[2] A. Greenbaum and L. Gurvits, *Max-Min Properties of Matrix Factor Norms*, SIAM Jour. Sci. Comput., to appear.

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

[4] A. Greenbaum and L. N. Trefethen, *Do the pseudospectra of a matrix determine its behavior?*, to appear.

[5] P. Henrici, *Bounds for iterates, inverses, spectral variation and fields of values of non-normal matrices*, Num. Math. 4, (1962), 24-40.

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

[7] W. A. Joubert, *A Robust GMRES-based Adaptive Polynomial Preconditioning Algorithm for Nonsymmetric Linear Systems*, SIAM Jour. Sci. Comput., to appear.

[8] H. O. Kreiss, *Über die Stabilitätsdefinition für Differenzengleichungen die partielle Differenzialgleichungen Approximieren*, BIT 2, (1962), 153-181.

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

[10] M. N. Spijker, *On a Conjecture by LeVeque and Trefethen Related to the Kreiss Matrix Theorem*, BIT 31, (1991), 551-555.

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

[12] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford University Press, New York, 1965.

$\lambda_1, \ldots, \lambda_m$ and the matrix $\mathcal{W}$ are poorly determined, yet for most polynomials p, the Frobenius norm of $p(A)$ is well-determined. From formula (7) it follows that the quantities $\varrho^* \mathcal{W} \varrho$ are well-determined. For matrices that are close to nondiagonalizable matrices, the Frobenius norm of $p(A)$ could probably be determined from some fixed set of quantities of this form, that would be less sensitive to small changes in A than the eigenvalues and the matrix $\mathcal{W}$ itself.

In this paper we have concentrated on the Frobenius norm, because that is a norm for which we could give an explicit expression for the ratio

$$\sup_p \ \|p(A)\|/\|p(\Lambda)\|. \tag{24}$$

In applications, it is more often the 2-norm that is of interest, and while these norms differ by no more than a factor of $\sqrt{n}$, this factor can be significant for large matrices. Additionally, it means that these results cannot be applied to general linear operators. It is an open question whether an explicit expression can be given for the quantity in (24), when the norm there is the 2-norm, and whether the minimal 2-norm condition number $\min_V \kappa_2(V)$ is close to this optimal constant, when n is large. Based on the analysis here, we can state that

$$\sup_p \|p(A)\|_2/\|p(\Lambda)\|_2 \ \leq \ \min_V \kappa_2(V) \ \leq \ n^{3/2} \ \sup_p \|p(A)\|_2/\|p(\Lambda)\|_2,$$

since

$$\min_V \kappa_2(V) \leq \min_V \kappa_F(V) \leq n \ \gamma_F(A) = n \ \sup_p \frac{\|p(A)\|_F}{\|p(\Lambda)\|_F} \leq n^{3/2} \ \sup_p \frac{\|p(A)\|_2}{\|p(\Lambda)\|_2}.$$

Acknowledgments

The author thanks Nick Higham, Zdenek Strakos, and Nick Trefethen for helpful comments on the manuscript.

References

[1] F. Chatelin and V. Fraysee, *QUALITATIVE COMPUTING: Elements of a theory for finite precision computation*, CERFACS, Thomson-CSF, Orsay, France, June, 1993.

eigenvalue multiplicities $|S_k|$, $k = 1, \ldots, m$, and the row sums of $T^{1/2}\hat{\mathcal{W}}T^{1/2}$ are the eigenvalue multiplicities $|T_k|$, $k = 1, \ldots, m$, it follows that $|S_k| = |T_k|$, $k = 1, \ldots, m$, and $\mathcal{W} = \hat{\mathcal{W}}$. $\square$

Corollary. The eigenvalues of a diagonalizable matrix A and their multiplicities, together with the strict lower triangle of the matrix $\mathcal{W}$ defined in (6), are necessary and sufficient information to determine $\|p(A)\|_F$, for any polynomial p.

Proof: Since $\mathcal{W}$ can be determined from its strict lower triangle and the eigenvalue multiplicities, it follows that the information in the corollary is equivalent to that in (21). $\square$

If A has n distinct eigenvalues, the preceding corollary lists $\frac{n(n+1)}{2}$ numbers – the eigenvalues of A and the strict lower triangle of $\mathcal{W}$ – that completely determine $\|p(A)\|_F$ for all polynomials p and that *must* be known in order to determine $\|p(A)\|_F$. This is in contrast to a Schur form of A, which also contains $\frac{n(n+1)}{2}$ pieces of information, but is *not* necessary in order to determine $\|p(A)\|_F$. (The fact that the Schur form is not necessary can be seen, for example, from our construction of example 1 in Section 2.1. There we specified the matrix V^*V and hence were able to determine $\mathcal{W}$. But there were many choices for the eigenvector matrix V, not all of which gave rise to matrices A that were unitarily similar to each other.) An interesting open question is whether the information in the corollary can be determined from some smaller set of numbers. Not all sets of numbers are possible as the strict lower triangle of a matrix of the form of $\mathcal{W}$; i.e., a matrix of the form $C \circ C^{-T}$, where C is Hermitian positive definite. It is not known whether a smaller set of numbers might provide the same information. An alternative set of necessary and sufficient information is described in [4], where it is shown that the set of resolvent norms $\|(zI - A)^{-1}\|_F$, for all complex numbers z, provide necessary and sufficient information to determine $\|p(A)\|_F$ for all p.

The preceding theorem and corollary apply only to diagonalizable matrices. The set of diagonalizable matrices is dense in the set of all square matrices, but as the matrix A approaches a nondiagonalizable matrix, some elements of the matrix $\mathcal{W}$ become infinite. In this case, the eigenvalues

$$\hat{\mathcal{W}}_{k\ell} = |T_k|^{-1/2}|T_\ell|^{-1/2} \sum_{i\in T_k}\sum_{j\in T_\ell} \hat{W}_{ij}, \quad k,\ell = 1,\ldots,q.$$

A necessary and sufficient condition to ensure $\|p(A)\|_F = \|p(B)\|_F$ for all polynomials p is that

$$m = q, \quad \lambda_{S_k} = \hat{\lambda}_{T_k} \text{ and } |S_k| = |T_k|,\ k = 1,\ldots,m, \quad \text{and} \quad \mathcal{W} = \hat{\mathcal{W}}. \quad (21)$$

Proof: Sufficiency is clear, since it follows from (7) that

$$\|p(A)\|_F^2 = \varrho^* \mathcal{W} \varrho, \qquad \|p(B)\|_F^2 = \hat{\varrho}^* \hat{\mathcal{W}} \hat{\varrho}, \quad (22)$$

where

$$\varrho = S^{1/2} \begin{pmatrix} p(\lambda_{S_1}) \\ \vdots \\ p(\lambda_{S_m}) \end{pmatrix}, \qquad \hat{\varrho} = T^{1/2} \begin{pmatrix} p(\hat{\lambda}_{T_1}) \\ \vdots \\ p(\hat{\lambda}_{T_q}) \end{pmatrix},$$

$S = \mathrm{diag}(|S_1|,\ldots,|S_m|)$, and $T = \mathrm{diag}(|T_1|,\ldots,|T_q|)$. If (21) holds then the two expressions in (22) are identical.

To prove necessity, first note that if the distinct eigenvalues of A and B are different; i.e., if A has an eigenvalue that is not an eigenvalue of B or vice versa, then the minimal polynomial of one of these matrices will map that matrix to zero but not the other one, and so A and B will not satisfy $\|p(A)\|_F = \|p(B)\|_F$ for this minimal polynomial p. Hence if $\|p(A)\|_F = \|p(B)\|_F$ for all p, then $m = q$ and $\lambda_{S_k} = \hat{\lambda}_{T_k}$, $k = 1,\ldots,m$. Equations (22) can be written in the form

$$\|p(A)\|_F^2 = \sigma^*(S^{1/2}\mathcal{W}S^{1/2})\sigma, \qquad \|p(B)\|_F^2 = \hat{\tau}^*(T^{1/2}\hat{\mathcal{W}}T^{1/2})\hat{\tau}, \quad (23)$$

where $\sigma = S^{-1/2}\varrho$ and $\hat{\tau} = T^{-1/2}\hat{\varrho}$. Since $\sigma = \hat{\tau}$, equality of the two expressions in (23) implies

$$\sigma^*(S^{1/2}\mathcal{W}S^{1/2} - T^{1/2}\hat{\mathcal{W}}T^{1/2})\sigma = 0,$$

for all m-vectors σ. Since $S^{1/2}\mathcal{W}S^{1/2} - T^{1/2}\hat{\mathcal{W}}T^{1/2}$ is Hermitian this implies that $S^{1/2}\mathcal{W}S^{1/2} = T^{1/2}\hat{\mathcal{W}}T^{1/2}$. Since the row sums of $S^{1/2}\mathcal{W}S^{1/2}$ are the to the Frobenius norm of that function of the eigenvalues. We have also pointed out cases in which this simple bound cannot provide a reasonable estimate. In such cases more or different information is needed.

An interesting open question is how much information must one know about an n by n matrix A in order to precisely determine the norm of $p(A)$ for any given polynomial p. For normal matrices, the n eigenvalues are sufficient to determine $\|p(A)\|$ for any unitarily invariant norm, since $\|p(A)\| = \|p(\Lambda)\|$. For general matrices, the $\frac{n(n+1)}{2}$ entries of a Schur triangular form are sufficient to determine $\|p(A)\|$ for any unitarily invariant norm, but they are not necessary in order to determine the Frobenius norm of $p(A)$, as we will show. If the matrix A is diagonalizable, then it is clear from equality (7) that knowledge of the distinct eigenvalues of A, their multiplicities, and the matrix $\mathcal{W}$ is sufficient to determine $\|p(A)\|_F$ for any polynomial p. Since $\mathcal{W}$ is Hermitian and maps the vector $(|S_1|^{1/2}, \ldots, |S_m|^{1/2})^T$ into itself, $\mathcal{W}$ is completely determined by its strict lower triangle and the multiplicities of the eigenvalues. Hence, if A has n distinct eigenvalues, then the n eigenvalues of A together with the $\frac{n(n-1)}{2}$ entries of the strict lower triangle of $\mathcal{W}$ provide sufficient information to determine $\|p(A)\|_F$ for all p. The following theorem and corollary show that this information is also necessary.

Theorem. Let A and B be n by n diagonalizable matrices with eigendecompositions $A = V\Lambda V^{-1}$ and $B = \hat{V}\hat{\Lambda}\hat{V}^{-1}$. Define W and $\hat{W}$ by

$$W = (V^*V) \circ (V^*V)^{-T}, \quad \hat{W} = (\hat{V}^*\hat{V}) \circ (\hat{V}^*\hat{V})^{-T}$$

where $\circ$ denotes the Hadamard product. Assume that A has m distinct eigenvalues $\lambda_{S_1}, \ldots, \lambda_{S_m}$, where S_k denotes the set of indices of eigenvalues equal to λ_{S_k} and $|S_k|$ denotes the number of indices in S_k. Assume that B has q distinct eigenvalues $\hat{\lambda}_{T_1}, \ldots, \hat{\lambda}_{T_q}$, where T_k and $|T_k|$ are defined analogously and where $\hat{\lambda}_{T_1}, \ldots \hat{\lambda}_{T_q}$ are ordered in the same way as $\lambda_{S_1}, \ldots \lambda_{S_m}$; say, in increasing order by absolute value, with eigenvalues of the same absolute value ordered in increasing order of imaginary parts. Define $\mathcal{W}$ and $\hat{\mathcal{W}}$ by

$$\mathcal{W}_{k\ell} = |S_k|^{-1/2}|S_\ell|^{-1/2} \sum_{i\in S_k} \sum_{j\in S_\ell} W_{ij}, \quad k, \ell = 1, \ldots, m,$$

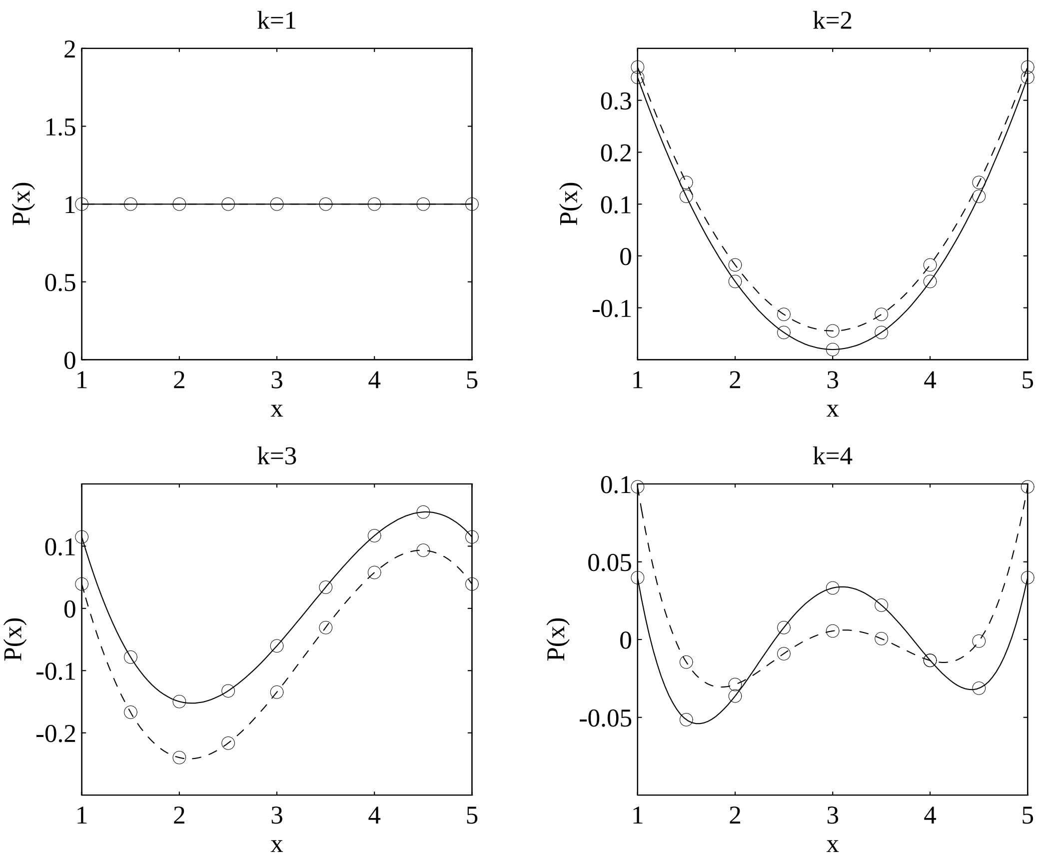

Figure 1: Minimum Frobenius Norm Polynomials $P_k^{(F)}$ (solid lines) and GMRES Polynomials $P_k^{(r^0)}$ for a random initial residual r^0 (dashed lines).

Note that in both of these examples the matrix V of eigenvectors of A is highly ill-conditioned, but it has only one small singular value and one small eigenvalue. The difference in the two examples cannot be explained in terms of the singular values or eigenvalues of V, but it is accounted for by the difference in the eigenvalues (which are the singular values) of W. In the first example W has only one large eigenvalue, while in the second example all of the eigenvalues of W (except the eigenvalue 1) are large.

4 Conclusions and Further Remarks

We have shown how the Frobenius norm of a function of a matrix A can often be estimated in terms of the eigenvalues of A and the sharp bound $\gamma_F(A)$ on the ratio of the Frobenius norm of any analytic function of A

where e_i is the vector with a 1 in position i and zeros elsewhere. In this case, the only large element (in fact, the only nonzero element) in the strict lower triangle of W is the $(n,1)$ element, which is $-1/\delta^2$. The polynomials $P_k^{(F)}$, $k = 1, 2, \ldots$, essentially minimize $\|p(\Lambda)\|_F$ subject to the constraint that $p(\lambda_1) = p(\lambda_n)$. This means that $P_1^{(F)}$ will be almost identically equal to 1, and that the higher degree polynomials $P_k^{(F)}$, $k > 1$, will be as small as possible on the set of eigenvalues of A, subject to the constraint that $P_k^{(F)}(1) = P_k^{(F)}(5)$. Figure 1 shows the polynomials $P_1^{(F)}$ through $P_4^{(F)}$ (solid lines) and the GMRES polynomials $P_1^{(r^0)}$ through $P_4^{(r^0)}$ for a random initial residual r^0 (dashed lines), using a matrix A of order $n = 9$. The circles mark the values of the polynomials at the eigenvalues of A. The polynomial that minimizes $\|p(\Lambda)\|_F$ subject to the constraint that $p(\lambda_1) = p(\lambda_n)$ is also plotted, but it is indistinguishable from P_k^F. While the GMRES polynomials differ from the minimum Frobenius norm polynomials, they are of the same order of magnitude, and each polynomial has nearly equal values at $x = 1$ and $x = 5$. (It is not entirely clear why the GMRES polynomials should satisfy this constraint, but they appear to do so for a variety of random initial residuals, and this phenomenon is currently being investigated.)

In contrast, consider a matrix A with the same eigenvalues equally spaced between 1 and 5, but with eigenvectors

$$e_1, \ldots e_{n-1}, \ \sum_{i=1}^{n-1} e_i + \delta e_n, \quad \delta = 10^{-5}.$$

In this case, every entry of the n^{th} row of the strict lower triangle of W is equal to $-1/\delta^2$. This forces the polynomials $P_k^{(F)}$ to have nearly equal values at all the eigenvalues of A, and consequently these polynomials are all almost identically equal to 1, for $k = 1, \ldots, n-1$. The polynomials P_k that minimize the 2-norm of $p(A)$ display similar behavior, and the GMRES polynomials $P_k^{(r^0)}$ associated with a random initial vector r^0 also show little deviation from 1 until the degree k is equal to n. This means that the GMRES algorithm (or any algorithm for which the k^{th} residual vector r^k is of the form $p(A)r^0$, where $p \in \mathcal{P}_k$) makes almost no progress toward solving a linear system with this coefficient matrix, until step n, when the exact solution is obtained.

3.1 Examples

In some cases the constraints described in (18) take on a particularly simple form. For ease of notation we will assume that the matrix A has n distinct eigenvalues so that equality (2) applies, but the modifications for multiple eigenvalues are straightforward. If the matrix W happens to have all real entries then equality (2) can be written in the form

$$\begin{aligned} \|p(A)\|_F^2 &= \sum_{i=1}^n \overline{p(\lambda_i)} \sum_{j=1}^n W_{ij}\ p(\lambda_j) \\ &= \sum_{i=1}^n |p(\lambda_i)|^2 \ + \ \sum_{i=1}^n \overline{p(\lambda_i)} \sum_{j=1}^n W_{ij}\ (p(\lambda_j) - p(\lambda_i)) \\ &= \|p(\Lambda)\|_F^2 \ + \ \sum_i \sum_{j<i} \overline{p(\lambda_i)}\ W_{ij}\ (p(\lambda_j) - p(\lambda_i)) \ + \\ & \quad \sum_j \sum_{i<j} \overline{p(\lambda_i)}\ W_{ij}\ (p(\lambda_j) - p(\lambda_i)) \\ &= \|p(\Lambda)\|_F^2 \ + \ \sum_i \sum_{j<i} |p(\lambda_i) - p(\lambda_j)|^2\ (-W_{ij}) \end{aligned} \tag{20}$$

If the elements of the strict lower triangle of W happen to be nonpositive – which would be the case, for instance, if V^*V or $(V^*V)^{-1}$ were a Stieltjes matrix – then (20) shows that the squared Frobenius norm of $p(A)$ is equal to the squared Frobenius norm of $p(\Lambda)$ plus a weighted sum of squares of differences of the values of p at different eigenvalues of A. If just a few of the elements $-W_{ij}$, $j < i$, are large, then $P_k^{(F)}$ will essentially minimize $\|p(\Lambda)\|_F$ over the set of polynomials p in $\mathcal{P}_k$ such that $p(\lambda_i) = p(\lambda_j)$ for all i, j such that $-W_{ij}$ is large. If many elements $-W_{ij}$, $j < i$, are large, then $P_k^{(F)}$ will be forced to have nearly equal values at all or many of the eigenvalues of A and since its value at the origin is 1, $P_k^{(F)}$ will be almost identically equal to 1 until the degree k is close to n.

As a simple example of the first situation, let A have real eigenvalues equally distributed between $\lambda_1 = 1$ and $\lambda_n = 5$, and let the eigenvectors of A be

$$e_1, \ldots e_{n-1},\ e_1 + \delta e_n, \quad \delta = 10^{-5},$$

where $\omega_1, \ldots, \omega_q$ are the eigenvectors of $\mathcal{W}$ corresponding to the large eigenvalues. Usually this means that the polynomials that minimize $\|p(A)\|_F$ over the set $\mathcal{P}_{k+q\pm}$, containing polynomials of degree $k+q$ plus or minus a few, will achieve approximately the same minimum as the polynomials that minimize $\|p(\Lambda)\|_F$ over $\mathcal{P}_k$:

$$\|P^F_{k+q\pm}(A)\|_F \approx \min_{p \in \mathcal{P}_k} \|p(\Lambda)\|_F.$$

- If the matrix $\mathcal{W}$ has many eigenvalues ranging over many orders of magnitude, then more information about $\mathcal{W}$, or different information about the matrix A, is needed to describe the behavior of the polynomials $P_k^{(F)}$.

Let P_k be the polynomial in $\mathcal{P}_k$ that minimizes $\|p(A)\|_2$ over all polynomials p in $\mathcal{P}_k$. Then the 2-norm of $P_k(A)$ can be related to the Frobenius norm of $P_k^{(F)}(A)$ as follows:

$$n^{-1/2}\,\|P_k^{(F)}(A)\|_F \le n^{-1/2}\,\|P_k(A)\|_F \le \|P_k(A)\|_2 \le \|P_k^{(F)}(A)\|_2 \le \|P_k^{(F)}(A)\|_F.$$

Since $\|P_k(A)\|_2$ differs from $\|P_k^{(F)}(A)\|_F$ by no more than a factor of $\sqrt{n}$, the previous qualitative statements about $\|P_k^F(A)\|_F$ apply also to $\|P_k(A)\|_2$, provided $\sqrt{n}$ is not too large.

In solving a system of linear equations $Ax = b$, the GMRES algorithm generates approximate solutions x^k, $k = 1, 2, \ldots$, such that the residual vectors $r^k = b - Ax^k$ satisfy

$$r^k = P_k^{(r^0)}(A) r^0,$$

where $P_k^{(r^0)}$ minimizes $\|p(A)r^0\|_2$ over all polynomials p in $\mathcal{P}_k$. It follows that the 2-norm of r^k is less than or equal to that of $P_k(A)r^0$, where P_k is the polynomial that minimizes $\|p(A)\|_2$:

$$\|r^k\|_2 \;\le\; \|P_k(A)r^0\|_2 \;\le\; \|P_k(A)\|_2 \cdot \|r^0\|_2. \tag{19}$$

It is believed (but not proved) [2,3,7] that for each k there is an initial residual r^0 for which equality holds in (19). If this is the case, then the qualitative statements about the norms of the polynomials $P_k^{(F)}$ and P_k can also be thought of as describing the norms of the GMRES residuals at step k, assuming the worst possible initial residual of norm one.

varied only between $3.4e+5$ and $1.3e+6$, while, as before, $\gamma_F(A) = 3.5e+6$.

3 Polynomials that Minimize $\|p(A)\|$

As an example of a class of polynomials for which the vectors ϱ can be expected to be orthogonal to the eigenvectors of $\mathcal{W}$ corresponding to the largest eigenvalues, consider the sequence of polynomials $P_k^{(F)}$, $k = 1, 2, \ldots$, where $P_k^{(F)}$ minimizes $\|p(A)\|_F$ over all polynomials p in $\mathcal{P}_k = \{k^{th}$ degree polynomials with value one at the origin$\}$. Based on equality (16), we can make the following *qualitative* statements about the behavior of the polynomials $P_k^{(F)}$, $k = 1, 2, \ldots$:

- If all eigenvalues of $\mathcal{W}$ are close to 1 then $P_k^{(F)}$ is essentially like the polynomial that minimizes $\|p(\Lambda)\|_F$ and
$$\|P_k^{(F)}(A)\|_F \approx \|P_k^{(F)}(\Lambda)\|_F \approx \min_{p \in \mathcal{P}_k} \|p(\Lambda)\|_F.$$
- If all eigenvalues of $\mathcal{W}$ are very large (except the eigenvalue 1), then, until the degree k is great enough so that $\mathcal{P}_k$ contains a polynomial that is correspondingly small at all the eigenvalues of A (that is, until $\|\varrho\|_2$ can be chosen to be on the order of $\|\mathcal{W}\|_2^{-1/2}$ or smaller), the vectors ϱ will be almost orthogonal to the eigenvectors of $\mathcal{W}$ corresponding to large eigenvalues, which means they will be almost parallel to the vector ω_m. This means the polynomial $P_k^{(F)}$ will be approximately equal to 1 and $\|P_k^{(F)}(A)\| \approx \|I\|$.
- If $\mathcal{W}$ has just a few very large eigenvalues of roughly the same order of magnitude, and the other eigenvalues are of moderate size, then the polynomials $P_k^{(F)}$ will be such that the corresponding vectors ϱ are almost orthogonal to the eigenvectors of $\mathcal{W}$ corresponding to the large eigenvalues. That is, the polynomials $P_k^{(F)}$ will essentially solve the constrained least squares problem
$$\min_{\substack{p \in \mathcal{P}_k \\ \varrho \perp \omega_1, \ldots, \omega_q}} \|p(\Lambda)\|_F = \|P_k^{(F)}(A)\|_F \tag{18}$$

k	$\|A^k\|_F/\|\Lambda^k\|_F$
1	9.4e+0
2	8.7e+1
3	7.9e+2
4	7.1e+3
5	6.1e+4
6	5.0e+5
7	3.5e+6
8	1.0e+0

Table 1: Ratios of norms of powers of A and powers of Λ

ϱ in (7) is parallel to the eigenvector of $\mathcal{W}$ corresponding to the eigenvalue one. In fact, it turns out that for this matrix the vectors ϱ obtained from the powers of Λ are almost equal to the eigenvectors of $\mathcal{W}$! The numbers in Table 1 are the square roots of the corresponding eigenvalues of $\mathcal{W}$.

This is an example for which the eigenvalues of A, together with a single number bounding the ratio between the norm of a function of A and the norm of that function of Λ, cannot provide reasonable estimates of the norms of powers of A. More information about the matrix $\mathcal{W}$ or different information about A is needed to obtain such estimates. Equality (16) shows why this is so. It is because the eigenvalues of $\mathcal{W}$ vary over orders of magnitude and because the vectors ϱ associated with the powers of Λ are highly correlated with the eigenvectors of $\mathcal{W}$.

Example 3. In this example, we consider a matrix with the same eigenvectors as the matrix in example 2, but with slightly different eigenvalues. Instead of the eigenvalues of example 2, which are uniformly distributed around the circle of radius .1 about the origin, we chose the eigenvalues to be randomly distributed on this same circle. This small change to the eigenvalues resulted in a much larger change to the matrix A, whose Frobenius norm was now on the order of $\gamma_F(A)$ times the Frobenius norm of Λ. Changing the eigenvalues in this way destroyed the high correlation between the vectors ϱ and eigenvectors of the matrix $\mathcal{W}$. In this case the upper bound in (11) provides a much better estimate of the ratios of the norms of the powers of A to the norms of the powers of Λ. Now these ratios of the columns of V (which are the square roots of the diagonal elements of V^*V) are all equal and the 2-norms of the columns of V^{-*} (which are the square roots of the diagonal elements of $(V^*V)^{-1}$) are all equal, the inequality (14) is also an equality for this problem, and so the first upper bound in (13) is met. In this case, then, the bound (1) becomes

$$\|p(A)\|_F \;\leq\; \sqrt{2}\; n \sqrt{1 - \frac{1}{n+1}}\; \|p(\Lambda)\|_F \;<\; \sqrt{2}\; n\; \|p(\Lambda)\|_F,$$

while the sharp bound (11) shows

$$\|p(A)\|_F \;\leq\; \sqrt{2 - \frac{1}{n+1}}\; \|p(\Lambda)\|_F \;<\; \sqrt{2}\; \|p(\Lambda)\|_F.$$

Taking $n = 40$, choosing V to be the upper triangular Cholesky factor of (17), and taking the eigenvalues of A to be randomly distributed in the unit circle, with the largest eigenvalue on the unit circle, we computed the powers A^k and Λ^k, for $k = 1, \ldots, n$. In this case, the parameter $\gamma_F(A)$ in (11) was 1.41, and the ratios $\|A^k\|_F \;/\; \|\Lambda^k\|_F$ ranged from 1.39 to 1.41, showing very close agreement with the upper bound in (11).

Example 2. As an example in which the simple bound (11) is not adequate to explain the behavior of powers of a matrix, we chose A to be a slightly perturbed Jordan block:

$$A = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & & & 1 \\ \epsilon & 0 & 0 & \ldots & 0 \end{pmatrix}$$

The eigenvalues of this matrix are the n^{th} roots of ϵ: $\epsilon^{1/n} \cdot \exp(2\pi i \ell / n)$, $\ell = 0, 1, \ldots, n-1$. Taking $n = 8$ and $\epsilon = 10^{-8}$, the parameter $\gamma_F(A)$ was found to be $3.5e+6$. For comparison, the bound $\min_V \kappa_F(V)$ of (1) was computed using formula (12) and found to be $1.0e+7$. The ratios $\|A^k\|_F \;/\; \|\Lambda^k\|_F$ are given in Table 1, for $k = 1, \ldots, n$.

Note that for $k = 7$, the bound (11) is met. Yet for $k = 8$, the ratio is 1. This is because Λ^8 has all equal diagonal elements, $1.0e-8$, so the vector

$\mathcal{W}$, or some other kind of information about A, it is impossible to say where, between the upper and lower bound in (11), the actual norm of $p(A)$ will lie. Unfortunately, this type of example arises very often in practice. If A has elements of size $O(1)$ and is very close to a nondiagonalizable matrix, then the matrix $\mathcal{W}$ associated with A will have very large norm, and the vector of eigenvalues of A will be nearly orthogonal to the eigenvectors of $\mathcal{W}$ corresponding to the largest eigenvalues.

2.1 Examples

Example 1. An example in which the upper bounds (11) and (1) differ by the largest possible factor, given in (13), is the following. Let A be an n by n matrix with n distinct eigenvalues and assume that an eigenvector matrix V of A satisfies

$$V^*V = I + uu^* \tag{17}$$

where u is the n-vector of all ones. (Note that we need not specify an eigenvector matrix V explicitly. It could be the symmetric square root of the positive definite matrix in (17) or it could be the Cholesky factor, for example.) Then, from the Sherman-Morrison formula, we have

$$(V^*V)^{-1} = I - \frac{1}{1 + u^*u}\, uu^*.$$

A simple computation shows that $\kappa_F^2(V) = 2n^2(1 - \frac{1}{n+1}) \approx 2n^2$, for large n, and, because the diagonal elements of both V^*V and $(V^*V)^{-1}$ are constant, it can also be seen that this is the minimum over all diagonal matrices D of $\kappa_F^2(VD) = \mathrm{tr}(\bar{D}V^*VD) \cdot \mathrm{tr}(D^{-1}(V^*V)^{-1}\bar{D}^{-1})$. Thus, this is the minimum Frobenius norm condition number for any eigenvector matrix of A. The matrix $\mathcal{W}$ is given by

$$\mathcal{W} = \left(2 - \frac{1}{n+1}\right) I - \frac{1}{n+1}\, uu^*,$$

and all of the eigenvalues of this matrix, except the eigenvalue 1 corresponding to the eigenvector u, are equal to $2 - \frac{1}{n+1}$. Consequently, the inequality in (15) is an equality for this problem. Additionally, since the 2-norms large. In this case, one can use equality (7) to characterize those matrices A and classes of polynomials p for which the upper bound in (11) will also be a reasonable *estimate* of $\|p(A)\|_F$ and those for which it might be a large overestimate. Let $\omega_1, \ldots, \omega_m$ denote the normalized eigenvectors of $\mathcal{W}$ and let $\alpha_1 \geq \ldots \geq \alpha_{m-1} \geq \alpha_m = 1$ denote the corresponding eigenvalues. Equality (7) can be written in the form

$$\begin{aligned} \|p(A)\|_F^2 &= \varrho^* \varrho + \varrho^* (\mathcal{W} - I) \varrho \\ &= \|p(\Lambda)\|_F^2 \cdot \left(1 + \sum_{j=1}^{m-1} (\alpha_j - 1) \left| \frac{\omega_j^* \varrho}{\|\varrho\|_2} \right|^2 \right) . \end{aligned} \tag{16}$$

If *all* eigenvalues of $\mathcal{W}$ (except the eigenvalue 1) are very large, then the upper bound in (11) is a realistic estimate of $\|p(A)\|_F$ for any polynomial p such that a significant component of the vector ϱ is orthogonal to the vector $\omega_m = (|S_1|^{1/2}, \ldots, |S_m|^{1/2})^T$. Unless p has very nearly equal values at the eigenvalues of A, which means $p(A)$ is approximately equal to a multiple of the identity, the vector ϱ will have such a component. In this case, then, for most interesting classes of polynomials p, the upper bound in (11) will provide a reasonable estimate of $\|p(A)\|_F$.

Suppose $\mathcal{W}$ has some very large eigenvalues and some eigenvalues on the order of 1. If the class of polynomials under consideration is such that the vectors ϱ have nonnegligible components in the direction of the largest eigenvector of $\mathcal{W}$, then the upper bound in (11) will provide a reasonable estimate of $\|p(A)\|_F$. In contrast, if the vectors ϱ are almost orthogonal to the eigenvectors of $\mathcal{W}$ corresponding to the largest eigenvalues, then the upper bound in (11) will be a large overestimate of the actual norm of $p(A)$. For example, suppose that the polynomials being considered are of fairly low degree and that the Frobenius norm of the matrix A is much less than $\gamma_F(A)$ times the Frobenius norm of Λ. Then the vector $(\lambda_{S_1}, \ldots, \lambda_{S_m})^T$ *must be* nearly orthogonal to the eigenvectors of $\mathcal{W}$ corresponding to the largest eigenvalues. Similarly, if the Frobenius norms of $A^2, \ldots, A^k$ are much less than $\gamma_F(A)$ times the Frobenius norms of $\Lambda^2, \ldots, \Lambda^k$, then the vectors ϱ corresponding to *any* polynomial of degree k or less will also be almost orthogonal to the eigenvectors of $\mathcal{W}$ corresponding to the largest eigenvalues. In this case, without additional information about the matrix

matrix of A [12, p. 89]. Let $v_1, \ldots, v_n$ denote the columns of an eigenvector matrix V, and let $y_1, \ldots, y_n$ denote the columns of V^{-*}. Then $v_1, \ldots, v_n$ are right eigenvectors and $y_1, \ldots, y_n$ are left eigenvectors of A, and we have

$$\min_{\tilde{V}} \kappa_F(\tilde{V}) = \sum_{j=1}^{n} \|v_j\|_2 \cdot \|y_j\|_2, \tag{12}$$

where the minimum is over all eigenvector matrices $\tilde{V}$, and $\kappa_F(\tilde{V}) = \|\tilde{V}\|_F \cdot \|\tilde{V}^{-1}\|_F$. Using this expression we prove the following theorem relating $\gamma_F(A)$ and $\min_{\tilde{V}} \kappa_F(\tilde{V})$.

Theorem 2. For any n by n matrix A with n distinct eigenvalues,

$$\gamma_F(A) \;\le\; \min_{\tilde{V}} \kappa_F(\tilde{V}) \;\le\; \sqrt{n((n-1)\gamma_F^2(A)+1)} \;\le\; n\ \gamma_F(A), \tag{13}$$

where the minimum is over all eigenvector matrices $\tilde{V}$ of A, and $\kappa_F(\tilde{V}) = \|\tilde{V}\|_F \cdot \|\tilde{V}^{-1}\|_F$.

Proof: The first inequality in (13) follows immediately from Theorem 1 and inequality (1). To obtain the second inequality in (13), first note from expression (12) that

$$\min_{\tilde{V}} \kappa_F^2(\tilde{V}) = \left(\sum_{j=1}^{n} \|v_j\|_2 \cdot \|y_j\|_2 \right)^2 \le n \sum_{j=1}^{n} \|v_j\|_2^2 \cdot \|y_j\|_2^2. \tag{14}$$

The sum on the right in (14) is just the trace of W, and since the trace is equal to the sum of the eigenvalues and we know that W has one eigenvalue equal to 1, we can write

$$\sum_{j=1}^{n} \|v_j\|_2^2 \cdot \|y_j\|_2^2 \le (n-1)\ \gamma_F^2(A) + 1 \tag{15}$$

Combining (14) and (15) gives the desired result (13). $\Box$

If $\gamma_F(A)$ is not much greater than one, then expression (11) gives fairly tight upper and lower bounds on $\|p(A)\|_F$. If $\gamma_F(A)$ is very large, however, then there are polynomials p for which the ratio $\|p(A)\|_F / \|p(\Lambda)\|_F$ is very multiple of the identity), while the upper bound is attained for polynomials p for which the corresponding vector ϱ is parallel to the eigenvector of $\mathcal{W}$ corresponding to the largest eigenvalue.

For an arbitrary square matrix A, define $\gamma_F(A)$ by

$$\gamma_F(A) = \begin{cases} \|\mathcal{W}\|_2^{1/2} & \text{if A is diagonalizable} \\ +\infty & \text{otherwise} \end{cases} \tag{10}$$

We then have the following theorem:

Theorem 1. For any square matrix A with eigenvalue matrix Λ,

$$\|p(\Lambda)\|_F \ \leq \ \|p(A)\|_F \ \leq \ \gamma_F(A) \cdot \|p(\Lambda)\|_F, \tag{11}$$

for all polynomials p, and $\gamma_F(A)$ is the smallest number γ, depending only on A, such that $\|p(A)\|_F \leq \gamma \|p(\Lambda)\|_F$ for all p.

Proof: If A is diagonalizable, then (11) is just a restatement of (9), which we have already proved. If A is not diagonalizable, then the minimal polynomial of A has a nonlinear factor while that of Λ does not. If p is the minimal polynomial of Λ, then $p(\Lambda) = 0$ but $p(A) \neq 0$, so no finite number can bound the ratio $\|p(A)\|_F / \|p(\Lambda)\|_F$. $\square$

Note that $\gamma_F(A)$, and hence the ratio $\|p(A)\|_F / \|p(\Lambda)\|_F$, can be large even when the *departure from normality* based on the Schur form is small. For example, if A is given by

$$A = \begin{pmatrix} 1 & \epsilon \\ 0 & 1 + \epsilon^2 \end{pmatrix}$$

then $\gamma_F(A) = \sqrt{1 + 2/\epsilon^2}$, and if $p(x) = x - (1 + \epsilon^2/2)$ then $\|p(A)\|_F / \|p(\Lambda)\|_F = \gamma_F(A)$. Yet the departure from normality in the Frobenius norm is only ϵ. Bounds based on the departure from normality are usually absolute bounds on $\|p(A)\|^2 - \|p(\Lambda)\|^2$, whereas $\gamma(A) - 1$ and $\kappa(V) - 1$ bound the relative difference $(\|p(A)\| - \|p(\Lambda)\|) / \|p(\Lambda)\|$.

To see how the sharp bound in (11) compares with the bound in (1), we first note that if A has n distinct eigenvalues, then an explicit expression is known for the minimum Frobenius norm condition number of an eigenvector corresponding to the eigenvector $(|S_1|^{1/2}, \ldots, |S_m|^{1/2})^T$. Finally note that the matrix $\mathcal{W}$ is independent of the scaling of the eigenvectors and of the basis chosen to represent the subspace corresponding to a multiple eigenvalue. To see this, assume that the equal eigenvalues of A are ordered consecutively in Λ, and suppose the eigenvector matrix V is replaced by VB, where B is an invertible block diagonal matrix with m blocks of size $|S_1|, \ldots, |S_m|$. Then for $i \in S_k$ and $j \in S_\ell$, the (i,j) element of the matrix W in (3) becomes

$$
\begin{aligned}
W_{ij} &= (B^*V^*VB)_{ij} \cdot (B^{-1}(V^*V)^{-1}B^{-*})_{ji} \\
&= \left(\sum_{q \in S_k} \sum_{r \in S_\ell} B^*_{iq}(V^*V)_{qr} B_{rj} \right) \cdot \left(\sum_{s \in S_k} \sum_{t \in S_\ell} B^{-1}_{jt}(V^*V)^{-1}_{ts} B^{-*}_{si} \right).
\end{aligned}
$$

Substituting this into expression (6) for $\mathcal{W}_{k,\ell}$ and changing the order of summation we find

$$
\begin{aligned}
\mathcal{W}_{k,\ell} &= |S_k|^{-1/2}|S_\ell|^{-1/2} \sum_{q \in S_k} \sum_{r \in S_\ell} \sum_{s \in S_k} \sum_{t \in S_\ell} (V^*V)_{qr}(V^*V)^{-1}_{ts} \cdot \\
& \quad \big(\sum_{i \in S_k} B^{-*}_{si} B^*_{iq}\big) \cdot \big(\sum_{j \in S_\ell} B_{rj} B^{-1}_{jt}\big).
\end{aligned} \tag{8}
$$

Since the product of the last two factors in (8) is zero unless $s = q$ and $t = r$, in which case these factors are one, the expression (8) is equal to that in (6). It is now easily seen from (6) and (3) that if A is normal then $\mathcal{W}$ is the identity, since we can take the eigenvector matrix V to be unitary.

It follows from (7) that for any polynomial p, the Frobenius norm of $p(A)$ can be bounded in terms of the 2-norm of ϱ, which is the Frobenius norm of $p(\Lambda)$, and the square root of the 2-norm of $\mathcal{W}$:

$$
\|p(\Lambda)\|_F \;\le\; \|p(A)\|_F \;\le\; \|\mathcal{W}\|_2^{1/2} \cdot \|p(\Lambda)\|_F \tag{9}
$$

Again, both the lower and upper bounds in (9) are attainable for certain polynomials p (with $p(A) \neq 0$). The lower bound is attained for any polynomial p with equal values at all the eigenvalues of A (i.e., for $p(A)$ a

$$\|p(\Lambda)\|_F \;\leq\; \|p(A)\|_F \;\leq\; \|W\|_2^{1/2} \cdot \|p(\Lambda)\|_F. \tag{4}$$

Note that both bounds in (4) are sharp, in the sense that there are polynomials p (with $p(A) \neq 0$) for which these bounds are met. If p has equal values at all eigenvalues of A (in which case $p(A)$ is just a multiple of the identity), then the lower bound is met. If p is such that the vector ρ of values of p at the eigenvalues of A is parallel to the largest eigenvector of W, then the upper bound is attained.

Suppose now that A is diagonalizable but has multiple eigenvalues; say, A has m distinct eigenvalues, $\lambda_{S_1}, \ldots, \lambda_{S_m}$, where S_k, $k = 1, \ldots m$, denotes the set of indices of eigenvalues equal to λ_{S_k}. Let $|S_k|$ denote the number of indices in the set S_k. Equation (2) can be written in the form

$$\|p(A)\|_F^2 = \sum_{k=1}^{m} \sum_{\ell=1}^{m} \overline{p(\lambda_{S_k})} p(\lambda_{S_\ell}) \sum_{i \in S_k} \sum_{j \in S_\ell} W_{ij}. \tag{5}$$

Defining the m by m matrix $\mathcal{W}$ by

$$\mathcal{W}_{k,\ell} = |S_k|^{-1/2} |S_\ell|^{-1/2} \sum_{i \in S_k} \sum_{j \in S_\ell} W_{ij}, \qquad k, \ell = 1, \ldots m, \tag{6}$$

and the m-vector ϱ by

$$\varrho = \begin{pmatrix} |S_1|^{1/2} \; p(\lambda_{S_1}) \\ \vdots \\ |S_m|^{1/2} \; p(\lambda_{S_m}) \end{pmatrix},$$

equation (5) can be written in the form

$$\|p(A)\|_F^2 = \varrho^* \mathcal{W} \varrho. \tag{7}$$

The matrix $\mathcal{W}$ is also Hermitian, and it follows from (7) that $\mathcal{W}$ is positive definite, since $\|p(A)\|_F^2$ is greater than or equal to zero, with equality if and only if p is zero at the distinct eigenvalues of A, which means the vector ϱ is zero. Thus, the Frobenius norm of $p(A)$ is the $\mathcal{W}$-norm of ϱ.

It follows further from (7) that the eigenvalues of $\mathcal{W}$ are all greater than or equal to one, since $\|p(A)\|_F^2$ is greater than or equal to $\|p(\Lambda)\|_F^2$, which is $\|\varrho\|_2^2$. Equation (6), together with (3), implies that $\mathcal{W}$ has an eigenvalue 1, where ρ is the vector of values of the polynomial p at the eigenvalues of A and W is the Hadamard product of V^*V with $(V^*V)^{-T}$:

$$\rho = \begin{pmatrix} p(\lambda_1) \\ \vdots \\ p(\lambda_n) \end{pmatrix}, \qquad W_{ij} = (V^*V)_{ij} \cdot (V^*V)^{-1}_{ji}, \quad i,j = 1, \ldots, n. \qquad (3)$$

The matrix W is Hermitian, and it follows from (2) that W is positive definite, since $\|p(A)\|_F^2$ is greater than or equal to zero, with equality if and only if the vector ρ of values of p at the eigenvalues of A is zero. Thus, the Frobenius norm of $p(A)$ is the W-norm of ρ.

Note from definition (3) that if A is normal then W is the *identity* matrix. It can also be seen that W is *independent* of the *column scalings* of the eigenvector matrix V, since for any diagonal matrix D, we have

$$((VD)^*(VD))_{ij} \cdot ((VD)^*(VD))^{-1}_{ji} = (V^*V)_{ij} \cdot (V^*V)^{-1}_{ji}.$$

In addition, W is *independent* of *unitary similarity transformations* of A; that is, if A is replaced by Q^*AQ, where Q is a unitary matrix, then W and Λ are unchanged. Note also from formula (2) that all eigenvalues of W must be greater than or equal to one, since if W had an eigenvalue α less than one and if ρ were equal to the corresponding eigenvector, then we would have $\|p(A)\|_F^2 = \alpha\rho^*\rho = \alpha\|p(\Lambda)\|_F^2$, $\alpha < 1$. This is impossible since the Frobenius norm of $p(A)$ is always greater than or equal to that of $p(\Lambda)$. This can also be proved as in [6, p. 323], where a more general discussion of matrices of the form $C \circ C^{-T}$ is given. It can be seen from (3) that the row sums of W are all equal to one and so W has an eigenvalue 1, corresponding to the eigenvector of all ones:

$$W \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Using expression (2), the Frobenius norm of $p(A)$ can be bounded in terms of the 2-norm of ρ, which is the Frobenius norm of $p(\Lambda)$, and the square root of the 2-norm of W:

stand for any unitarily invariant norm. The trace of a square matrix A is written $\mathrm{tr}(A)$. A superscript * denotes the Hermitian transpose of a vector or matrix, while a superscript ^{T} denotes the ordinary transpose. The Hadamard, or, elementwise product of two matrices A and B is denoted $A \circ B$: $(A \circ B)_{ij} = A_{ij} B_{ij}$.

2 A Sharp Bound on $\|p(A)\|_F / \|p(\Lambda)\|_F$

Let A be an n by n matrix and let $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ be a diagonal matrix of eigenvalues of A. In this section we determine the smallest number γ such that

$$\|p(A)\|_F \leq \gamma \|p(\Lambda)\|_F,$$

for all polynomials p.

Assume first that A has n distinct eigenvalues (and hence a complete set of eigenvectors). Then we can write

$$A = V \Lambda V^{-1}$$

where V is a matrix of eigenvectors (unique up to column scalings). For any polynomial p, we have

$$p(A) = V p(\Lambda) V^{-1},$$

and the Frobenius norm of $p(A)$ satisfies

$$\|p(A)\|_F^2 = \mathrm{tr}(V^{-*} \overline{p(\Lambda)} V^* V p(\Lambda) V^{-1}).$$

Changing the order of factors inside the trace, this becomes

$$\|p(A)\|_F^2 = \mathrm{tr}((V^* V)^{-1} \overline{p(\Lambda)} (V^* V) p(\Lambda)),$$

which can be written in the form

$$\|p(A)\|_F^2 = \rho^* W \rho, \tag{2}$$

than or equal to the constant $\kappa(V)$ in (1), when κ is taken to be the Frobenius norm condition number, it is shown that if A has distinct eigenvalues then $\min_V \kappa_F(V)$ is within a factor of n of $\gamma_F(A)$, when the minimum is taken over all eigenvector matrices V of A and κ_F denotes the Frobenius norm condition number. Thus, the well-known bound (1) is almost sharp. Unlike (1), however, our sharp bound is derived from an equality that enables us to characterize those matrices A and polynomials p for which the upper bound is also a reasonable estimate of $\|p(A)\|_F$ and those for which it might be a large overestimate.

If the sharp bound $\gamma_F(A) \cdot \|p(\Lambda)\|_F$ is determined to be a large overestimate of $\|p(A)\|_F$ for some particular class of polynomials p, then different techniques are needed to estimate $\|p(A)\|_F$ for this class of functions. One possible approach is to base estimates on the eigenvalues of A and the *departure from normality* (sometimes called the *defect from normality*), which is defined as the infimum, over all Schur triangularizations of A, of the norm of the strictly upper triangular part of the Schur form. Bounds based on this measure of nonnormality are given in [5], for example, but, unless the departure from normality is very small, they do not generally provide good estimates of the actual norm of $p(A)$. A variety of related measures of nonnormality are discussed and compared in [1]. Another idea is to base estimates of $\|p(A)\|$, not on the eigenvalues and nonnormality of A, but on completely different quantities associated with A. This idea seems reasonable, since the eigenvalues of some matrices are highly sensitive to small perturbations in the matrix, while the norm of $p(A)$ may be much less sensitive. Interesting estimates have been derived in terms of the resolvent norms $\|(zI - A)^{-1}\|$, where z ranges over all complex numbers [8,10], or, equivalently, in terms of the ϵ-pseudospectra of A, where, for each positive number ϵ, the ϵ-pseudospectrum of A is the set of points z such that $\|(zI-A)^{-1}\| \geq \epsilon^{-1}$ [11]. A discussion of these estimates is beyond the scope of this paper, but we will point out when the eigenvalues of A, together with $\gamma_F(A)$, provide a good estimate of the Frobenius norm of a function of A, and when additional, or different, information is needed in order to obtain such an estimate.

Throughout the paper, $\|\cdot\|_2$ will denote the 2-norm of a vector or the corresponding spectral norm of a matrix, and $\|\cdot\|_F$ will denote the Frobenius norm of a matrix. The symbol $\|\cdot\|$ without any subscript will

One would like to be able to determine or estimate the norm of $p(A)$ using only a small amount of information about A – information that might be derived from consideration of the physical problem being modeled. For normal matrices A, the eigenvalues are sufficient to determine $\|p(A)\|$, for any unitarily invariant norm. For nonnormal matrices, the eigenvalues provide a lower bound on the norm of $p(A)$, but no reasonable upper bound can be given in terms of eigenvalues alone. One might, however, attempt to provide an upper bound on the norm of $p(A)$, based on the eigenvalues of A and a single number bounding the ratio of the norm of any analytic function of A to the norm of that function of the eigenvalues. Since any analytic function of an n by n matrix A can be written as a polynomial of degree $n-1$ or less in A [6, p. 412], we need only consider polynomial functions.

One very simple approach to obtaining such a bound is the following. Assume that A is diagonalizable and that $A = V\Lambda V^{-1}$ where Λ is a diagonal matrix of eigenvalues and V is a matrix of eigenvectors. For any polynomial p, we can write $p(A) = Vp(\Lambda)V^{-1}$, and the norm of $p(A)$ satisfies

$$\|p(A)\| \leq \kappa(V)\ \|p(\Lambda)\|, \tag{1}$$

where $\kappa(V) = \|V\| \cdot \|V^{-1}\|$ is the condition number of V.

There are several difficulties with the simple bound (1). First, the condition number of V depends on the scaling of the eigenvectors, and, if there are any multiple eigenvalues, it depends on the basis chosen to represent the corresponding eigenspace. To obtain the best possible bound in (1) one must determine the optimal scaling and basis choice. Second, even with the best conditioned eigenvector matrix V, the bound (1) is not sharp. For some matrices A, there are no polynomials p, for which equality holds in (1), as will be illustrated later. Third, there are important classes of polynomials p for which the bound in (1) is a large overestimate of the actual norm of $p(A)$, and it is difficult to determine for which matrices A and which polynomials p this is the case.

In this paper we derive a sharp upper bound on the ratio $\|p(A)\|_F/\|p(\Lambda)\|_F$, where $\|\cdot\|_F$ denotes the Frobenius norm. The bound is sharp in the sense that for any matrix A, there are polynomials p for which the bound is attained. We denote this quantity $\gamma_F(A)$. While $\gamma_F(A)$ is necessarily less

Norms of Functions of Matrices

Anne Greenbaum*

August, 1993

Abstract

Let A be an n by n matrix and let Λ be a diagonal matrix of eigenvalues of A. An expression is given for the smallest number γ such that $\|p(A)\|_F \leq \gamma \cdot \|p(\Lambda)\|_F$, for all polynomials p, where $\|\cdot\|_F$ denotes the Frobenius norm. It is shown that if A has distinct eigenvalues, then the well-known bound – $\gamma = \min_V \kappa_F(V)$, where κ_F denotes the Frobenius norm condition number and the minimum is taken over all eigenvector matrices V of A – is within a factor of n of the optimal constant γ. Classes of matrices A and polynomials p for which the sharp upper bound is also a reasonable estimate of $\|p(A)\|_F$ are described.

1 Introduction

In analyzing differential equations and various numerical algorithms, it is frequently necessary to determine or estimate the norm of $p(A)$, where A is an n by n matrix and p is a given polynomial or analytic function. For example, the stability of an evolution process governed by A is determined by the norms of the powers of A, $\|A^k\|$, $k = 1, 2, \ldots$, or by $\|e^{tA}\|$ for increasing time t [9]. The convergence rate of the GMRES algorithm for solving linear systems is determined by the norm of $P_k(A)$, $k = 1, 2, \ldots$, where P_k is the k^{th} degree polynomial with value one at the origin that minimizes $\|p(A)\|$ over all such polynomials p [3].

*Courant Institute of Mathematical Sciences, 251 Mercer St., New York, NY 10012.
This work was supported in part by DOE contract DEFG0288ER25053.