494–499.

[6] F. Chaitin-Chatelin and V. Frayssé, *Lectures on Finite Precision Computations*, SIAM, Philadelphia, PA, 1996.

[7] F. Chatelin, *Eigenvalues of Matrices*, Wiley, Chichester, UK, 1993.

[8] J. K. Cullum and R. A. Willoughby, *A QL procedure for computing the eigenvalues of complex symmetric tridiagonal matrices*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 83–109.

[9] G. Cybenko, *On the eigenstructure of Toeplitz matrices*, IEEE Trans. Acoust., Speech, Signal Processing, ASSP-32 (1984), pp. 918–921.

[10] A. S. Deif, *Realistic a priori and a posteriori error bounds for computed eigenvalues*, IMA J. Numer. Anal., 9 (1990), pp. 323–329.

[11] Z. Drmač, *A tangent algorithm for computing the generalized singular value decomposition*, SIAM J. Numer. Anal., 35 (1998), pp. 1804–1832.

[12] H. Elman and D. Silvester, *Fast nonsymmetric iterations and preconditioning for Navier-Stokes equations*, SIAM J. Sci. Comput., 17 (1996), pp. 33–46.

[13] L. Elsner, I. Koltracht, M. Neumann, and D. Xiao, *On accurate computations of the Perron root*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 456–467.

[14] V. Frayssé and V. Toumazou, *A note on the normwise perturbation theory for the regular generalized eigenproblem* $Ax = \lambda Bx$, Numer. Linear Algebra Appl., 5 (1998), pp. 1–10.

[15] A. J. Geurts, *A contribution to the theory of condition*, Numer. Math., 39 (1982), pp. 85–96.

[16] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed., Johns Hopkins University Press, Baltimore, MD, 1989.

[17] D. J. Higham and N. J. Higham, *Backward error and condition of structured linear systems*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 162–175.

[18] N. J. Higham, *A survey of componentwise perturbation theory in numerical linear algebra*, in Mathematics of Computation 1943–1993: A Half Century of Computational Mathematics, W. Gautschi, ed., Proc. Sympos. Appl. Math. 48, AMS, Providence, RI, 1994, pp. 49–77.

[19] N. J. Higham, *The Test Matrix Toolbox for* Matlab *(version* 3.0*)*, Numerical Analysis Report No. 276, Manchester Centre for Computational Mathematics, Manchester, England, 1995.

[20] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, London, 1985.

[21] B. Kågström and P. Poromaa, *Computing eigenspaces with specified eigenvalues of a regular matrix pair* (A, B) *and condition estimation: Theory, algorithms and software*, Numer. Algorithms, 12 (1996), pp. 369–407.

[22] W. Kahan, B. N. Parlett, and E. Jiang, *Residual bounds on approximate eigensystems of nonnormal matrices*, SIAM J. Numer. Anal., 19 (1982), pp. 470–484.

[23] C. B. Moler and G. W. Stewart, *An algorithm for generalized matrix eigenvalue problems*, SIAM J. Numer. Anal., 10 (1973), pp. 241–256.

[24] W. Oettli and W. Prager, *Compatibility of approximate solution of linear equations with given error bounds for coefficients and right-hand sides*, Numer. Math., 6 (1964), pp. 405–409.

[25] B. N. Parlett, *The Symmetric Eigenvalue Problem*, Prentice–Hall, Englewood Cliffs, NJ, 1980. Reprinted by SIAM, Philadelphia, PA, 1998.

[26] J. L. Rigal and J. Gaches, *On the compatibility of a given solution with the data of a linear system*, J. Assoc. Comput. Mach., 14 (1967), pp. 543–548.

[27] Y. Saad, *Numerical Methods for Large Eigenvalue Problems*, Manchester University Press, Manchester, and Halsted Press, New York, 1992.

[28] A. Smoktunowicz, *The Strong Stability of Algorithms for Solving the Symmetric Eigenproblem*, manuscript, 1995.

[29] G. W. Stewart, *Perturbation theory for the generalized eigenvalue problem*, in Recent Advances in Numerical Analysis, C. de Boor and G. H. Golub, eds., Academic Press, New York, 1978, pp. 193–206.
[30] G. W. Stewart and Ji-guang Sun, *Matrix Perturbation Theory*, Academic Press, London, 1990.
[31] P. M. Van Dooren, *Structured linear algebra problems in digital signal processing*, in Numerical Linear Algebra, Digital Signal Processing and Parallel Algorithms, G. H. Golub and P. M. Van Dooren, eds., NATO Adv. Sci. Inst. Ser. F Comput. Systems Sci. 70, Springer-Verlag, Berlin, 1991, pp. 361–384.
[32] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford University Press, London, 1965.