IMA Journal of Numerical Analysis (2003) **23**, 359–375

Robust stability and a criss-cross algorithm for pseudospectra

J. V. BURKE†
Department of Mathematics, University of Washington, Seattle, WA 98195, USA

A. S. LEWIS‡
Department of Mathematics, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

AND

M. L. OVERTON§
Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, USA

[Received on 22 October 2002; revised on 23 December 2002]

A dynamical system $\dot{x} = Ax$ is robustly stable when all eigenvalues of complex matrices within a given distance of the square matrix A lie in the left half-plane. The 'pseudospectral abscissa', which is the largest real part of such an eigenvalue, measures the robust stability of A. We present an algorithm for computing the pseudospectral abscissa, prove global and local quadratic convergence, and discuss numerical implementation. As with analogous methods for calculating $\mathbf{H}_\infty$ norms, our algorithm depends on computing the eigenvalues of associated Hamiltonian matrices.

Keywords: pseudospectrum; eigenvalue optimization; robust control; stability; spectral abscissa; $\mathbf{H}_\infty$ norm; robust optimization; Hamiltonian matrix.

1. Introduction

The spectrum of a square matrix provides crucial information about the asymptotic behaviour of associated dynamical systems. Consider a fixed n-by-n real or complex matrix A, and denote the *spectral abscissa* of A (the largest of the real parts of the eigenvalues) by α. Then trajectories $z(t)$ in $\mathbf{R}^n$ or $\mathbf{C}^n$ satisfying $\dot{z} = Az$ converge to the origin faster than $e^{\beta t}$ (for given real $\beta < 0$) if and only if $\alpha < \beta$. In other words, the spectral abscissa of a matrix measures the asymptotic rate of decay of associated trajectories.

Unfortunately however, the spectrum alone may be misleading as an indicator of dynamic properties associated with A. Even when $\alpha < 0$ and so the corresponding dynamical system is asymptotically stable, if complex matrices close to A have eigenvalues with positive real parts, the trajectories $z(t)$ may exhibit large transient peaks before converging to the origin, and may be unstable under small nonlinearities or forcing terms in the dynamical system. By examining the eigenvalues of all complex matrices in given

†Email: burke@math.washington.edu
‡Email: aslewis@sfu.ca
§Email: overton@cs.nyu.edu

© The Institute of Mathematics and its Applications 2003

neighbourhoods of A (the so-called *pseudospectra*), we obtain a more robust indication of stability. For a recent survey on pseudospectra, including a discussion of some of these issues, see Trefethen (1999) and the references therein. The 'pseudospectral gateway' web-site (Pseudospectral Gateway, 2003, http://web.comlab.ox.ac.uk/projects/pseudospectra/) is a good general source.

Motivated by this pseudospectral philosophy, our aim in this paper is to compute the *pseudospectral abscissa* of A, namely

$$\alpha_\epsilon = \max\{\mathrm{Re}\,\lambda : \lambda \text{ an eigenvalue of } X \text{ and } \|X - A\| \leqslant \epsilon\},$$

where the real constant $\epsilon > 0$ is fixed at the outset, and $\|\cdot\|$ denotes the usual matrix operator or 2-norm. In other words, we wish to maximize the real part of a point in the ϵ-pseudospectrum of A.

Clearly, the pseudospectral abscissa α_ϵ is negative exactly when all matrices within a distance ϵ of A are stable (that is, have all eigenvalues in the open left half-plane), or in other words, when the *distance to instability* of A is larger than ϵ. This distance, introduced in Van Loan (1985), is also commonly called the *complex stability radius* (Hinrichsen & Pritchard, 1986): computing it is a special case of calculating the $\mathbf{H}_\infty$ norm of a transfer matrix. Specifically, the $\mathbf{H}_\infty$ norm of the function $H(s) = (sI - A)^{-1}$ is defined as $\|H\|_\infty = \sup_{\omega\in\mathbf{R}} \|H(\mathrm{i}\omega)\|$, and this quantity is just the reciprocal of the distance to instability (see for example Boyd *et al.*, 1989). In our notation,

$$\alpha_\epsilon < 0 \quad \Leftrightarrow \quad \|H\|_\infty < \frac{1}{\epsilon}.$$

The interest of the distance to instability and the $\mathbf{H}_\infty$ norm in robust control has encouraged broad study of algorithms for computing these quantities. The basic bisection method for the distance to instability appeared in Byers (1988) (see also Hinrichsen & Motscha, 1988), and was generalized to the $\mathbf{H}_\infty$ norm in Boyd *et al.* (1989). Quadratically convergent versions soon followed (Boyd & Balakrishnan, 1990; Bruinsma & Steinbuch, 1990), and research on faster methods continues—see for example Genin *et al.* (1998), He & Watson (1998).

Although a very simple relationship links the pseudospectral abscissa on the one hand with the complex stability radius on the other, using this relationship to convert an effective algorithm for the latter into one for the former does not seem immediate. A basic bisection method for the pseudospectral abscissa, akin to that of Byers (1988), appears in Burke *et al.* (2003).

In this paper we present a globally and locally quadratically convergent algorithm for computing the pseudospectral abscissa, motivated by the algorithms referenced above for the complex stability radius and $\mathbf{H}_\infty$ norm. Like them, the new algorithm depends heavily on computing eigenvalues of Hamiltonian matrices. Unlike those algorithms, however, the new method relies on a 'criss-cross' procedure to explore the possibly complicated two-dimensional geometry of the pseudospectrum. Extensive numerical tests suggest the algorithm is fast, accurate and reliable, and should prove a helpful tool in robust stability analysis.

2. Cross-sections of the pseudospectrum

We begin with some notation. As before, we consider a fixed n-by-n matrix A, and a fixed real $\epsilon > 0$. The (ϵ-)pseudospectrum of A is then just the set of eigenvalues of complex matrices whose 2-norm distance from A is no more than ϵ. However, while this is a conceptually simple approach to the pseudospectrum, an equivalent definition is more helpful computationally.

Let us define a function $h : \mathbf{R}^2 \to \mathbf{R}$ by

$$h(x, y) = \sigma_{\min}(A - (x + \mathrm{i}y)I) - \epsilon,$$

where $\sigma_{\min}$ denotes the smallest singular value and i is the imaginary unit. We can then define the (ϵ-)pseudospectrum of A equivalently as the set

$$\{(x, y) \in \mathbf{R}^2 : h(x, y) \leqslant 0\}$$

(see for example Trefethen, 1999). Analogously, we define the *strict pseudospectrum* as

$$\{(x, y) \in \mathbf{R}^2 : h(x, y) < 0\}.$$

With this notation, the pseudospectral abscissa of A is given by

$$\alpha_\epsilon = \max\{x : (x, y) \in \mathbf{R}^2,\ h(x, y) \leqslant 0\}.$$

The algorithm we present relies heavily on a step we call a *vertical search*: for a fixed real x, we look for real zeros of the function $h(x, \cdot)$. For this purpose, the following easy and well-known result (going back to Byers, 1988) is fundamental for us, both theoretically and computationally.

LEMMA 2.1 (Hamiltonian eigenvalues) For real numbers x and y, the number $\epsilon > 0$ is a singular value of the matrix $A - (x + \mathrm{i}y)I$ if and only if $\mathrm{i}y$ is an eigenvalue of the Hamiltonian matrix

$$H(x) = \begin{bmatrix} xI - A^* & \epsilon I \\ -\epsilon I & A - xI \end{bmatrix}. \tag{2.1}$$

This holds in particular if $h(x, y) = 0$.

Consider a fixed real x. It follows immediately from this lemma that the function $h(x, \cdot)$ can have at most $2n$ real zeros. To find them all, we compute all the imaginary eigenvalues $\{\mathrm{i}y_j\}$ of $H(x)$, discarding any for which $\sigma_{\min}(A - (x + \mathrm{i}y_j)I) < \epsilon$. The resulting $\{y_j\}$ are all the desired zeros of $h(x, \cdot)$.

We distinguish two types of zeros of the continuous real function $y \mapsto h(x, y)$: *crossing* zeros, where the function changes sign, and *non-crossing* zeros, where it does not. Notice that $h(x, y) > 0$ whenever $|y|$ is sufficiently large. So if we write out a non-decreasing list of the zeros, writing non-crossing zeros twice, then the list will be even in length, say $2m(x)$, and we can write it as

$$l_1(x) \leqslant u_1(x) \leqslant l_2(x) \leqslant u_2(x) \leqslant \cdots \leqslant u_{m(x)}(x). \tag{2.2}$$

If one of the inequalities holds with equality, any immediately neighbouring inequality is strict. We have $h(x, y) < 0$ for y in the set

$$\bigcup_{j=1}^{m(x)} (l_j(x), u_j(x))$$

(which we can think of as a vertical 'cross-section' of the strict pseudospectrum), and $h(x, y) > 0$ for y in the set

$$(-\infty, l_1(x)) \cup \bigcup_{j=1}^{m(x)-1} (u_j(x), l_{j+1}(x)) \cup (u_{m(x)}, \infty).$$

The following result is useful for distinguishing crossing and non-crossing zeros. Recall that a singular value σ of an n-by-n matrix B is *simple* if σ^2 is a simple eigenvalue of B^*B. If $\sigma > 0$, this is equivalent to σ being a simple eigenvalue of the Hermitian matrix

$$\begin{bmatrix} 0 & B \\ B^* & 0 \end{bmatrix}.$$

We remark that, generically, we expect the matrix A to have the property that $A - zI$ has all simple singular values for all complex z. To see this, we argue as follows. Denote the real vector space of n-by-n Hermitian matrices by $\mathbf{H}^n$. This vector space has dimension n^2. The subset $\mathbf{H}_0^n$ of matrices with a multiple eigenvalue has dimension $n^2 - 3$ (a result going back to von Neumann & Wigner, 1929). Hence for a generic matrix A, the two-dimensional manifold

$$\{(A - zI)^*(A - zI) : z \in \mathbf{C}\} \subset \mathbf{H}^n$$

will not intersect $\mathbf{H}_0^n$, and so $A - zI$ has all simple singular values for all complex z.

In the case when the matrix A is real, the argument needs slightly more care. In this case, the relevant space consists of the n-by-n real symmetric matrices, which has dimension $n(n + 1)/2$. The subset of those matrices having a multiple eigenvalue equal to ϵ^2 has dimension $n(n + 1)/2 - 3$, and hence again will typically miss the two-dimensional manifold above. Thus for a generic real matrix A, at every complex z satisfying $\sigma_{\min}(A - zI) = \epsilon$ this smallest singular value is simple.

Of course, there are non-generic examples of matrices A, real and complex, for which a multiple $\sigma_{\min}(A - zI)$ occurs for some z. In particular, this occurs when A is normal ($AA^* = A^*A$). In this case, the pseudospectrum is simply the union of discs with radius ϵ and so, as long as A has at least two distinct eigenvalues, the boundary of the pseudospectrum must be non-smooth for sufficiently large ϵ, with the non-smoothness occurring at points z for which $\sigma_{\min}(A - zI)$ is multiple.

LEMMA 2.4 (crossing versus non-crossing zeros) Given real numbers x and y_0, suppose that $\mathrm{i}y_0$ is an eigenvalue of the Hamiltonian matrix $H(x)$, and that the singular value $\sigma_{\min}(A - (x + \mathrm{i}y_0)I)$ is simple and equals ϵ. Then y_0 is a crossing zero of the function $h(x, \cdot)$ if and only if the eigenvalue $\mathrm{i}y_0$ has odd algebraic multiplicity.

Proof. If the eigenvalue $\mathrm{i}y_0$ has algebraic multiplicity m, then the function $p : \mathbf{R} \to \mathbf{C}$ defined by

$$p(y) = \det(H(x) - \mathrm{i}yI) = \det \begin{bmatrix} xI - A^* - \mathrm{i}yI & \epsilon I \\ -\epsilon I & A - xI - \mathrm{i}yI \end{bmatrix}$$

satisfies

$$p(y) \sim \beta(y - y_0)^m \quad \text{for } y \text{ close to } y_0,$$

for some non-zero $\beta \in \mathbf{C}$. But by multiplying the first n rows of the above determinant by -1 and then interchanging them with the corresponding second n rows, we obtain

$$\begin{aligned} p(y) &= \det \begin{bmatrix} -\epsilon I & A - (x + \mathrm{i}y)I \\ (A - (x + \mathrm{i}y)I)^* & -\epsilon I \end{bmatrix} \\ &= (-1)^n \prod_{j=1}^{n} [\sigma_j(A - (x + \mathrm{i}y)I) - \epsilon]\,[\sigma_j(A - (x + \mathrm{i}y)I) + \epsilon] \end{aligned}$$

where $\sigma_j(\cdot)$ denote the singular values, listed by multiplicity. In particular p is real-valued, and so changes sign at y_0 if and only if the multiplicity m is odd.

On the other hand, for all real y close to y_0, each factor in the above product is strictly positive, except the single factor

$$\sigma_{\min}(A - (x + \mathrm{i}y)I) - \epsilon = h(x, y).$$

Hence for such y, the signs of $p(y)$ and $h(x, y)$ are related by a factor of $(-1)^n$, and the result now follows. □

In addition to the vertical search, our algorithm also relies on a *horizontal search*: for fixed real y, we look for the largest real zero of the function $h(\cdot, y)$. For this purpose, the following result, quite analogous to Lemma 2.1 (Hamiltonian eigenvalues), is fundamental.

LEMMA 2.5 (horizontal search) For real numbers x and y, the number $\epsilon > 0$ is a singular value of the matrix $A - (x + \mathrm{i}y)I$ if and only if $\mathrm{i}x$ is an eigenvalue of the Hamiltonian matrix

$$\widetilde{H}(y) = \begin{bmatrix} \mathrm{i}A^* - yI & \epsilon I \\ -\epsilon I & \mathrm{i}A + yI \end{bmatrix}.$$

This holds in particular if $h(x, y) = 0$. Furthermore, x is the largest real zero of $h(\cdot, y)$ if and only if $\mathrm{i}x$ is the imaginary eigenvalue of $\widetilde{H}(y)$ with largest imaginary part.

Proof. The singular values of the matrix $A - (x + \mathrm{i}y)I$ are the same as those of

$$i(A - (x + \mathrm{i}y)I) = \mathrm{i}A - (-y + \mathrm{i}x)I.$$

Now applying Lemma 2.1 (Hamiltonian eigenvalues) with A, x, and y replaced by $\mathrm{i}A$, $-y$, and x respectively proves the first statement. The second is an immediate consequence.

To prove the last statement, we first make a subsidiary claim: if $\mathrm{i}x'$ is an imaginary eigenvalue of the matrix $\widetilde{H}(y)$, then the function $h(\cdot, y)$ has a real zero greater than or equal to x'. To see this, observe that ϵ is a singular value of the matrix $A - (x' + \mathrm{i}y)I$, so

$$\epsilon \geqslant \sigma_{\min}(A - (x' + \mathrm{i}y)I) = h(x', y) + \epsilon,$$

and hence $h(x', y) \leqslant 0$. Since the function $x \in \mathbf{R} \mapsto h(x, y)$ is continuous, and positive for large values of x, it must have a real zero $x \geqslant x'$.

Returning to the last statement, suppose x is the largest real zero of $h(\cdot, y)$. Then $\mathrm{i}x$ is an imaginary eigenvalue of the matrix $\widetilde{H}(y)$. Furthermore, for any other imaginary eigenvalue $\mathrm{i}x'$, the claim we proved above shows the existence of a real zero $x'' \geqslant x'$ for $h(\cdot, y)$. But by assumption, $x \geqslant x''$, so $\mathrm{i}x$ must be the imaginary eigenvalue with largest imaginary part.

Conversely, suppose $\mathrm{i}x'$ is the imaginary eigenvalue of the matrix $\widetilde{H}(y)$ with largest imaginary part. Again using our claim, the function $h(\cdot, y)$ has a real zero $x \geqslant x'$. But then $\mathrm{i}x$ is an imaginary eigenvalue of $\widetilde{H}(y)$, so by assumption we must in fact have $x = x'$. We have therefore shown that x' is a real zero of $h(\cdot, y)$. Finally, if x'' is any other real zero, then $\mathrm{i}x''$ is an imaginary eigenvalue of $\widetilde{H}(y)$, so by assumption, $x'' \leqslant x$, which concludes the proof. □

To analyse our algorithm we need the following technical result. It states that any vertical line of points with first coordinate strictly between the spectral abscissa and the pseudospectral abscissa must intersect the strict pseudospectrum.

THEOREM 2.6 (non-degenerate components) For any real x in the interval $(\alpha, \alpha_\epsilon)$, there exists a real y such that $h(x, y) < 0$.

Proof. Using Burke *et al.* (2003, Theorems 4.5 and 5.1), there exists a continuous path in the complex plane $\mathbf{C}$ from an eigenvalue of A to an endpoint with real part α_ϵ, and with the exception of this endpoint, the path lies entirely in the strict pseudospectrum (using the natural identification of $\mathbf{R}^2$ and $\mathbf{C}$). Clearly the vertical line of points with first component x must intersect this curve, and the result follows. □

3. Computing the pseudospectral abscissa

The simple method we describe in this section is motivated by an algorithm of Boyd & Balakrishnan (1990). Consider the minimization problem

$$\bar{\mu} = \min_{y \in \mathbf{R}} \sigma_{\min}(A - y\mathrm{i}I)$$

(a special case of the problem they consider). Their approach, in outline, is as follows. Given a current estimate $\mu \geqslant \bar{\mu}$, find all real y satisfying $\sigma_{\min}(A - \mathrm{i}yI) = \mu$, using Lemma 2.1 (Hamiltonian eigenvalues). Denote the solutions by $y_1 \leqslant y_2 \leqslant \cdots \leqslant y_{2m}$, listing non-crossing solutions twice, analogously to the list (2.2). Now update

$$\mu := \min_j \sigma_{\min}\Big(A - \frac{y_{2j-1} + y_{2j}}{2}\,\mathrm{i}I\Big),$$

and iterate. Geometrically, this method amounts to considering intersection points of the graph of the function $y \mapsto \sigma_{\min}(A - yiI)$ with various horizontal and vertical lines. The iterates converge globally and locally quadratically to $\bar{\mu}$ (Boyd & Balakrishnan, 1990).

Our algorithm for computing the pseudospectral abscissa α_ϵ uses a similar approach. It depends on finding zeros of the functions $h(x, \cdot)$ (for fixed real x) and $h(\cdot, y)$ (for fixed real y) as we described in the previous section, by computing the imaginary eigenvalues of associated Hamiltonian matrices. Figures illustrating the behaviour of the algorithm on two examples may be found in Section 6, where implementation issues are also discussed.

ALGORITHM 3.1 (criss-cross method)

1. *Initialize:* $x^1 = \alpha$ and $r = 1$.
2. *Vertical search:* Find all zeros

$$l_1^r \leqslant u_1^r \leqslant l_2^r \leqslant u_2^r \leqslant \cdots \leqslant u_{m^r}^r$$

of the function $h(x^r, \cdot)$, listing non-crossing zeros twice.
3. *Horizontal searches:* For each $j = 1, 2, \ldots, m^r$, define

$$y_j^r = \frac{l_j^r + u_j^r}{2},$$

and find the largest zero x_j^r of the function $h(\cdot, y_j^r)$.
4. *Update:* Define

$$x^{r+1} = \max\{x_j^r : j = 1, 2, \ldots, m^r\},$$

increment r by one, and return to Step 2.

Notice that, with the notation of the list (2.2), we have $m^r = m(x^r)$, $l_j^r = l_j(x^r)$, and $u_j^r = u_j(x^r)$. Notice also that if k is any maximizing choice of the index j in Step 4 (Update) of the rth iteration of the algorithm, then in Step 2 (Vertical search) of the next iteration, y_k^r must appear in the list of zeros of the function $h(x^{r+1}, \cdot)$.

The vertical search is, as already noted, accomplished by computing the imaginary eigenvalues of $H(x^r)$ and discarding those that correspond to singular values larger than ϵ. This avoids unnecessary horizontal searches in the next step. On the other hand, there is no need to check the correspondence between imaginary eigenvalues of $\widetilde{H}(y_j^r)$ and singular values during horizontal searches; all we need is the imaginary eigenvalue with largest imaginary part, by Lemma 2.5 (horizontal search).

THEOREM 3.2 (global convergence) The criss-cross method generates increasing iterates x^r, with limit the pseudospectral abscissa α_ϵ.

Proof. First observe that the new iterate x^{r+1} generated in Step 3 is a zero of $h(\cdot, y_j^r)$ for some index j, and so $x^{r+1} \leqslant \alpha_\epsilon$. It follows by induction that $x^r \leqslant \alpha_\epsilon$ for all r.

At each iteration r, if $l_j^r = u_j^r$ for each j then the vertical line of points with first component x^r misses the strict pseudospectrum entirely, so by Theorem 2.6 (non-degenerate components) we deduce $x^r = \alpha_\epsilon$ and we have nothing more to prove. If, on the

other hand, $l_j^r < u_j^r$ for some j, then $h(x^r, y_j^r) < 0$, and so $x^{r+1} > x^r$. So we can assume that the sequence (x^r) is strictly increasing, and bounded above strictly by α_ϵ.

Suppose, by way of contradiction, that x^r approaches a limit $x^\infty < \alpha_\epsilon$. Since the numbers m_r are uniformly bounded (by $2n$), we can choose a subsequence S of the natural numbers $\mathbf{N}$ such that m_r equals some constant m for all indices r in S, and such that for some cluster points l_j^∞, u_j^∞ and y_j^∞ (for $j = 1, 2, \ldots, m$) we have

$$l_j^r \to l_j^\infty, \quad u_j^r \to u_j^\infty, \quad \text{and} \quad y_j^r \to y_j^\infty \quad \text{as } r \to \infty \text{ in } S.$$

For any real $\mu \in [0, 1]$ we know

$$h(x^r, \mu l_j^r + (1 - \mu) u_j^r) \leqslant 0,$$

so by continuity we deduce

$$h(x^\infty, \mu l_j^\infty + (1 - \mu) u_j^\infty) \leqslant 0.$$

Hence

$$h(x^\infty, y) \leqslant 0 \quad \text{whenever } l_j^\infty \leqslant y \leqslant u_j^\infty.$$

If $l_j^\infty < u_j^\infty$ for some index j, then $l_j^\infty < y_j^\infty < u_j^\infty$, so

$$l_j^\infty < y_j^r < u_j^\infty \quad \text{for all large } r \in S.$$

For such an index r we deduce $h(x^\infty, y_j^r) \leqslant 0$, so by the definition of the horizontal search process we must have the contradiction $x^{r+1} \geqslant x^\infty$.

Thus we can assume

$$l_j^\infty = u_j^\infty \quad \text{for each } j = 1, 2, \ldots, m. \tag{3.1}$$

But since $\alpha < x^\infty < \alpha_\epsilon$, Theorem 2.6 (non-degenerate components) implies the existence of a real y^∞ satisfying $h(x^\infty, y^\infty) < 0$. By continuity, there exists $\delta > 0$ such that

$$h(x, y) < 0 \quad \text{whenever } |x - x^\infty| < \delta \text{ and } |y - y^\infty| < \delta.$$

But then for all large r we have

$$h(x^r, y) < 0 \quad \text{whenever } |y - y^\infty| < \delta,$$

so there must exist an index j such that $u_j^r - l_j^r \geqslant 2\delta$, contradicting (3.1). □

A bisection method for computing the pseudospectral abscissa was presented in Burke *et al.* (2003). As we show below, the criss-cross method has the added merit of local quadratic convergence.

4. The boundary of the pseudospectrum

To prove quadratic convergence, we need to study the boundary of the pseudospectrum near a local maximizer for the *pseudospectral abscissa problem*

$$\alpha_\epsilon = \max\{x : (x, y) \in \mathbf{R}^2,\ h(x, y) \leqslant 0\}.$$

The following result is central to our analysis. It gives a local description of the pseudospectrum, using a single real-analytic function.

THEOREM 4.1 (simple singular values) For any point $(x_0, y_0) \in \mathbf{R}^2$, if the smallest singular value of the matrix $A - (x_0 + \mathrm{i}y_0)I$ is simple, then the function h is real-analytic near (x_0, y_0), with gradient

$$\nabla h(x_0, y_0) = [-\mathrm{Re}\,(u^* v), \mathrm{Im}\,(u^* v)] \tag{4.1}$$

for any consistent pair of unit left and right singular vectors $u, v \in \mathbf{C}^n$ corresponding to the smallest singular value. Suppose furthermore that $h(x_0, y_0) = 0$, and $\mathrm{Re}\,(u^* v) < 0$. Then there is a function $f : \mathbf{R} \to \mathbf{R}$, real-analytic near zero, such that

$$f(0) = 0, \quad f'(0) = -\frac{\mathrm{Im}\,(u^* v)}{\mathrm{Re}\,(u^* v)},$$

and the signs of the functions $h(x, y)$ and $x - x_0 + f(y - y_0)$ coincide for all points $(x, y) \in \mathbf{R}^2$ close to (x_0, y_0).

Proof. The matrix

$$S(p, q) = \begin{bmatrix} 0 & A - (p + \mathrm{i}q)I \\ A^* - (p - \mathrm{i}q)I & 0 \end{bmatrix}$$

has a simple eigenvalue $\sigma_{\min}(A - (x_0 + \mathrm{i}y_0)I)$ when the parameter $(p, q) \in \mathbf{C}^2$ equals (x_0, y_0). Since $S(p, q)$ depends linearly on (p, q), this eigenvalue $\lambda(p, q)$ is a holomorphic function of (p, q) (see Kato, 1982). Furthermore, for real x and y we have

$$\lambda(x, y) = \sigma_{\min}(A - (x + \mathrm{i}y)I) = h(x, y) + \epsilon.$$

Hence the function h is real-analytic near (x_0, y_0).

The unit (left and right) eigenvector of the Hermitian matrix $S(x_0, y_0)$ corresponding to the eigenvalue $\sigma_{\min}(A - (x_0 + \mathrm{i}y_0)I)$ is

$$w = \frac{1}{\sqrt{2}} \begin{bmatrix} u \\ v \end{bmatrix}.$$

Now standard perturbation theory (Kato, 1982) shows that the partial derivatives of the functions $\lambda : \mathbf{C}^2 \to \mathbf{C}$ and $h : \mathbf{R}^2 \to \mathbf{R}$ at (x_0, y_0) are

$$\lambda_p(x_0, y_0) \ = \ h_x(x_0, y_0) = w^* \begin{bmatrix} 0 & -I \\ -I & 0 \end{bmatrix} w \ = \ -\mathrm{Re}\,(u^* v)$$

$$\lambda_q(x_0, y_0) \ = \ h_y(x_0, y_0) = w^* \begin{bmatrix} 0 & -\mathrm{i}I \\ \mathrm{i}I & 0 \end{bmatrix} w \ = \ \mathrm{Im}\,(u^* v)$$

as required.

Now, since by assumption the partial derivative $\lambda_p(x_0, y_0)$ is non-zero, the equation $\lambda(p, q) = \epsilon$ defines $p(q)$ as a holomorphic function of q near the point $(x_0, y_0) \in \mathbf{C}^2$ (see Dienes, 1957, for example). On the other hand, since $h_x(x_0, y_0)$ is non-zero, we could equally well apply the classical implicit function theorem to the real equation $h(x, y) = 0$, to define $x(y)$ as a real function of y near the point $(x_0, y_0) \in \mathbf{R}^2$. The two functions $p(y)$ and $x(y)$ must coincide for real y, so the function we obtain is real-analytic.

Thus there is a function $f : \mathbf{R} \to \mathbf{R}$, real-analytic near zero, such that $f(0) = 0$ and

$$h(x_0 - f(y - y_0), y) = 0$$

for all real y close to y_0. The formula for $f'(0)$ follows easily by the chain rule.

Finally, notice that

$$h_x(x_0 - f(y - y_0), y) > 0$$

for all real y close to y_0, since the left-hand side is continuous in y and strictly positive when $y = y_0$. Hence we deduce

$$h(x, y) > 0 \quad \Leftrightarrow \quad x > x_0 - f(y - y_0)$$

holds for all points $(x, y) \in \mathbf{R}^2$ close to (x_0, y_0). □

To paraphrase, under the assumptions of the theorem the pseudospectrum is defined near the point (x_0, y_0) by the real-analytic inequality

$$x - x_0 \leqslant -f(y - y_0), \tag{4.2}$$

and locally the function h is zero exactly at points (x, y) where this holds with equality. (Clearly, with identical assumptions except that $\mathrm{Re}\,(u^*v) > 0$, the same result holds, except that now the functions $h(x, y)$ and $x - x_0 + f(y - y_0)$ have opposite signs.)

We now introduce the idea of a regular point for a matrix. Regular points are 'non-degenerate' in the sense of Burke *et al.* (2003). However, unlike non-degenerate points, a regular point must correspond to a simple smallest singular value.

DEFINITION 4.4 (regularity) We call a point $(x_0, y_0) \in \mathbf{R}^2$ *regular* for the matrix A if the matrix $A - (x_0 + \mathrm{i}y_0)I$ has a simple smallest singular value with corresponding unit left and right singular vectors $u, v \in \mathbf{C}^n$ satisfying $u^*v \neq 0$.

If the point (x_0, y_0) is regular, we see by Theorem 4.1 (smallest singular values) that the function h is real-analytic around (x_0, y_0), with non-zero gradient. Our main interest is in maximizers for the pseudospectral abscissa problem: when a local maximizer is a regular point, the local structure of the pseudospectrum is easy to describe, as the following result shows.

COROLLARY 4.5 (local maximizers) Suppose that the point $(x_0, y_0) \in \mathbf{R}^2$ is a local maximizer of the pseudospectral abscissa problem, and is regular. Then there is a function $f : \mathbf{R} \to \mathbf{R}$, real-analytic around zero, such that

$$f(0) = 0, \;\; f'(0) = 0, \ldots, \; f^{(2k-1)}(0) = 0, \;\; f^{(2k)}(0) > 0 \;\text{ where } k \in \mathbf{N}, \tag{4.3}$$

and such that the signs of the functions $h(x, y)$ and $x - x_0 + f(y - y_0)$ coincide for all points $(x, y) \in \mathbf{R}^2$ close to (x_0, y_0).

Proof. By Theorem 4.1 (simple singular values), the function h is real-analytic around (x_0, y_0). If a vector $d \in \mathbf{R}^2$ satisfies $d \cdot \nabla h(x_0, y_0) < 0$ then by the chain rule, for all small real $t > 0$ we have

$$h((x_0, y_0) + td) = h(x_0, y_0) + t(d \cdot \nabla h(x_0, y_0)) + \mathrm{O}(t^2) < 0$$

so the point $(x_0, y_0) + td$ lies in the strict pseudospectrum. Since (x_0, y_0) is a local maximizer, we deduce $d_1 < 0$. It follows easily that $\nabla h(x_0, y_0)$ is a positive multiple of the vector $(1, 0)$.

Now by (4.1), for unit left and right singular vectors u and v corresponding to the smallest singular value ϵ, we know u^*v is real and negative. Hence Theorem 4.1 shows the existence of the real-analytic function f satisfying $f(0) = 0 = f'(0)$ such that, near the point (x_0, y_0), the pseudospectrum is defined by inequality (4.2). But now we know that the origin is a local maximizer for the problem

$$\max\{s : s \leqslant -f(t),\ s, t \in \mathbf{R}\},$$

and property (4.3) now follows by considering the Taylor expansion of the function f around zero. □

5. Quadratic convergence

Our technique for proving quadratic convergence is based on an analogous idea in Boyd & Balakrishnan (1990). The key tool is the following result.

LEMMA 5.1 (bisection) Consider a function $f : \mathbf{R} \to \mathbf{R}$ that is real-analytic around zero and satisfies property (4.3). Construct a positive real sequence (v_r) iteratively by choosing an initial small positive v_0 and then, for each index $r = 0, 1, 2, \ldots$, defining

$$v_{r+1} = f\Big(\frac{\hat{u} + \bar{u}}{2}\Big),$$

where $\hat{u}$ and $\bar{u}$ are the two small solutions of the equation $f(u) = v_r$. Then there exists a constant M such that $0 \leqslant v_{r+1} \leqslant Mv_r^2$ for all r, so the sequence (v_r) decreases quadratically to zero.

Proof. For small real u we have the Taylor expansion

$$f(u) = \beta u^{2k} + \mathrm{O}(u^{2k+1})$$

for some real $\beta > 0$. For some small $\delta > 0$ we know that the function f is strictly decreasing on $[-\delta, 0]$ and strictly increasing on $[0, \delta]$. Hence f has inverse functions $p : [0, f(-\delta)] \to [-\delta, 0]$ and $q : [0, f(\delta)] \to [0, \delta]$. These functions have Puiseux series expansions (Dienes, 1957, p. 246), so a simple calculation shows

$$p(v) = -\beta^{-1/2k}\, v^{1/2k} + \mathrm{O}(v^{1/k})$$
$$q(v) = \beta^{-1/2k}\, v^{1/2k} + \mathrm{O}(v^{1/k})$$

for small $v \geqslant 0$. Since

$$\frac{\hat{u} + \bar{u}}{2} = \frac{p(v_r) + q(v_r)}{2} = \mathrm{O}(v_r^{1/k}),$$

we deduce $v_{r+1} = \mathrm{O}((v_r^{1/k})^{2k}) = \mathrm{O}(v_r^2)$, as required. □

By applying this result at each global maximum in the pseudospectral abscissa problem, we obtain a quadratic convergence result for the criss-cross method, assuming regularity (see Definition 4.4).

THEOREM 5.2 (quadratic convergence) If each global maximizer in the pseudospectral abscissa problem is regular, then the criss-cross method produces iterates x^r that increase to the pseudospectral abscissa α_ϵ with locally quadratic rate.

Proof. Theorem 3.2 (global convergence) shows that x^r increases to α_ϵ. Suppose the global maximizers are (α_ϵ, y_j) (for $j = 1, 2, \ldots, m$). By Corollary 4.5 (local maximizers), near the point (α_ϵ, y_j) the pseudospectrum can be defined by a single inequality of the form

$$x - \alpha_\epsilon \leqslant -f_j(y - y_j),$$

where each function $f_j : \mathbf{R} \to \mathbf{R}$ is real-analytic around zero and satisfies property (4.3).

We now apply Lemma 5.1 (bisection) to each of these functions f_j. We deduce the existence of constants $M_1, M_2, \ldots, M_m$ such that

$$\alpha_\epsilon - x^{r+1} \leqslant M_j(\alpha_\epsilon - x^r)^2$$

for each index j and for all large r. The result now follows. □

Remarkably, quadratic convergence for the algorithm of Boyd & Balakrishnan (1990) needs no regularity assumption. Analogously, we conjecture that the criss-cross method always converges quadratically even without the regularity assumption of the above result.

6. Numerical implementation

We have implemented the criss-cross method in MATLAB and tested it extensively. The key computational step is finding the eigenvalues of the Hamiltonian matrices $H(x)$ (which is real if A is) and $\widetilde{H}(y)$ (which is complex even if A is real). Hamiltonian matrices have eigenvalues that are symmetric with respect to the imaginary axis, so all non-imaginary eigenvalues occur in pairs and the number of imaginary eigenvalues is always even (since the order of the matrix is even). We use real and complex implementations (Benner *et al.*, 1999, 2000) of Van Loan's square-reduced algorithm (Van Loan, 1984), which takes advantage of Hamiltonian structure and delivers eigenvalues that are exactly symmetric with respect to the imaginary axis. As noted in Byers' original paper (Byers, 1988), this property is important if one wishes to be able to reliably determine whether a Hamiltonian matrix has an imaginary eigenvalue. To be more specific, suppose that for a given iteration r, the zeros $\{l_j^r, u_j^r\}$ are all crossing zeros (so are distinct), and furthermore that their numerical values are well separated. Suppose also that for each zero, the corresponding imaginary eigenvalue of the Hamiltonian matrix $H(x^r)$ is simple. Then the specialized

Hamiltonian eigenvalue solver delivers the imaginary eigenvalues with real parts exactly equal to zero, uncontaminated by rounding errors, since small Hamiltonian perturbations to $H(x^r)$ cannot move a simple imaginary eigenvalue off the imaginary axis.

However, this observation does not hold in the limit as the algorithm converges. Let $(\alpha_\epsilon, \tilde{y})$ denote a maximizer of the real part over the pseudospectrum. Clearly, $\tilde{y}$ is a non-crossing zero of $h(\alpha_\epsilon, \cdot)$, and consequently the imaginary eigenvalue $\mathrm{i}\tilde{y}$ of $H(\alpha_\epsilon)$ has multiplicity two (or possibly higher in non-generic cases). For all $x^r < \alpha_\epsilon$, there is a pair of crossing zeros, say l^r, u^r, with $u^r - l^r \to 0$ as $r \to \infty$. As the algorithm converges, it becomes progressively more difficult to detect numerically whether the Hamiltonian matrix has two distinct but close imaginary eigenvalues, a double imaginary eigenvalue, or, because of rounding errors, a pair of distinct but close nonimaginary eigenvalues with the same imaginary part; all these scenarios are consistent with the Hamiltonian structure. In our implementation, we simply terminate the algorithm as soon as *either* the eigensolver fails to return an exactly imaginary eigenvalue in the vertical search of Step 2, *or* the horizontal search of Step 3 returns $x^{r+1} \leqslant x^r$. The occurrence of either event indicates that the accuracy limits inherent in floating point arithmetic have been reached. Furthermore, one of these events *must* occur eventually in floating point arithmetic, and, because of the quadratic convergence, typically occurs within 3 to 5 iterations. We have not found it necessary to use tolerances for either test.

A crucial advantage of the specialized Hamiltonian eigensolvers is that the property that the number of imaginary eigenvalues is even holds in the presence of rounding errors. This is essential for the criss-cross algorithm, because any loss of parity in the pair structure displayed in (2.2) may cause the algorithm to fail. The vertical search in the ideal algorithm without rounding discards imaginary eigenvalues $\mathrm{i}y_j$ for which $\sigma_{\min}(A - (x + \mathrm{i}y_j)I) < \epsilon$. In the presence of rounding, it is tempting to introduce a tolerance into this inequality, but this can create trouble if there is more than one singular value close to ϵ. It is much more reliable to check whether $\sigma_{\min}(A - (x + \mathrm{i}y_j)I)$ is the singular value of $A - (x + \mathrm{i}y_j)I$ that is closest to ϵ, discarding the $\mathrm{i}y_j$ for which this is not the case. Discards must take place in pairs, a property that is automatic if the matrix is real, since then the eigenvalues are also symmetric with respect to the real axis. Clearly, when the matrix is real, it is not necessary to carry out horizontal searches in both the lower and upper half-planes.

There is a further subtlety, which is well illustrated by an example of Demmel (1987),

$$A = \begin{bmatrix} -1 & -5 & -25 & -125 & -625 \\ 0 & -1 & -5 & -25 & -125 \\ 0 & 0 & -1 & -5 & -25 \\ 0 & 0 & 0 & -1 & -5 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}.$$

The matrix A has only one eigenvalue, -1. The smooth curve in Fig. 1 shows the boundary of the pseudospectrum of A for $\epsilon = 0{\cdot}01$, generated by the software package EigTool (Wright, 2002). (The legend on the right indicates that $\log_{10}(\epsilon) = -2$.) The plot allows us to immediately estimate the value of $\alpha_{0{\cdot}01}(A)$ to low accuracy, but not to high accuracy, since the computation depends on interpolation of $\sigma_{\min}(A-(x+\mathrm{i}y)I)$ on a two-dimensional grid. Superimposed on this pseudospectral plot are line segments, crosses and plus signs that clearly illustrate the vertical and horizontal searches in Steps 2 and 3 of the criss-cross method to compute $\alpha_{0{\cdot}01}(A)$. In this example, the first vertical search finds two

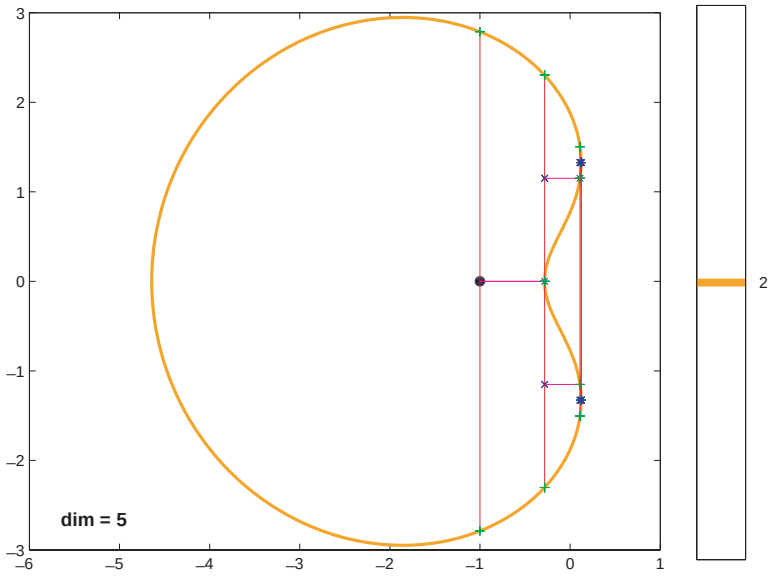

FIG. 1. Computing the pseudospectral abscissa for the Demmel example.

TABLE 1 *Iterates for the Demmel example*

r	x^r	m^r	y_1^r	y_2^r
1	−1·000000000000000	1	0·000000000000000	—
2	−0·283 307 773 738 337	2	−1·151 342 702 176 112	1·151 342 702 176 112
3	0·110 378 777 480 711	2	−1·328 011 519 739 448	1·328 011 519 739 448
4	0·122 855 725 365 556	2	−1·327 743 418 800 731	1·327 743 418 800 731
5	0·122 855 754 071 588	2	−1·327 743 418 079 968	1·327 743 418 079 968
6	0·122 855 754 072 281	0	—	—

crossing zeros, with, since A is real, $l_1^1 = -u_1^1$, so $y_1^1 = 0$ and hence the first horizontal search takes place along the real axis. For many real matrices, the resulting iterate $(x^2, 0)$ would be optimal and the algorithm would correctly terminate, but in this example $(x^2, 0)$ is a stationary point, but not a maximizer, of the real part over the pseudospectrum. Since A is real, 0 is a non-crossing zero of $h(x^2, \cdot)$, and 0 is a double eigenvalue of the Hamiltonian matrix $H(x^2)$. In addition, $h(x^2, \cdot)$ has two crossing zeros respectively at the top and bottom of Fig. 1. Suppose rounding errors cause the eigensolver to return only two imaginary eigenvalues, instead of the correct number of four, counting the zero eigenvalue twice. We would then have $l_2^1 = -u_2^1$, and the algorithm would terminate with $x^3 = x^2$, which is the wrong answer.

A safeguard to avoid this failure is easily enacted: in the update at Step 4 of iteration r of the algorithm, save the y value(s) (say y_k^r) corresponding to the maximum value x^{r+1} and then check whether the list of zeros generated by the vertical search in Step 2 of iteration $r + 1$ includes y_k^r. More specifically, since we cannot expect the agreement to be exact in

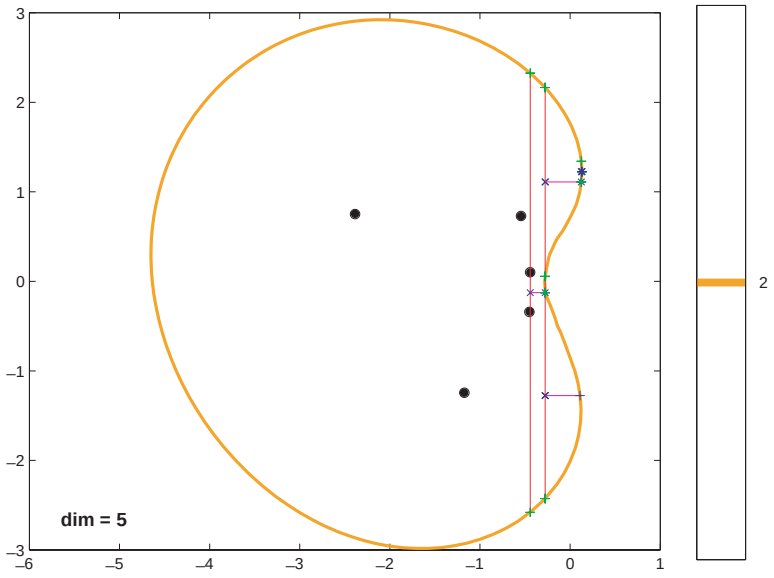

FIG. 2. Computing the pseudospectral abscissa for the perturbed Demmel example.

the presence of rounding errors, we do the following: if the list of zeros at iteration $r+1$ is non-empty and

$$y_k^r \in \left[l_j^{r+1} + \delta_j^{r+1}, u_j^{r+1} - \delta_j^{r+1} \right],$$

with $\delta_j^{r+1} = \tau(u_j^{r+1} - l_j^{r+1})$, for some $j \in \{1, \ldots, m^{r+1}\}$ and a fixed $\tau << 1$, then break the pair (l_j^{r+1}, u_j^{r+1}) into two pairs (l_j^{r+1}, y_k^r) and (y_k^r, u_j^{r+1}), substituting two horizontal searches defined by the two new midpoints for the one that would otherwise take place defined by the midpoint of (l_j^{r+1}, u_j^{r+1}). A reasonable choice for τ is 0·01, small enough that the breaking of one pair into two will take place when it is needed, as in the Demmel example, but large enough that the cost of an additional horizontal search will be incurred only when necessary. In order to avoid an unnecessary additional horizontal search just before termination, the safeguard just described may be invoked only if the quantity $u_j^{r+1} - l_j^{r+1}$ is not very small. Table 1 shows the iterates for the Demmel example to 16 decimal digits, the limit of the standard IEEE double-precision floating-point format.

For a second example, we introduce a complex perturbation to the Demmel example, changing the 5, 1 entry from 0 to 0·001i, so that the pseudospectrum is no longer symmetric with respect to the real axis. Figure 2 shows the pseudospectrum for $\epsilon = 0{\cdot}01$, along with the perturbed eigenvalues (shown as heavy dots) and the iterates of the criss-cross method. The numerical values of the iterates are given in Table 2.

We have also tested the criss-cross method on much larger matrices. The computational time is dominated, of course, by the cost of computing the eigenvalues of $H(x)$ and $\widetilde{H}(y)$. Typically, there are only a small number of eigenvalue computations, because the algorithm

TABLE 2 *Iterates for the perturbed Demmel example*

r	x^r	m^r	y_1^r	y_2^r
1	−0·442 437 235 099 235	1	−0·126 565 274 824 285	—
2	−0·276 899 490 766 155	2	−1·276 126 382 561 665	1·110 482 209 299 681
3	0·124 058 778 702 449	1	1·226 173 510 815 744	—
4	0·130 272 463 584 219	1	1·225 424 774 480 370	—
5	0·130 272 723 577 035	0	—	—

converges so rapidly. This number is further reduced if only the first few digits of the pseudospectral abscissa are required.

An analogous algorithm to compute the pseudospectral radius has recently been developed by Emre Mengi. The MATLAB code implementing our algorithm is freely available at http://www.cs.nyu.edu/faculty/overton/software/ and has been incorporated as a feature of EigTool (Wright, 2002), along with codes to compute the pseudospectral radius and related quantities.

Acknowledgements

Many thanks to Carsten Scherer for pointing out Boyd & Balakrishnan (1990), setting us on the course that led to this work. Thanks also to Peter Benner for providing the Hamiltonian eigenvalue software and to Emre Mengi for providing a MATLAB interface for the complex Hamiltonian code. We particularly thank Nick Trefethen and Tom Wright for their enthusiasm for adding the criss-cross algorithm to EigTool, and both Emre Mengi and Tom Wright for carrying this out in short order. The first author's research was supported in part by National Science Foundation Grant DMS-0203175. The second author's research was supported in part by NSERC. The third author's research was supported in part by National Science Foundation Grant CCR-0098145.

REFERENCES

BENNER, P., BYERS, R. & BARTH, E. (2000) Algorithm 800. Fortran 77 subroutines for computing the eigenvalues of a Hamiltonian matrix. *ACM Transactions on Mathematical Software*, **26**, 49–77.

BENNER, P., MEHRMANN, V. & XU, H. (1999) A note on the numerical solution of complex Hamiltonian and skew-Hamiltonian eigenvalue problems. *Electronic Trans. Numer. Anal.*, **8**, 115–126.

BOYD, S. & BALAKRISHNAN, V. (1990) A regularity result for the singular values of a transfer matrix and a quadratically convergent algorithm for computing its L_∞-norm. *Systems and Control Letters*, **15**, 1–7.

BOYD, S., BALAKRISHNAN, V. & KATAMBA, P. (1989) A bisection method for computing the H_∞ norm of a transfer matrix and related problems. *Math. Cont. Sig. Sys.*, **2**, 207–219.

BRUINSMA, N. A. & STEINBUCH, M. (1990) A fast algorithm to compute the $\mathbf{H}_\infty$-norm of a transfer function matrix. *Systems and Control Letters*, **14**, 287–293.

BURKE, J. V., LEWIS, A. S. & OVERTON, M. L. (2003) Optimization and pseudospectra, with applications to robust stability. *SIAM J. Matrix Analysis and Applications*, to appear.

BYERS, R. (1988) A bisection method for computing the distance of a stable matrix to the unstable matrices. *SIAM J. Scientific and Statistical Computing*, **9**, 875–881.

DEMMEL, J. W. (1987) A counterexample for two conjectures about stability. *IEEE Trans. Auto. Cont.*, **32**, 340–342.

DIENES, P. (1957) *The Taylor Series*. New York: Dover.

GENIN, Y., VAN DOOREN, P. & VERMAUT, V. (1998) Convergence of the calculation of $\mathbf{H}_\infty$ norms and related questions. *Proceedings Math. Theory of Networks and Systems-98*. (A. Beght, L. Finesso & G. Picci, eds). Padua: Il Poligrafo.

HE, C. & WATSON, G. A. (1998) An algorithm for computing the distance to instability. *SIAM J. Matrix Analysis and its Applications*, **20**, 101–116.

HINRICHSEN, D. & MOTSCHA, M. (1988) Optimization problems in the robustness analysis of linear state space systems. *Approximation and Optimization (Havana, 1987)* (A. Gomez, F. Guerra, M. A. Jimenez & G. Lopez, eds). Lecture Notes in Mathematics 1354. Berlin: Springer, pp. 54–78.

HINRICHSEN, D. & PRITCHARD, A. J. (1986) Stability radii of linear systems. *Systems and Control Letters*, **7**, 1–10.

KATO, T. (1982) *A Short Introduction to Perturbation Theory for Linear Operators*. New York: Springer.

TREFETHEN, L. N. (1999) Computation of pseudospectra. *Acta Numerica*, **8**, 247–295.

VAN LOAN, C. F. (1984) A symplectic method for approximating all the eigenvalues of a Hamiltonian matrix. *Linear Algebra and its Applications*, **61**, 233–251.

VAN LOAN, C. F. (1985) How near is a stable matrix to an unstable matrix? *Contemporary Mathematics*, **47**, 465–477.

VON NEUMANN, J. & WIGNER, E. (1929) Uber das Verhalten von Eigenwerten bei adiabatischen Prozessen. *Physik. Zeitschr.*, **30**, 467–470.

WRIGHT, T. G. (2002) EigTool: a graphical tool for nonsymmetric eigenproblems, Oxford University Computing Laboratory. http://www.comlab.ox.ac.uk/pseudospectra/eigtool/.