Do the Pseudospectra of a Matrix Determine its Behavior?

Anne Greenbaum*
Lloyd N. Trefethen**

TR 93-1371
August 1993

Department of Computer Science
Cornell University
Ithaca, NY 14853-7501

* Supported by DOE Contract DEFG0288ER25053.
** Supported by NSF Grant DMS-9116110.

Do the pseudospectra of a matrix determine its behavior?

Anne Greenbaum*
Courant Institute of Mathematical Sciences
251 Mercer St.
New York, NY 10012
greenbau@cims.nyu.edu

Lloyd N. Trefethen†
Department of Computer Science
Cornell University
Ithaca, NY 14853
LNT@cs.cornell.edu

Abstract. Let A and B be square matrices. It is shown that the condition *(R)* $\|(zI-A)^{-1}\| = \|(zI-B)^{-1}\|$ for all $z \in \mathbb{C}$ is equivalent to the condition *(P)* $\|p(A)\| = \|p(B)\|$ for all polynomials p if $\|\cdot\|$ is the Frobenius norm, but not if $\|\cdot\|$ is the 2-norm.

* Supported by DOE Contract DEFG0288ER25053.
† Supported by NSF Grant DMS-9116110.

1. The problem

Let A and B be complex square matrices, not necessarily of the same dimension, and let $\|\cdot\|$ be a matrix norm. Consider the conditions

$$\|(zI-A)^{-1}\| = \|(zI-B)^{-1}\| \qquad \forall z \in \mathbb{C} \tag{R}$$

and

$$\|p(A)\| = \|p(B)\| \qquad \forall p, \tag{P}$$

where p is a polynomial. Are these conditions equivalent?

Condition *(R)* asserts that A and B have the same resolvent norms throughout the complex plane. Equivalently, A and B have the same *pseudospectra,* where the ϵ-pseudospectrum of A is defined for each $\epsilon \geq 0$ by

$$\Lambda_\epsilon(A) = \{z \in \mathbb{C} : \ \|(zI-A)^{-1}\| \geq \epsilon^{-1}\}, \tag{1}$$

with the convention $\|(zI-A)^{-1}\| = \infty$ if z is an eigenvalue of A [10].

Condition *(P)* asserts that A and B have the same *behavior,* if behavior is measured by norms of polynomials, or equivalently, since analytic functions of matrices reduce to polynomials, by norms of functions. This notion of behavior may seem restrictive, but it captures much of what one wants to know about a matrix in applications. For example, the stability of an evolution process governed by A is determined by the norms $\|A^n\|$ in the discrete case or $\|e^{tA}\|$ in the continuous case [8]. Similarly, the convergence of matrix iterative algorithms for solving linear equations or finding eigenvalues, such as the GMRES and Arnoldi iterations [2], depends on how fast the norms $\|p_n^*(A)\|$ decrease as n increases, where p_n^* is the polynomial that minimizes $\|p(A)\|$ in a suitably normalized class of polynomials of degree n [3].

If $\|\cdot\|$ is the 2-norm and A is normal (i.e., A has a complete set of orthogonal eigenvectors), then $\|(zI-A)^{-1}\| = \mathrm{dist}(z, \Lambda(A))^{-1}$ and $\|p(A)\| = \sup_{z\in\Lambda(A)} |p(z)|$, where $\Lambda(A)$ denotes the spectrum of A, and thus *(R)* and *(P)* reduce to scalar conditions that are easily seen to be equivalent. If A is not normal, however, then $\|(zI-A)^{-1}\|$ and $\|p(A)\|$ are not determined by $\Lambda(A)$. Of course they are both determined by $(zI-A)^{-1}$, thanks to the Cauchy integral

$$p(A) = \frac{1}{2\pi i}\int_\Gamma (zI-A)^{-1}\, p(z)\, dz, \tag{2}$$

where Γ is any contour enclosing $\Lambda(A)$ [6]. If only the norm $\|(zI-A)^{-1}\|$ is known, however, then (2) can be applied to derive upper bounds for $\|p(A)\|$, but they are not in general sharp. The best known and most refined example of such a bound is the Kreiss Matrix Theorem, which can be derived from (2) by integration by parts combined with a few other tricks [7,8,11].

2. (P) $\Rightarrow$ (R)

In one direction our problem is straightforward.

Theorem 1. *(P) implies (R).*

Proof. If $\|p(A)\| = \|p(B)\|$ for all polynomials p, then $\|p(A)\| = 0$ if and only if $\|p(B)\| = 0$, and this implies that A and B have the same minimal polynomial, $m(z)$. In particular A and B have the same eigenvalues, and thus $\|(zI-A)^{-1}\| = \infty$ if and only if $\|(zI-B)^{-1}\| = \infty$. On the other hand let $z \in \mathbb{C}$ be a number distinct from these eigenvalues. Then $(zI-A)^{-1} = q(A)$, where $q(w)$ is any polynomial that interpolates $(z-w)^{-1}$ in the zeros of m, counted with multiplicity (i.e., the derivatives of orders $0, 1, \ldots, \mu-1$ are interpolated in the case of a zero of multiplicity μ) [5, Thm. 6.2.9]. Likewise, $(zI-B)^{-1} = q(B)$ for the same polynomial q. By *(P)*, we now have $\|q(A)\| = \|q(B)\|$, that is, $\|(zI-A)^{-1}\| = \|(zI-B)^{-1}\|$. ∎

3. (R) $\not\Rightarrow$ (P) in the 2-norm

Let $\|\cdot\|$ be the 2-norm. In this case the pseudospectra can be characterized in various other ways besides (1), such as

$$\Lambda_\epsilon(A) = \{z \in \mathbb{C} : \ z \in \Lambda(A+E) \text{ for some } E \text{ with } \|E\| \le \epsilon\} \tag{3}$$

$$= \{z \in \mathbb{C} : \ \sigma_{\min}(A) \le \epsilon\}, \tag{4}$$

where $\sigma_{\min}(A)$ denotes the smallest singular value of A. The first equality is valid for any matrix norm induced by a vector norm [1]; the second applies to the 2-norm only.

We thought at first that *(R)* certainly could not imply *(P)*, since the norm of the resolvent surely could not "contain enough information" to determine behavioral quantities exactly. Yet the information contents of *(R)* and *(P)* are better matched than they may at first appear. For example, both pseudospectra and norms of polynomials take no notice of eigenvalue multiplicities, as is illustrated by the matrices

$$A = \begin{pmatrix} 1 & \\ & 2 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & & \\ & 2 & \\ & & 2 \end{pmatrix}.$$

(Here and below, blank matrix entries are zero.) Both *(R)* and *(P)* are also satisfied if one takes $B = A^T$ or $B = UAU^*$ with U unitary. Furthermore, it is easily shown that *(R)* implies *(P)* if A and B are of dimension ≤ 2.

Nevertheless, the following counterexample shows that our first intuition was correct.

Theorem 2. *If* $\|\cdot\|$ *is the 2-norm, then (R) does not imply (P).*

Proof. Consider the Jordan blocks

$$J_1 = \begin{pmatrix} 0 & 1 & \\ & 0 & 1 \\ & & 0 \end{pmatrix}, \qquad J_2 = \begin{pmatrix} 0 & \alpha \\ & 0 \end{pmatrix}$$

with $\alpha \in \mathbb{C}$. We have $\|J_1\| = 1$ and $\|J_2\| = |\alpha|$, so if $\alpha > 1$, then $\|J_2\| > \|J_1\|$. On the other hand if $\alpha \leq \sqrt{2}$, then

$$\|(zI - J_2)^{-1}\| \leq \|(zI - J_1)^{-1}\| \qquad \forall z \in \mathbb{C}. \tag{5}$$

To show this we note first that for any Jordan block J, $\|(zI - J)^{-1}\|$ depends only on $|z|$, i.e., the pseudospectra are disks about the origin. (This can be proved by a diagonal similarity transformation.) Thus it is enough to establish (5) for z real and positive. Rather than providing an algebraic proof we simply present Figure 1, which illustrates strict inequality in (5) for $\alpha = 1.1$.

The proof is completed by taking

$$A = J_1, \qquad B = \begin{pmatrix} J_1 & \\ & J_2 \end{pmatrix},$$

with $1 < |\alpha| \leq \sqrt{2}$. We then have $\|B\| = |\alpha| > \|A\| = 1$ but $\|(zI - A)^{-1}\| = \|(zI - B)^{-1}\| = \|(zI - J_1)^{-1}\|$ for all $z \in \mathbb{C}$. Thus these matrices provide a counterexample to *(R)* $\Rightarrow$ *(P)* with $p(z) = z$. ∎

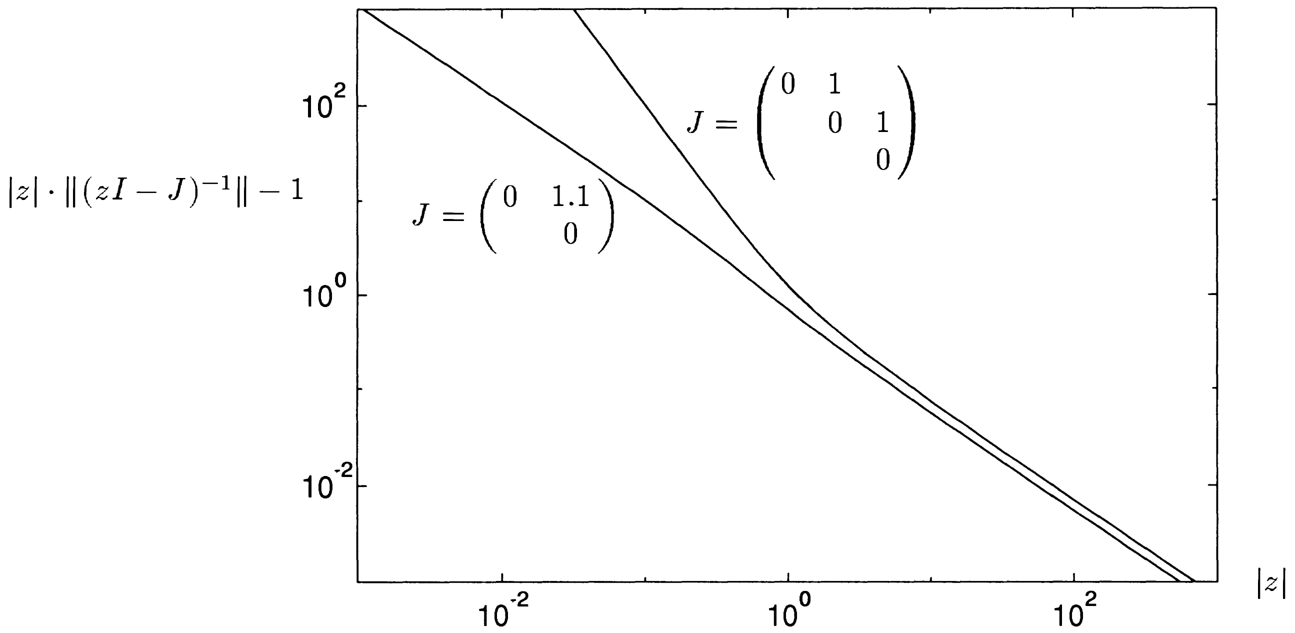

Figure 1. Illustration of strict inequality in (5); the ordinate is taken as $|z| \cdot \|(zI - J)^{-1}\| - 1$ instead of $\|(zI - J)^{-1}\|$ because this makes the behavior in both limits $|z| \to \infty$ and $|z| \to 0$ clear. The upper curve corresponds to a matrix with smaller norm, but larger resolvent norm $\|(zI - J)^{-1}\|$ for all nonzero $z \in \mathbb{C}$.

In the counterexample just given, the matrices A and B have different dimensions (3 and 5), but this is inessential. The dimensions could be made equal, for example, by padding A with zeros.

4. (R) $\Rightarrow$ (P) in the Frobenius norm

Our third theorem shows that in the Frobenius norm, the pseudospectra do determine the behavior of a matrix.

Theorem 3. *If* $\|\cdot\|$ *is the Frobenius norm, then (R) implies (P). In this case we have* $\dim(A) = \dim(B)$.

Proof. The Frobenius norm of $(zI - A)^{-1}$ is given by

$$\|(zI - A)^{-1}\|^2 = |z|^{-2} \operatorname{tr}\left[(I - z^{-1}A)^{-*}(I - z^{-1}A)^{-1})\right],$$

where $\operatorname{tr}(\cdot)$ denotes the trace. If $|z| > \rho(A)$, where $\rho(\cdot)$ denotes the spectral radius, then this expression can be expanded in a convergent Neumann series as

$$\|(zI - A)^{-1}\|^2 = |z|^{-2} \operatorname{tr}\left((\sum_{k=0}^{\infty} \bar{z}^{-k} A^{*^k}) \cdot (\sum_{k=0}^{\infty} z^{-k} A^k) \right)$$

$$= |z|^{-2} \operatorname{tr}\left(\sum_{\ell=0}^{\infty} \sum_{k=0}^{\ell} \bar{z}^{-k} z^{k-\ell} A^{*^k} A^{\ell-k} \right).$$

For $|z| > \max\{\rho(A), \rho(B)\}$, then, condition *(R)* implies that we can write

$$\sum_{\ell=0}^{\infty} \sum_{k=0}^{\ell} \bar{z}^{-k} z^{k-\ell} \operatorname{tr}(A^{*^k} A^{\ell-k}) = \sum_{\ell=0}^{\infty} \sum_{k=0}^{\ell} \bar{z}^{-k} z^{k-\ell} \operatorname{tr}(B^{*^k} B^{\ell-k}). \tag{6}$$

Taking the limit as $|z| \to \infty$ implies that the $\ell = 0$ terms of (6) must be equal,

$$\operatorname{tr}(I_A) = \operatorname{tr}(I_B),$$

where I_A and I_B denote the identities of dimensions $\dim(A)$ and $\dim(B)$, respectively. This implies $\dim(A) = \dim(B)$, as claimed. Now subtract these $\ell = 0$ term from both sides of (6), multiply by z, and take the limit as $|z| \to \infty$ again to obtain

$$\operatorname{tr}(A) + \frac{z}{\bar{z}} \operatorname{tr}(A^*) = \operatorname{tr}(B) + \frac{z}{\bar{z}} \operatorname{tr}(B^*).$$

Choosing two different values for $z/\bar{z}$, say, 1 and -1, we find

$$\operatorname{tr}(A) = \operatorname{tr}(B), \qquad \operatorname{tr}(A^*) = \operatorname{tr}(B^*).$$

Continuing in this way, suppose it has been shown that the traces involved in the terms with $\ell < m$ are all equal. To show that the traces involved in the $\ell = m$ terms are equal, first subtract the known equal terms from both sides of (6), then multiply by z^m and take the limit as $|z| \to \infty$ to obtain

$$\sum_{k=0}^{m} (\frac{z}{\bar{z}})^k \operatorname{tr}(A^{*^k} A^{m-k}) = \sum_{k=0}^{m} (\frac{z}{\bar{z}})^k \operatorname{tr}(B^{*^k} B^{m-k}).$$

Choosing $m+1$ different values on the unit circle for $z/\bar{z}$, say, $u_0, \dots, u_m$, gives a Vandermonde system of $m+1$ independent homogeneous equations for the $m+1$ quantities $\operatorname{tr}(A^{*^k} A^{m-k}) - \operatorname{tr}(B^{*^k} B^{m-k})$:

$$\begin{pmatrix} 1 & u_0 & \dots & u_0^m \\ 1 & u_1 & \dots & u_1^m \\ \vdots & \vdots & & \vdots \\ 1 & u_m & \dots & u_m^m \end{pmatrix} \begin{pmatrix} \operatorname{tr}(A^m) - \operatorname{tr}(B^m) \\ \operatorname{tr}(A^* A^{m-1}) - \operatorname{tr}(B^* B^{m-1}) \\ \vdots \\ \operatorname{tr}(A^{*^m}) - \operatorname{tr}(B^{*^m}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

It follows that $\operatorname{tr}(A^{*^k}A^\ell) = \operatorname{tr}(B^{*^k}B^\ell)$ for all k and ℓ, and this implies $\|p(A)\| = \|p(B)\|$ for all p since

$$\|p(A)\|^2 = \sum_k \sum_\ell \bar{c}_k c_\ell \operatorname{tr}(A^{*^k}A^\ell), \quad \|p(B)\|^2 = \sum_k \sum_\ell \bar{c}_k c_\ell \operatorname{tr}(B^{*^k}B^\ell),$$

if $p(x) = \sum_k c_k x^k$. ∎

Unfortunately, in the Frobenius norm the pseudospectra, defined by (1), are not characterized by (3). Even for the simplest example, where A is the zero matrix of dimension N, $\Lambda_\epsilon(A)$ is the closed disk about the origin of radius $\sqrt{N}\epsilon$, whereas the set defined in (3),

$$S_\epsilon = \{z \in \mathbb{C} : \; z \in \Lambda(E) \text{ for some } E \text{ with } \|E\| \le \epsilon\},$$

is the disk of radius ϵ. In general, for any A and ϵ we have

$$S_\epsilon^{(F)}(A) \;\subseteq\; \Lambda_\epsilon^{(2)}(A) \;\subseteq\; \Lambda_\epsilon^{(F)}(A),$$

where the superscripts indicate Frobenius or 2-norm. The first inequality follows from $\|E\|_F \ge \|E\|_2$ for the perturbation matrix E, and the second from $\|(zI - A)^{-1}\|_F \ge \|(zI - A)^{-1}\|_2$.

5. Discussion

The problem of relating information in the complex plane to the behavior of a matrix or linear operator A is an old one. Many results are known, of which the most important connect the spectrum of A with the asymptotic behavior of A^n and e^{tA} as $n, t \to \infty$, and the numerical range of A with the initial behavior of A^n and e^{tA} as $n, t \to 0$ [1]. Since the spectrum is determined by the ϵ-pseudospectra in the limit $\epsilon \to 0$ and the numerical range by the (2-norm) ϵ-pseudospectra in the limit $\epsilon \to \infty$, these two examples can both be regarded as special cases of pseudospectral information [4]. In effect this paper has considered the question of how much more can be gained if one knows $\Lambda_\epsilon(A)$ for finite values of ϵ in addition to the limits $\epsilon \to 0$ and $\epsilon \to \infty$.

A related collection of results is associated with von Neumann's theory of spectral sets, described for example in [9].

Since the 2-norm is more useful for most applications than the Frobenius norm, we regard the main result of this paper as negative: exact information about matrix behavior cannot be inferred from the norm of the resolvent (Theorem 2). It remains to be seen, however, whether the gap between these two kinds of information may be quantifiable. For example, the gap between the norms of powers $\|A^n\|$ and the bound provided by the Kreiss Matrix Theorem is known to be linear in the dimension of the matrix—remarkably small, in view of the exponential factors that appear in the process of taking powers. It would be interesting to know whether analogous results may hold for more general functions $p(A)$.

References

[1] J. L. M. van Dorsselaer, J. F. B. M. Kraaijevanger, and M. N. Spijker, *Linear stability analysis in the numerical solution of initial value problems,* Acta Numerica 1993 (1993), 199–237.

[2] R. W. Freund, G. H. Golub, and N. M. Nachtigal, *Iterative solution of linear systems,* Acta Numerica 1992 (1992), 57–100.

[3] A. Greenbaum and L. N. Trefethen, *GMRES/CR and Arnoldi/Lanczos as matrix approximation problems,* SIAM J. Sci. Comp., to appear.

[4] D. J. Higham and L. N. Trefethen, *Stiffness of ODEs,* BIT, to appear.

[5] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis,* Cambridge University Press, Cambridge, 1991.

[6] T. Kato, *Perturbation Theory for Linear Operators,* Springer-Verlag, New York, 1976.

[7] H.-O. Kreiss, *Über die Stabilitätsdefinition für Differenzengleichungen die partielle Differentialgleichungen approximieren,* BIT 2 (1962), 153–181.

[8] R. D. Richtmyer and K. W. Morton, *Difference Methods for Initial-Value Problems,* Wiley-Interscience, New York, 1967.

[9] F. Riesz and B. Sz.-Nagy, *Functional Analysis,* Frederick Ungar, New York, 1955.

[10] L. N. Trefethen, *Pseudospectra of matrices,* in D. F. Griffiths and G. A. Watson, eds., Numerical Analysis 1991, Longman, 1992.

[11] E. Wegert and L. N. Trefethen, *From the Buffon needle problem to the Kreiss matrix theorem,* Amer. Math. Monthly, to appear.