ELSEVIER Linear Algebra and its Applications 358 (2003) 5–22

LINEAR ALGEBRA AND ITS APPLICATIONS

www.elsevier.com/locate/laa

Bounds for eigenvalues of matrix polynomials ☆

Nicholas J. Higham *,[1], Françoise Tisseur

Department of Mathematics, University of Manchester, Manchester M13 9PL, UK

Received 16 February 2001; accepted 5 March 2001

Submitted by B.N. Parlett

Abstract

Upper and lower bounds are derived for the absolute values of the eigenvalues of a matrix polynomial (or λ-matrix). The bounds are based on norms of the coefficient matrices and involve the inverses of the leading and trailing coefficient matrices. They generalize various existing bounds for scalar polynomials and single matrices. A variety of tools are used in the derivations, including block companion matrices, Gershgorin's theorem, the numerical radius, and associated scalar polynomials. Numerical experiments show that the bounds can be surprisingly sharp on practical problems.
© 2002 Elsevier Science Inc. All rights reserved.

AMS classification: 65F15; 15A18

Keywords: Polynomial eigenvalue problem; λ-Matrix; Matrix polynomial; Block companion matrix; Gershgorin's theorem; Numerical radius

1. Introduction

Polynomial root finding is an old subject on which much has been written. In particular, many bounds are available for roots of polynomials, comprehensive surveys being given in [17,21]. When the coefficients of the polynomial are generalized

☆ This work was supported by Engineering and Physical Sciences Research Council grant GR/L76532.
* Corresponding author.

E-mail addresses: higham@ma.man.ac.uk (N.J. Higham), ftisseur@ma.man.ac.uk (F. Tisseur).

URL: http://www.ma.man.ac.uk/~higham/ (N.J. Higham), http://www.ma.man.ac.uk/~ftisseur/ (F. Tisseur).
[1] Supported by a Royal Society Leverhulme Trust Senior Research Fellowship.

0024-3795/02/$ - see front matter © 2002 Elsevier Science Inc. All rights reserved.
PII: S0024-3795(01)00316-0

from scalars to matrices we obtain the polynomial eigenvalue problem, which has also received much attention—see, for example, the books [6,13]. However to our knowledge little or nothing has been published on bounds for eigenvalues of matrix polynomials, except for special classes of coefficient matrices (cf. [13, Chapter 9] and [22]). In this work we derive upper and lower bounds for the absolute values of the eigenvalues of general matrix polynomials, concentrating on bounds that are of practical use. Thus we aim for bounds that can be computed with much less computational effort than is required to solve the eigenproblem, especially for large problems. All our bounds are generalizations, in one way or another, of bounds for the eigenvalues of a single matrix and of bounds for the roots of a scalar polynomial. Our treatment is selective: many other bounds can be derived by generalizing results for the special cases just mentioned, but the bounds we present are probably sufficient for most purposes, especially when combined with scaling, similarity and eigenvalue-reciprocating transformations.

Motivation for developing bounds for the eigenvalues of matrix polynomials is readily found. Information about the location of eigenvalues is valuable when computing them by an iterative method, for example to aid m the choice of shifts [22]. When computing pseudospectra of matrix polynomials, which provide information about the global sensitivity of the eigenvalues [8,23], a particular region of the (possibly extended) complex plane must be identified that contains the eigenvalues of interest, and bounds clearly help to determine such region.

To set the notation, we consider the matrix polynomial (or λ-matrix)

$$P(\lambda) = \lambda^m A_m + \lambda^{m-1} A_{m-1} + \cdots + A_0, \tag{1.1}$$

where $A_k \in \mathbb{C}^{n\times n}$, $k = 0 : m$. The polynomial eigenvalue problem is to find an eigenvalue λ and corresponding nonzero eigenvector x satisfying $P(\lambda)x = 0$. If A_m is singular, then P has an infinite eigenvalue, while if A_0 is singular, then 0 is an eigenvalue. Therefore all our upper bounds on $|\lambda|$ require A_m to be nonsingular and the lower bounds require A_0 to be nonsingular; we will not repeatedly state these nonsingularity conditions as they are usually clear from the context. Note that it is of interest to bound the largest finite eigenvalue or smallest nonzero eigenvalue, but to do so requires more sophisticated and computationally expensive estimates than the matrix norm-based ones that we employ here.

To simplify the exposition, we introduce two new matrix polynomials associated with $P(\lambda)$:

$$P_{\mathrm{U}}(\lambda) = \lambda^m I + \lambda^{m-1} U_{m-1} + \cdots + U_0, \tag{1.2}$$

where $U_i = A_m^{-1} A_i$, so that $P(\lambda) = A_m P_{\mathrm{U}}(\lambda)$, and

$$P_{\mathrm{L}}(\lambda) = \lambda^m I + \lambda^{m-1} L_1 + \cdots + L_m, \tag{1.3}$$

where $L_1 = A_0^{-1} A_i$ so that $\lambda^m P(\lambda^{-1}) = A_0 P_{\mathrm{L}}(\lambda)$. The polynomials P and P_{U} have the same eigenvalues, whereas the eigenvalues of P_{L} are the reciprocals of the eigenvalues of P. The tow polynomials P_{U} and P_{L} are monic polynomials whose eigenvalues are easily shown to be the eigenvalues of the $mn \times mn$ block companion matrices

$$C_\mathrm{U} \equiv \begin{bmatrix} 0 & I & 0 & \cdots & 0 \\ \vdots & 0 & I & & \vdots \\ \vdots & & 0 & \ddots & \vdots \\ 0 & & & \ddots & I \\ -U_0 & -U_1 & \cdots & \cdots & -U_{m-1} \end{bmatrix} \tag{1.4}$$

and

$$C_\mathrm{L} \equiv \begin{bmatrix} 0 & I & 0 & \cdots & 0 \\ \vdots & 0 & I & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & & & \ddots & I \\ -L_m & -L_{m-1} & \cdots & \cdots & -L_1 \end{bmatrix}, \tag{1.5}$$

respectively. The matrices C_U and C_L are exploited in Section 2, in which we derive bounds for the eigenvalues of P from the eigenvalues and singular values of C_U and C_L. In Section 3 we use the roots of an associated scalar polynomial to bound the eigenvalues of P, by generalizing a bound of Cauchy. Some additional bounds are given in Section 4, including generalizations of two bounds of Mohammad. Numerical experiments presented in Section 5 show that the bounds can be surprisingly sharp, even on practical problems. Conclusions are given in Section 6.

Throughout this paper $\|\cdot\|$ denotes a subordinate matrix norm. Also, we write

$$U = [U_0, U_1, \ldots, U_{m-1}], \quad L = [L_m, L_{m-1}, \ldots, L_1]. \tag{1.6}$$

2. Bounds from the block companion matrix

In this section we bound the eigenvalues of P by applying various eigenvalue bounds to the two companion matrices C_U and C_L and exploiting their block structure.

2.1. *Bounds based on norms of* C_U *and* C_L

A basic tool is the following standard result.

Lemma 2.1. *Every eigenvalue* λ *of* $A \in \mathbb{C}^{n\times n}$ *satisfies* $|\lambda| \leqslant \|A\|$ *for any matrix norm.*

An application of Lemma 2.1 gives the following bound. Here, $\|\cdot\|_p$ denotes a matrix norm subordinate to a vector p-norm.

Lemma 2.2. *Every eigenvalue of* λ *of* P *satisfies*

$$\left(1+\sum_{j=1}^{m}\|L_j\|_p\right)^{-1} \leqslant |\lambda| \leqslant 1+\sum_{j=-0}^{m-1}\|U_j\|_p, \quad 1\leqslant p\leqslant\infty.$$

Proof. Write C_{U} in the form $\sum_{j=0}^{m-1} f_{m,j+1}(U_j)+V$, where $f_{ij}(U)$ has (i, j) block U and is otherwise zero and where V is the upper triangular part of C_{U}. Now take norms to obtain the upper bound. The proof of the lower bound is similar. □

Stronger bounds can be obtained for $p=1,2,\infty$ by exploiting the properties of the norms. Recall that U and L are defined in (1.6).

Lemma 2.3. *Every eigenvalue* λ *of* P *satisfies*

$$\begin{aligned}&\max\left(\|L_m\|_1, 1+\max_{i=1:m-1}\|L_i\|_1\right)^{-1}\\ &\quad\leqslant |\lambda| \leqslant \max\left(\|U_0\|_1, 1+\max_{i=1:m-1}\|U_i\|_1\right),\end{aligned} \tag{2.1}$$

$$\max(1,\|L\|_\infty)^{-1} \leqslant |\lambda| \leqslant \max(1,\|U\|_\infty), \tag{2.2}$$

$$\|I+LL^*\|_2^{-1/2} \leqslant |\lambda| \leqslant \|I+U\,U^*\|_2^{1/2}. \tag{2.3}$$

Proof. The first two bounds are obtained by using the explicit formulae for the 1 and ∞ norms, respectively. For the 2-norm we write

$$C_{\mathrm{U}} = \begin{bmatrix} 0 & I & & \\ & 0 & \ddots & \\ & & \ddots & I \\ & & & 0 \end{bmatrix} + \begin{bmatrix} 0 & \dots & 0 \\ \vdots & & \vdots \\ -U_0 & \dots & -U_{m-1} \end{bmatrix} \equiv X+Y \tag{2.4}$$

and note that $X^*Y=Y^*X=0$. Thus

$$\begin{aligned}\|C_{\mathrm{U}}\|_2^2 &= \|C_{\mathrm{U}}^*C_{\mathrm{U}}\|_2 = \|(X+Y)^*(X+Y)\|_2\\ &= \|X^*X+Y^*Y\|_2\\ &\leqslant \|I+Y^*Y\|_2 = \|I+YY^*\|_2\\ &= \|I+UU^*\|_2,\end{aligned}$$

as required. □

An alternative derivation of (2.2) and (2.1) is to apply Gershgorin's theorem [9, Theorem. 6.1.1] to C_{L} and C_{U} and their transposes, respectively. The result in the following corollary are weaker, but they directly generalize bounds for the roots of scalar polynomials summarized in [9, p. 316 ff.], [17, Section 27] and [21, Chapter II].

Corollary 2.4. *Every eigenvalue* λ *of* P *satisfies*

$$\big(1+\max_{i=1:m}\|L_i\|_1\big)^{-1} \leqslant |\lambda| \leqslant 1+\max_{i=0:m-1}\|U_i\|_1, \tag{2.5}$$

$$\max\left(1,\sum_{j=1}^{m}\|L_j\|_\infty\right)^{-1} \leqslant |\lambda| \leqslant \max\left(1,\sum_{j=0}^{m-1}\|U_j\|_\infty\right), \tag{2.6}$$

$$\left(1+\sum_{j=1}^{m}\|L_j\|_2^2\right)^{-1/2} \leqslant |\lambda| \leqslant \left(1+\sum_{j=0}^{m-1}\|U_j\|_2^2\right)^{1/2}. \tag{2.7}$$

For $n = 1$, the upper bound in (2.5) is Cauchy's bound, that in (2.6) is Montel's bound, and that in (2.7) is Carmichael and Mason's bound.

A weakness of Lemma 2.3 and Corollary 2.4 is that the upper bounds are all at least 1 and the lower bounds are all at most 1, irrespective of the norms of the U_i and L_i. This property stems from the fact that the eigenvalues of a matrix are invariant under similarity transformations, while norms are not. However, we can apply a similarity transformation to C_{U} and C_{L} before taking their norms and thereby obtain different and potentially smaller bounds. A natural choice of similarity is the diagonal matrix $X = \mathrm{diag}(\alpha_1 I,\dots,\alpha_m I)$ where the α_i are positive parameters. We have

$$\begin{aligned} C_{\mathrm{U}}(\alpha) &= XC_{\mathrm{U}}X^{-1}\\ &= \begin{bmatrix} 0 & \frac{\alpha_1}{\alpha_2}I & 0 & \cdots & 0\\ \vdots & 0 & \frac{\alpha_2}{\alpha_3}I & & \vdots\\ \vdots & & 0 & \ddots & \vdots\\ 0 & & & \ddots & \frac{\alpha_{m-1}}{\alpha_m}I\\ -\frac{\alpha_m}{\alpha_1}U_0 & -\frac{\alpha_m}{\alpha_2}U_1 & \cdots & -\frac{\alpha_m}{\alpha_{m-1}}U_{m-2} & -U_{m-1} \end{bmatrix}. \end{aligned} \tag{2.8}$$

Adapting the proof of Lemma 2.3 to $C_{\mathrm{U}}(\alpha)$ leads to the next result. To save clutter we do not state the corresponding lower bounds obtained using the analogous $C_{\mathrm{L}}(\alpha)$.

Lemma 2.5. *Let* α_i, $i = 0:m$, *be positive with* $\alpha_m = 1$. *Every eigenvalue* λ *of* P *satisfies*

$$|\lambda| \leqslant \max\left(\frac{\|U_0\|}{\alpha_1}, \max_{i=1:m-1}\left(\frac{\alpha_i}{\alpha_{i+1}}+\frac{\|U_i\|_1}{\alpha_{i+1}}\right)\right), \tag{2.9}$$

$$|\lambda| \leqslant \max\left(\max_{i=1:m-1}\frac{\alpha_i}{\alpha_{i+1}}\left\|\left[\frac{U_0}{\alpha_1},\dots,\frac{U_{m-2}}{\alpha_{m-1}},U_{m-1}\right]\right\|_\infty\right), \tag{2.10}$$

$$|\lambda| \leqslant \left\| \sum_{i=1}^{m-1} \left(\frac{\alpha_i}{\alpha_{i+1}}\right)^2 + \sum_{i=0}^{m-1} \left(\frac{U_i}{\alpha_{i+1}}\right)^2 \right\|_2^{1/2}. \tag{2.11}$$

For $\alpha_i = \|U_i\|_1$, (2.9) yields

$$|\lambda| \leqslant \max\left(\frac{\|U_0\|_1}{\|U_1\|_1}, 2 \max_{i=1:m-1} \frac{\|U_i\|_1}{\|U_{i+1}\|_1}\right), \tag{2.12}$$

which, for $n = 1$, is Kojima's bound [9, p. 319], and $\alpha_i = \|U_i\|_\infty$, (2.10) yields

$$|\lambda| \leqslant \sum_{i=0}^{m-1} \frac{\|U_i\|_\infty}{\|U_{i+1}\|_\infty}. \tag{2.13}$$

Finally, we state a bound potentially smaller than (2.1) and (2.2). For a matrix $A \in \mathbb{C}^{n\times n}$ we define the row and column sums

$$s_i(A) = \sum_{j=1}^{n} |a_{ij}|, \quad t_j(A) = \sum_{i=1}^{n} |a_{ij}|.$$

The following lemma is a straightforward application of Ostrowski's theorem (cf. [9, Theorem 6.4.1] and [16, p. 151, (4)]) to C_{U} and C_{L}. (Ostrowski's theorem is a generalization of Gershgorin's theorem involving both row and column sums.)

Lemma 2.6. *Every eigenvalue* λ *of* P *satisfies*

$$\left(\max_{i=1:mn} s_i(C_{\mathrm{L}})^\beta t_i(C_{\mathrm{L}})^{1-\beta}\right)^{-1} \leqslant |\lambda| \leqslant \max_{i=1:mn} s_i(C_{\mathrm{U}})^\beta t_i(C_{\mathrm{U}})^{1-\beta} \tag{2.14}$$

for every $\beta \in [0, 1]$.

2.2. *Bounds from the singular values of* C_{U} *and* C_{L}

Most of the eigenvalue bounds above are based on bounds for the norms of the block companion matrices C_{U} and C_{L}. For the 1- and ∞-norms we can, of course, evaluate $\|C_{\mathrm{U}}\|$ and $\|C_{\mathrm{L}}\|$ exactly with little computational expense, as is done in (2.1) and (2.2). The cost of evaluating the 2-norm of an $mn \times mn$ matrix is usually prohibitive. We now investigate the singular values of C_{U}, with the aim of simplifying the task of evaluating or bounding $\|C_{\mathrm{U}}\|_2$. As usual, the singular values σ_i of $A \in \mathbb{C}^{n\times n}$ are ordered $\sigma_1 \geqslant \cdots \geqslant \sigma_n \geqslant 0$. The following result generalizes expressions for the singular values of the companion matrix of a scalar polynomial ($n = 1$) derived in [11,12].

Lemma 2.7. *The singular values* σ_i *of the companion matrix* C_{U} *fall into three groups*:

(i) $\sigma_i \leqslant 1, \quad i = 1:n,$
(ii) $\sigma_1 = 1, \quad i = n+1 : n(m-1) \quad$ (*if* $m \geqslant 3$),
(iii) $\sigma_i \geqslant 1, \quad i = n(m-1)+1 : nm \quad$ (*if* $m \geqslant 2$).

The $2n$ *singular values in groups* (i) *and* (iii) *are the square roots of the eigenvalues of the quadratic* λ*-matrix* $Q(\lambda) = \lambda^2 I - \lambda(UU^* + I) + U_0U_0^*$.

Proof. The singular values of C_{U} are the square roots of the eigenvalues of $C_{\mathrm{U}}C_{\mathrm{U}}^*$, so we consider

$$S(\lambda) = C_{\mathrm{U}}C_{\mathrm{U}}^* - \lambda I = \begin{bmatrix} (1-\lambda)I & 0 & \cdots & 0 & -U_1^* \\ 0 & (1-\lambda)I & \cdots & 0 & -U_2^* \\ \vdots & 0 & \ddots & \vdots & \vdots \\ \vdots & \vdots & & (1-\lambda)I & -U_{m-1}^* \\ -U_1 & -U_2 & \cdots & -U_{m-1} & UU^* - \lambda I \end{bmatrix}.$$

We need to evaluate $\det(S(\lambda))$, which we do with the aid of block Gaussian elimination. Premultiplying by lower triangular matrices that eliminate the $(m,1)$, $(m,2)$, $\ldots$, $(m, m-1)$ block entries in turn leads to

$$L(\lambda)S(\lambda) = \begin{bmatrix} (1-\lambda)I & 0 & \cdots & 0 & -U_1^* \\ 0 & (1-\lambda)I & \cdots & 0 & -U_2^* \\ \vdots & 0 & \ddots & \vdots & \vdots \\ \vdots & \vdots & & (1-\lambda)I & -U_{m-1}^* \\ 0 & 0 & \cdots & 0 & UU^* - \lambda I - (1-\lambda)^{-1}\sum_{i=1}^{m-1} U_iU_i^* \end{bmatrix},$$

where $L(\lambda)$ is unit lower triangular. Taking determinants gives

$$\begin{aligned} \det(S(\lambda)) &= (1-\lambda)^{n(m-1)} \det\left((UU^* - \lambda I) - (1-\lambda)^{-1}\sum_{i=1}^{m-1} U_iU_i^*\right) \\ &= (1-\lambda)^{n(m-2)} \det\left((1-\lambda)(UU^* - \lambda I) - \sum_{i=1}^{m-1} U_iU_i^*\right) \\ &= (1-\lambda)^{n(m-2)} \det\left(\lambda^2 I - \lambda(UU^* + I) + U_0U_0^*\right). \end{aligned} \tag{2.15}$$

We now know that there are $n(m-2)$ singular values equal to 1 and it remains to show that the remaining singular values fall into two groups of n, one bounded above by 1 and one bounded below by 1. The latter relations follow by applying the Cauchy interlace theorem (cf. [9, Theorem 4.3.15] and [20, Theorem 10.1.1]) to the matrix $C_{\mathrm{U}}C_{\mathrm{U}}^* = S(0)$. □

We now consider the quadratic λ-matrix $Q(\lambda)$ arising in Lemma 2.7. Recall that a quadratic $\lambda^2 A + \lambda B + C$ is *hyperbolic* [14] if A is Hermitian positive definite, B and C are Hermitian, and

$$(x^*Bx)^2 > 4(x^*Ax)(x^*Cx) \quad \text{for all } x \neq 0. \tag{2.16}$$

For $Q(\lambda)$, condition (2.16) requires the positivity of, for $\|x\|_2 = 1$,

$$\begin{aligned}
&(x^*(UU^* + I)x)^2 - 4x^*U_0U_0^*x \\
&= \left(1 + \sum_{i=0}^{m-1} x^*U_iU_i^*x\right)^2 - 4x^*U_0U_0^*x \\
&\geqslant \left(1 - x^*U_0U_0^*x\right)^2 + \sum_{i=1}^{m-1} \left(x^*U_iU_i^*x\right)^2.
\end{aligned}$$

We conclude that $Q(\lambda)$ is hyperbolic as long as at least one of $U_1, \ldots, U_{m-1}$ is nonsingular. The reason for verifying the hyperbolicity property is that it is known that for hyperbolic quadratics all $2n$ eigenvalues are real and there is a gap between the n largest and the n smallest [13, Section 7.6]. We therefore deduce the following result.

Lemma 2.8. *Let $m \geqslant 2$ and suppose that at least one of $U_1, \ldots, U_{m-1}$ is nonsingular. Then in Lemma* 2.7 *strict inequality holds throughout in atleast one of* (i) *and* (iii).

Lemma 2.7 implies the following bounds on the absolute values of the eigenvalues of P.

Corollary 2.9. *For any eigenvalues λ of P,*

$$\lambda_{\min}(Q(\lambda))^{1/2} = \sigma_{mn}(C_{\mathrm{U}}) \leqslant |\lambda| \leqslant \sigma_1(C_{\mathrm{U}}) = \lambda_{\max}(Q(\lambda))^{1/2}. \tag{2.17}$$

Proof. The eigenvalues of P are the eigenvalues of C_{U}, which are bounded in absolute value above and below by the largest and smallest singular values of C_{U}, respectively. The result then follows from Lemma 2.7. □

By their derivation, the bounds of Corollary 2.9 are sharper than those in (2.3), though of course they are more expensive to compute. A significant feature of the corollary is that it bounds the moduli of the eigenvalues of a general λ-matrix of degree m in terms of the eigenvalues of a Hermitian λ-matrix Q of degree 2 that is hyperbolic.

Finally, we note that obvious analogues of the results of this section hold for C_{L}.

2.3. Bound based on the numerical radius

An alternative to using a norm of $A \in \mathbb{C}^{n\times n}$ to bound the eigenvalues is to use the numerical radius ζ:

$$|\lambda| \leqslant \max\left\{|z^*Az| : z \in \mathbb{C}^n,\ \|z\|_2 = 1\right\} =: \zeta(A).$$

The inequality $\zeta(A) \leqslant \|A\|_2$ is immediate and it can be shown that $\zeta(A) \leqslant \|A\|_2/2$ [9, p. 331]; thus employing the numerical radius rather than norms can lead to an improvement, although by a limited amount. We will not attempt to evaluate $\zeta(C_{\mathrm{U}})$, but instead use the splitting (2.4) to obtain a slightly larger but easily expressed upper bound.

We need the following lemma.

Lemma 2.10. *Let $A, B \in \mathbb{C}^{m\times n}$ and let $z \in \mathbb{C}^n$ have unit 2-norm. Then*

$$|z^*BA^*z| \leqslant \frac{\|A^*B\|_2 + \|A\|_2\|B\|_2}{2}.$$

Proof. We have

$$\begin{aligned}\|A\|_2\|B\|_2 &\geqslant \|A^*(2zz^* - I)B\|_2\\ &= \|2A^*zz^*B - A^*B\|_2\\ &\geqslant 2\|A^*zz^*B\|_2 - \|A^*B\|_2\\ &= 2\|A^*z\|_2\|B^*z\|_2 - \|A^*B\|_2\\ &\geqslant 2|z^*BA^*z| - \|A^*B\|_2,\end{aligned}$$

which gives the result on rearranging. □

The following lemma generalizes a result for scalar polynomials in [5].

Lemma 2.11. *Every eigenvalue λ of P satisfies*

$$|\lambda| \leqslant \cos\left(\frac{\pi}{m+1}\right) + \frac{\|U_{m-1}\|_2 + \|U\|_2}{2}. \tag{2.18}$$

Proof. Using $|\lambda| \leqslant \zeta(C_{\mathrm{U}})$ and the splitting (2.4) we have

$$|\lambda| \leqslant \max\left\{|z^*Xz| : z^*z = 1\right\} + \max\left\{|z^*Yz| : z^*z = 1\right\}.$$

The first term can be shown to be $\cos(\pi/(m+1))$ [10, Problem 9, p. 25]. The second term is

$$\max\left\{|z^*[0 \quad \cdots \quad 0 \quad I]^*[U_0 \quad \cdots \quad U_{m-1}]z : z^*z = 1\right\},$$

which can be bounded using Lemma 2.10, to give the result. □

We see that, since $\|U\|_2 = \|UU^*\|_2^{1/2}$, (2.18) can be up to about a factor 2 smaller than (2.3) when U_{m-1} has much smaller norm than all the other U_i and at least one U_i has norm much bigger than 1.

3. Extension of Cauchy's theorem

Another approach is to use norms of the coefficient matrices of P to define a scalar polynomial whose roots provide information about the eigenvalues of P. The following result generalizes a result of Cauchy described in [17, Section 27] and [25, p. 209].

Lemma 3.1. *Assume that A_m and A_0 are nonsingular and define the two scalar polynomials associated with $P(\lambda)$*

$$u(\lambda) = \lambda^m \|A_m^{-1}\|^{-1} - \lambda^{m-1}\|A_{m-1}\| - \cdots - \|A_0\|,$$
$$\ell(\lambda) = \lambda^m \|A_m\| + \cdots + \lambda\|A_1\| - \|A_0^{-1}\|^{-1}.$$

Then every eigenvalue λ of P satisfies

$$r \leqslant |\lambda| \leqslant R, \tag{3.1}$$

where R and r are the unique positive real roots of $u(\lambda)$ and $\ell(\lambda)$, respectively.

Proof. By Descarte's rule of signs, $u(\lambda)$ and $\ell(\lambda)$ have unique positive real root (cf. [21, Theorem 5.3] and [25, p. 197]). Write

$$P(\lambda)x = \lambda^m\big(A_m x + \lambda^{-1}A_{m-1}x + \cdots + \lambda^{-m}A_0 x\big),$$

where $\|x\| = 1$. For $|\lambda| > R$ we have, using $u(R) = 0$;

$$\begin{aligned}&\|\lambda^{-1}A_{m-1}x + \cdots + \lambda^{-m}A_0 x\|\\ &\quad < R^{-1}\|A_{m-1}\| + \cdots + R^{-m}\|A_0\| = \|A_m^{-1}\|^{-1}.\end{aligned} \tag{3.2}$$

Now

$$\begin{aligned}\|P(\lambda)x\| &\geqslant |\lambda|^m\big(\|A_m x\| - \|\lambda^{-1}A_{m-1}x + \cdots + \lambda^{-m}A_0 x\|\big)\\ &\geqslant |\lambda|^m\big(\|A_m^{-1}\|^{-1} - \|\lambda^{-1}A_{m-1}x + \cdots + \lambda^{-m}A_0 x\|\big)\\ &> 0\end{aligned}$$

by (3.2), and it follows that λ is not an eigenvalue of P. Therefore all the eigenvalues satisfy $|\lambda| \leqslant R$. The proof of the lower bound is similar. □

Which choice of norm gives the tightest bounds in (3.1) is difficult to predict. Note that in contrast to all the earlier bounds these bounds are in terms of the original A_i and not $U_i = A_m^{-1}A_i$ or $L_i = A_0^{-1}A_i$.

By applying Lemma 3.1 and Corollary 2.9 we can obtain lower and upper bounds for the moduli of the eigenvalues in terms of the extremal singular values of U_0.

Lemma 3.2. *Let $\beta = \|UU^* + I\|_2$. Every eigenvalue λ of P satisfies*

$$\frac{1}{2}\left(-\beta + \sqrt{\beta^2 + 4\sigma_n(U_0)^2}\right) \leqslant |\lambda|^2 \leqslant \frac{1}{2}\left(\beta + \sqrt{\beta^2 + 4\sigma_1(U_0)^2}\right).$$

Proof. By Corollary 2.9 we have to bound the smallest and largest eigenvalues of $Q(\lambda) = \lambda^2 I - \lambda(UU^* + I) + U_0 U_0^*$. The upper and lower bounds are obtained by applying Lemma 3.1 for the 2-norm. □

Note that we can apply Lemma 3.1 to the polynomial P_{U} and P_{L} in (1.2) and (1.3) instead of P; in general we will obtain different bounds that can be better or worse, depending on the A_i.

The roots r and R in Lemma 3.1 can of course themselves be bounded by applying any of the explicit bounds from this paper with $n = 1$.

4. Other bounds

In this section we give three further types of bounds of a different flavour from those before.

4.1. Bounds from the characteristic polynomial

Potentially useful bounds for the eigenvalues can be obtained from the characteristic polynomial, $\det(P(\lambda))$. It is easily seen that if A_m is nonsingular, then the eigenvalues λ_i of P satisfy

$$(-1)^{mn}\prod_{i=1}^{mn}\lambda_i = \det\left(A_m^{-1}A_0\right) = \frac{\det(A_0)}{\det(A_m)},$$

$$-\sum_{i=1}^{mn}\lambda_i = \operatorname{trace}\left(A_m^{-1}A_{m-1}\right).$$

Hence

$$\min_i |\lambda_i| \leqslant \left|\frac{\det(A_0)}{\det(A_m)}\right|^{1/mn} \leqslant \max_i |\lambda_i|,$$

$$\max_i |\lambda_i| \geqslant \frac{\left|\operatorname{trace}\left(A_m^{-1}A_{m-1}\right)\right|}{mn}.$$

We note that if A_0, A_{m-1} and A_m are symmetric positive definite, then these bounds can be estimated relatively cheaply using Gaussian quadrature and Monte Carlo techniques from [2,3].

4.2. Bound involving fractional powers of norms

Using a variation on the proof of Lemma 3.1 we obtain the following generalization of a bound of Mohammad [18].

Lemma 4.1. *Every eigenvalue of P satisfies*

$$\left(1 + \|A_0^{-1}\|\right)^{-1} \min_{i=1:m} \|A_i\|^{-1/i} \leqslant |\lambda| \leqslant \left(1 + \|A_m^{-1}\|\right) \max_{i=0:m-1} \|A_i\|^{1/(m-i)}.$$

Proof. Let $\theta = \max_{i=0:m-1} \|A_i\|^{1/(m-i)}$. For any x of norm 1 we have

$$\begin{aligned}
\|P(\lambda)x\| &\geqslant |\lambda^m|\left(\|A_m x\| - \|(\lambda^{-1}A_{m-1} + \cdots + \lambda^{-m}A_0)x\|\right) \\
&\geqslant |\lambda^m| \left(\|A_m^{-1}\|^{-1} - \sum_{i=0}^{m-1} \frac{\|A_i\|}{|\lambda|^{m-i}} \right) \\
&= |\lambda^m| \left(\|A_m^{-1}\|^{-1} - \sum_{i=1}^{m} \frac{\theta^i}{|\lambda|^i} \right) \\
&\geqslant |\lambda^m| \left(\|A_m^{-1}\|^{-1} - \sum_{i=1}^{\infty} \frac{\theta^i}{|\lambda|^i} \right) \\
&\geqslant |\lambda^m| \left(\|A_m^{-1}\|^{-1} - \frac{\theta}{|\lambda| - \theta} \right) \\
&> 0 \quad \text{if } |\lambda| > \left(1 + \|A_m^{-1}\|\right)\theta.
\end{aligned}$$

Hence every eigenvalue λ must satisfy the upper bound of the theorem. The lower bound is proved similarly. □

Note that if $A_0 = A_1 = \cdots = A_{m-1} = 0$, then the upper bound in Lemma 4.1 correctly implies that all the eigenvalues are zero; most of the bounds above do not lead to this conclusion. On the other hand, unlike the eigenvalues, the bounds of the lemma are not invariant under the transformations $A_i \leftarrow \beta A_i$, so the bounds are scale-dependent.

4.3. Bound involving maximization over unit circle

Our final bound involves a little more computation, although still at the scalar level. This result generalizes one of Mohammad [19] for scalar polynomials.

Lemma 4.2. *Every eigenvalue of P satisfies* $|\lambda| \leqslant \max(\mu\|A_m^{-1}\|, 1)$, *where*

$$\mu = \max_{|z|=1} \|z^{m-1}A_{m-1} + \cdots + A_0\| = \max_{|z|=1} \|A_{m-1} + \cdots + z^{m-1}A_0\|.$$

Proof. Let x be an arbitrary vector of unit norm. Writing $w = z^{-1}$, we have

$$\begin{aligned}\|P(z)x\| &\geqslant |z^m|\big(\|A_m x\| - \|(wA_{m-1} + \cdots + w^m A_0)x\|\big)\\ &\geqslant |z^m|\big(\|A_m^{-1}\|^{-1} - \|wA_{m-1} + \cdots + w^m A_0\|\big).\end{aligned}$$

For $|w| \leqslant 1$,

$$\begin{aligned}\|wA_{m-1} + \cdots + w^m A_0\| &= |w|\|A_{m-1} + \cdots + w^{m-1}A_0\|\\ &\leqslant |w| \max_{|z|\leqslant 1} \|A_{m-1} + \cdots + z^{m-1}A_0\|\\ &= |w|\mu,\end{aligned}$$

where the last equality follows from a version of the maximum modulus principle. Thus

$$\|P(z)x\| \geqslant |z^m|\big(\|A_m^{-1}\|^{-1} - \mu|w|\big).$$

This lower bound is nonzero for $|w| < \|A_m^{-1}\|^{-1}/\mu$, that is, $|z| > \mu\|A_m^{-1}\|$, and hence any eigenvalue of P exceeding 1 in modulus must satisfy $|\lambda| \leqslant \mu\|A_m^{-1}\|$. $\square$

In the case where $A_m = I$, Lemma 4.2 reduces to $|\lambda| \leqslant \max(1, \mu)$, which clearly can be smaller than (2.2), for example. Indeed, consider $P(\lambda) = \lambda^m I - m^{-1}\lambda^{m-1}I - \cdots - m^{-1}I$, which clearly has I as an eigenvalue. For any subordinate matrix norm, $\mu = m$, and the bound of Lemma 4.2 is 1, whereas the bounds in (2.1)–(2.3) are $1 + 1/m$, 2, and $\sqrt{1 + 1/m}$, respectively.

In general, the task of computing μ is a 1-dimensional maximization over the unit circle.

5. Numerical experiments

Many variations on the bounds explicitly stated here are possible. For example, all the bounds in Section 3 and 4 can be applied to P_{U} in (1.2) and P_{L} in (1.3) as well as to P itself, and those bounds based on the block companion matrices can be applied to a diagonally scaled matrix, as in (2.8). It is therefore not possible to give here a full comparison of all the bounds; instead, we give some illuminating numerical examples. The experiments were performed using MATLAB 6.

As a first example, we consider a 5×5 matrix polynomial $P(\lambda)$ of degree $m = 9$ whose coefficient matrices are of the form

$$A_i = 10^{i-3}\texttt{randn}, \quad i = 0:8; \qquad A_9 = \texttt{randn},$$

Table 1
Upper bounds for first example, for which $\max_i |\lambda_i| = 1.01 \times 10^6$

Bound	Value	Comment
(2.3)	1.79×10^6	2-Norm based
(2.13)	2.82×10^6	∞-Norm based
(2.14)	1.94×10^6	Ostrowski, $\beta = 3/4$
(2.18)	1.78×10^6	Numerical radius-based
(3.1)	2.45×10^6	Cauchy applied to P, 2-norm
(3.1)	1.78×10^6	Cauchy applied to P_{U}, 2-norm
Lemma 4.1	2.74×10^6	2-Norm
Lemma 4.2	1.92×10^6	Applied to P_{U}, 2-norm

Table 2
Lower bounds for first example, for which $\min_i |\lambda_i| = 3.90 \times 10^{-2}$

Bound	Value	Comment
(2.12)	9.97×10^{-3}	1-Norm based
(2.14)	1.11×10^{-2}	Ostrowski applied to $C_{\mathrm{L}}(\alpha)$ with $\alpha_i = \|L_{m+1-i}\|_2$, $\beta = 1/4$
(3.1)	1.76×10^{-3}	Cauchy applied to P_{U}, 2-norm

where `randn` denotes a random matrix from the normal (0, 1) distribution. The minimal and maximal moduli of the 45 eigenvalues are

$$\min_i |\lambda_i| = 3.90 \times 10^{-2}, \quad \max_i |\lambda_i| = 1.01 \times 10^6.$$

All the upper bounds are of the correct order of magnitude, and some of them provide sharp estimates, as shown in Table 1. The lower bounds are of more variable quality, but again several of them are good estimates; see Table 2.

For our second example we consider the free vibration of a string clamped at both ends in a spatially inhomogeneous environment. The equation characterizing the wave motion can be described by

$$\begin{cases} u_{tt} + \epsilon a(x) u_t = \Delta u, & x \in [0, \pi], \epsilon > 0, \\ u(t, 0) = u(t, \pi) = 0. \end{cases}$$

Approximating

$$u(x, t) = \sum_{k=1}^{n} q_k(t) \sin(kx)$$

and applying the Galerkin method we obtain a second-order differential equation

$$M\ddot{q}(t) + \epsilon C\dot{q}(t) + Kq(t) = 0, \tag{5.1}$$

where

$$q(t) = [q_1(t), \dots, q_n(t)]^{\mathrm{T}}, \quad M = (\pi/2) I_n, \quad K = (\pi/2) \mathrm{diag}(j^2),$$

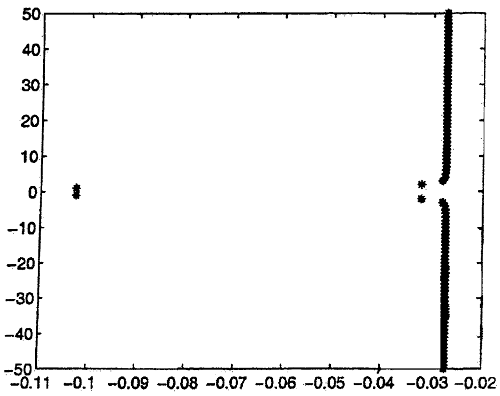

Fig. 1. Spectrum of $P(\lambda)$ for example based on (5.1).

and

$$C = (c_{kj}), \quad c_{kj} = \int_0^\pi a(x)\sin(kx)\sin(jx)\,\mathrm{d}x.$$

The polynomial of interest is $Q(\lambda) = \lambda^2 M + \lambda C + K$. In our experiments we take $n = 50$, $a(x) = x^2(\pi - x)^2 - \delta$, $\delta = 2.7$ and $\epsilon = 0.1$. Since M and K are diagonal, we make the estimation problem harder by multiplying Q on the left and the right by random orthogonal matrices. The spectrum is plotted in Fig. 1 and it satisfies

$$\min_i |\lambda_i| = 1.00, \quad \max_i |\lambda_i| = 50.0.$$

The upper bounds in (2.1)–(2.3), (2.12), (2.13) and (2.18) all exceed 1250. However, the Ostrowski bound (2.14) with $\beta = 1/2$ is 119 when applied to C_{U} and 110 when applied to $C_{\mathrm{U}}(\alpha)$ for $\alpha_i = \|U_i\|_1$, while the Cauchy bound (3.1) for the 2-norm is the remarkably sharp 50.2 (or 234 for the 1-norm).

For the lower bounds, the 2-norm bound (2.3) gives 0.70 and the Cauchy bound (3.1) gives 0.84 for the 2-norm. Even better are the bound 0.88 given by both (2.17), based on the 2-norm of C_{L}, and the numerical radius-based bound (2.18) applied to P_{L}.

Our final example is from a structural dynamics model representing a reinforced concrete machine foundation [4]. It is a sparse quadratic eigenvalue problem $Q(\lambda) = \lambda^2 M + \lambda C + K$ of dimension 3627 with complex symmetric C and K. The matrices M and C are diagonal and the sparsity pattern of K is shown in Fig. 2. To compute $\max_i |\lambda_i|$ we converted the problem to a generalized eigenvalue problem $Ax = \lambda Bx$ with a Hermitian positive definite B and used MATLAB's `eigs` function (an interface to the ARPACK package [15]) to compute the five eigenvalues of largest absolute value. This computation took 233 seconds on a 500 MHz Pentium III machine, yielding $\max_i |\lambda_i| = 2.12 \times 10^4$. For this problem, $\|K\| \gg \|C\| \gg \|M\|$ and $\|U_0\| \gg \|U_1\|$, and this causes the bound (2.1) to be a severe overestimate at

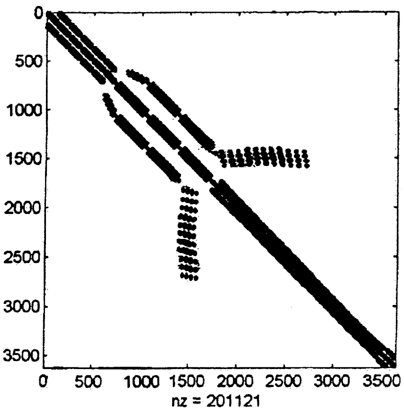

Fig. 2. Sparsity pattern of K for third example.

1.10×10^9. However, (2.12) yields 2.74×10^5 and (2.13) yields 2.22×10^5. The Cauchy bounds, applied to P_{U}, are even sharper: (3.1) yields 3.53×10^4 for the 1-norm and 3.17×10^4 for the ∞-norm. Each of these bounds is computed in less than half a second.

6. Conclusions

With the growing interest in the numerical solution of polynomial eigenvalue problems [1,24] the derivation of eigenvalue bounds is a timely topic of investigation. All our bounds are essentially norm-based, and therein lies a weakness, because even in the special case of a single matrix a norm can differ by an arbitrary amount from the eigenvalues. Nevertheless, our numerical experience shows that our bounds are often surprisingly good estimates, on practical as well as contrived problems. A reason for working with norms is that they can usually be computed or estimated, even for applications invlolving *very* large, sparse matrices, which are perhaps defined only implicitly as long as matrix–vector products can be computed then norms can be estimated [7].

All our bounds involve the inverse of the leading or trailing coefficient matrix. This is inevitable, since, for example, any upper bound can be finite only if A_m is nonsingular, and so computing an upper bound requires at least as much work as testing A_m for nonsingularity. In some applications A_m is diagonal, and in others it has structure (for example, diagonal dominance) that enables $\|A_m^{-1}\|$ to be bounded without computing A_m^{-1}, in which case our bounds are still appicable with minor modification.

We noted earlier that a huge variety of bounds can be obtained by combining the various techniques described here and generalizing other bounds from the scalar case. Exhaustively cataloguing bounds is not the most useful avenue of future research, but identifying bounds suited to particular classes of problems is an important topic on which progress would be valuable in applications.

References

[1] Z. Bai, J.W. Demmel, J.J. Dongarra, A. Ruhe, H.A. van der Vorst (Eds.), Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2000, xxix+410 pp., ISBN 0-89871-471-0.

[2] Z. Bai, M. Fahey, G.H. Golub, Some large-scale matrix computation problems, J. Comput. Appl. Math. 74 (1996) 71–89.

[3] Z. Bai, G.H. Golub, Bounds for the trace of the inverse and the determinant of symmetric positive definite matrices, Ann. Number. Math. 4 (1997) 29–38.

[4] A. Feriani, F. Perotti, V. Simoncini, Iterative system solvers for the frequency analysis of linear mechanical systems, Comput. Methods Appl. Mech. Eng. 19 (13–14) (2000) 1719–1739.

[5] M. Fujii, F. Kubo, Buzano's inequality and bounds for roots of algebraic equations, Proc. Amer. Math. Soc. 117 (2) (1993) 359–361.

[6] I. Gohberg, P. Lancaster, L. Rodman, Matrix Polynomials, Academic Press, New York, 1982, xiv+409 pp., ISBN 0-12-287160-X.

[7] N.J. Higham, F. Tisseur, A block algorithm for matrix 1-norm estimation, with an application to 1-norm pseudospectra, SIAM J. Matrix Anal. Appl. 21 (4) (2000) 1185–1201.

[8] N.J. Higham, F. Tisseur, More on pseudospectra for polynomial eigenvalue problems and applications in control theory, Numerical Analysis Report 372, Manchester Centre for Computational Mathematics, Manchester, UK, January 2001.

[9] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, 1985, xiii+561 pp., ISBN 0-521-30586-1.

[10] R.A. Horn, C.R. Johnson, Topics in Matrix Analysis, Cambridge University Press, 1991, viii+607 pp., ISBN 0-521-30587-X.

[11] C. Kenney, A.J. Laub, Controllability and stability radii for companion form systems, Math. Control Signals Systems 1 (1988) 239–256.

[12] F. Kittaneh, Singular values of companion matrices and bounds on zeros of polynomials, SIAM J. Matrix Anal. Appl. 16 (1) (1995) 333–340.

[13] P. Lancaster, Lambda Matrices and Vibrating Systems, Pergamon Press, Oxford, 1966, xiii+196 pp.

[14] P. Lancaster, Quadratic eigenvalue problems, Linear Algebra Appl. 150 (1991) 499–506.

[15] R.B. Lehoucq, D.C. Sorensen, C. Yang, ARPACK Users' Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1998, xv+142 pp., ISBN 0-89871-407-9.

[16] M. Marcus, H. Minc, A Survey of Matrix Theory and Matrix Inequalities, Prindle, Weber and Schmidt, Boston, 1964. Reprinted by Dover, New York, 1992, ISBN 0-486-67102-X.

[17] M. Marden, Geometry of Polynomials, Mathematical Surveys, vol. 3, second ed., AMS, Providence, RI, USA, 1966, xiii+243 pp.

[18] Q.G. Mohammad, On the zeros of polynomials, Amer. Math. Monthly 72 (1) (1965) 35–38.

[19] Q.G. Mohammad, On the zeros of polynomials, Amer. Math. Monthly 72 (6) (1965) 631–633.

[20] B.N. Parlett, The Symmetric Eigenvalue Problem, Society for Industrial and Applied Mathematics, Phiadelphia, PA, USA, 1998, xxiv+398 pp., Unabridged, amended version of book first published by Prentice-Hall in 1980, ISBN 0-89871-402-8.

[21] Bl. Sendov, A. Andreev, N. Kjurkchiev, Numerical solution of polynomial equations, in: P.G. Ciarlet, J.L. Lions (Eds.), Handbook of Numerical Analysis, volume III: Solution of Equations in $\mathbb{R}^n$ (Part 2), Elsevier, Amsterdam, 1994, pp. 625–778.

[22] V. Simoncini, F. Perotti, On the numerical solution of $(\lambda^2 A + \lambda B + C)x = b$ and appication to structural dynamics, Technical Report 1197, Istituto di Analisi Numerica del CNR, Pavia, Italy, December 2000.
[23] F. Tisseur, N.J. Higham, Strucutred pseudospectra for polynomial eigenvalue problems with applications, Numerical Analysis Report No. 359, Manchester Centre for Computational Mathematics, Manchester, UK, April 2000. To appear in SIAM J. Matrix Anal. Appl.
[24] F. Tisseur, K. Meerbergen, The quadratic eigenvalue problem, SIAM Rev. 43 (2) (2001).
[25] D.M. Young, R.T. Gregory, A survey of Numerical Mathematics, vol. 1, Addison-Wesley, Reading, MA, USA, 1972. x+492+appendices pp. Reprinted by Dover, New York, 1988, ISBN 0-486-65691-8.