Upper and Lower Bounds on Norms of Functions of Matrices

Anne Greenbaum*

June 17, 2008

Abstract

Given an n by n matrix A, we look for a set S in the complex plane and positive scalars m and M such that for all functions p bounded and analytic on S and throughout a neighborhood of each eigenvalue of A, the inequalities

$$m \cdot \inf\{\|f\|_{\mathcal{L}^\infty(S)} : f(A) = p(A)\} \le \|p(A)\| \le M \cdot \inf\{\|f\|_{\mathcal{L}^\infty(S)} : f(A) = p(A)\}$$

hold. We show that for 2 by 2 matrices, if S is the field of values, then one can take $m = 1$ and $M = 2$. We show that for a perturbed Jordan block – a matrix A that is an n by n Jordan block with eigenvalue 0 except that its $(n, 1)$-entry is ν, with $|\nu| \in (0, 1)$ – if S is the unit disk, then $m = M = 1$. We argue, however, that, in general, due to the behavior of minimal-norm interpolating functions, it may be very difficult or impossible to find such a set S for which the ratio M/m is of moderate size.

1 Introduction

In recent years there has been considerable interest in finding sets in the complex plane that can be associated with a given square matrix or bounded

*University of Washington, Mathematics Dept., Box 354350, Seattle, WA 98195. This work was supported in part by NSF grant DMS-0208353.

linear operator A to give more information than the spectrum alone can provide about the norms of functions of A. Examples include the *field of values* or *numerical range* [14, 3, 4], the *ϵ-pseudospectrum* [19], and the *polynomial numerical hull* of a given degree [16, 17, 11]. Let S be a set that contains the spectrum of A. One might look for a scalar M (which might or might not depend on A) such that for all functions $p \in \mathcal{H}^\infty(S)$, the Hardy space of bounded analytic functions on S with norm $\|p\|_{\mathcal{L}^\infty(S)} \equiv \sup_{z \in S} |p(z)|$ (and with the additional requirement that p be analytic in a neighborhood of each eigenvalue of A so that $p(A)$ is well-defined), the upper bound

$$\|p(A)\| \leq M \cdot \|p\|_{\mathcal{L}^\infty(S)} \tag{1}$$

holds.

In this paper, we restrict our attention to n by n matrices A and let $\|\cdot\|$ denote the 2-norm for vectors and the corresponding spectral norm for matrices: $\|A\| \equiv \sup_{\|\mathbf{v}\|=1} \|A\mathbf{v}\|$. We look for sets S where (1) holds and where there is a similar lower bound on $\|p(A)\|$ involving a positive scalar m.

One's first thought might be to look for a positive scalar m such that for all $p \in \mathcal{H}^\infty(S)$,

$$m \cdot \|p\|_{\mathcal{L}^\infty(S)} \leq \|p(A)\|. \tag{2}$$

If S is the spectrum of A, and if A is diagonalizable – $A = V \Lambda V^{-1}$, where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ is a diagonal matrix of eigenvalues and V a matrix whose columns are eigenvectors – then the following inequalities hold:

$$\|p\|_{\mathcal{L}^\infty(S)} \leq \|p(A)\| \leq \kappa(V) \cdot \|p\|_{\mathcal{L}^\infty(S)}, \quad \kappa(V) \equiv \|V\| \cdot \|V^{-1}\|. \tag{3}$$

Hence the scalars m and M in (2) and (1) can be taken to be 1 and $\kappa(V)$, respectively. The scalar $m = 1$ is best possible since, for example, if $p(z) \equiv 1$, then $p(A) = I$ and $\|I\| = 1$. If the columns of V are taken to have norm 1 and if the eigenvalues of A are distinct, then the scalar $M = \kappa(V)$ is within a factor of n of optimal, since if $p(\lambda_J) = 1$, where J is the index of a row of V^{-1} with maximal norm and $p(\lambda_i) = 0$ for $i \neq J$, then

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

where $V(:,J)$ denotes the Jth column of V and $V^{-1}(J,:)$ denotes the Jth row of V^{-1}, while

$$\kappa(V) \leq \|V\|_F \cdot \|V^{-1}\|_F = \sqrt{n} \cdot \left(\sum_{j=1}^{n} \|V^{-1}(j,:)\|^2 \right)^{1/2} \leq n \cdot \|V^{-1}(J,:)\|,$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

It is easy to see, however, that if the set S contains points outside the spectrum of A, then there is no positive scalar m for which (2) holds, since if p is the minimal polynomial of A, then $p(A) = 0$ but $p(z) = 0$ only if z is an eigenvalue of A.

One way to circumvent this difficulty is to note that if the degree of the minimal polynomial of A is r, then any function $p(A)$ can be written as a polynomial of degree at most $r-1$ in A: $p(A) = p_{r-1}(A)$, where p_{r-1} is the polynomial of degree at most $r-1$ that matches p at the eigenvalues of A and whose derivatives of orders up through $t-1$ also match those of p at eigenvalues corresponding to a t by t Jordan block. Hence one might look for a set S and a positive scalar m such that for all $p \in \mathcal{H}^\infty(S)$

$$m \cdot \|p_{r-1}\|_{\mathcal{L}^\infty(S)} \leq \|p(A)\|, \tag{4}$$

where p_{r-1} is the polynomial of degree at most $r-1$ satisfying $p_{r-1}(A) = p(A)$. The largest set S for which (4) holds with $m = 1$ is, by definition, the *polynomial numerical hull of degree* $r-1$ [16, 17, 11].

Since $p(A)$ can be represented in infinitely many ways as a function of A – if f is any function satisfying $f(z) = p_{r-1}(z) + \chi(z)h(z)$, where χ is the minimal polynomial of A and $h \in \mathcal{H}^\infty(S)$, then $f(A) = p(A)$ – it might seem somewhat arbitrary to represent it in the form $p_{r-1}(A)$. To have the best chance of finding a good lower bound, one might instead consider all functions $f \in \mathcal{H}^\infty(S)$ such that $f(A) = p(A)$ and choose one whose $\mathcal{L}^\infty$-norm on S is as small as possible. One might then ask for a positive scalar m such that for all $p \in \mathcal{H}^\infty(S)$,

$$m \cdot \inf\{\|f\|_{\mathcal{L}^\infty(S)} : \ f(A) = p(A)\} \leq \|p(A)\|. \tag{5}$$

Note that if the upper bound (1) holds for all such p, then it also must be the case that

$$\|p(A)\| \leq M \cdot \inf\{\|f\|_{\mathcal{L}^\infty(S)} : \ f(A) = p(A)\}, \tag{6}$$

since $p(A)$ could always be written as $\tilde{f}(A)$, where $\tilde{f}$ is a function that achieves the infimum in (6) or comes arbitrarily close to it.

If A is diagonalizable, with eigendecomposition $A = V\Lambda V^{-1}$, then $f(A) = p(A)$ if and only if $f(\Lambda) = p(\Lambda)$. If S is the unit disk, then for given values

$p(\Lambda)$, the infimum in (5) and (6) is attained by a unique function $\tilde{f}$ and the problem of finding this minimal-norm interpolating function is known as a **Pick-Nevanlinna interpolation problem**. The function $\tilde{f}$ is a scalar multiple of a **finite Blaschke product** and can be written in the forms

$$\tilde{f}(z) = \mu \prod_{k=1}^{n-1} \frac{z-\alpha_k}{1-\bar{\alpha}_k z}, \quad |\alpha_k| < 1, \tag{7}$$

$$= \mu \frac{\gamma_0 + \gamma_1 z + \ldots + \gamma_{n-1} z^{n-1}}{\bar{\gamma}_{n-1} + \bar{\gamma}_{n-2} z + \ldots + \bar{\gamma}_0 z^{n-1}}, \quad \gamma_{n-1} = 1. \tag{8}$$

Using the representation (8), it was shown in [9] that the function $\tilde{f}$ and its $\mathcal{L}^\infty$-norm can be found by solving a simple eigenvalue problem (actually a *coneigenvalue* problem [13, p. 245]). Similar results can be obtained when A is not diagonalizable, in which case the interpolation problem becomes one of finding the minimal-norm function $\tilde{f}$ that matches p at the eigenvalues and whose derivatives of order up through $t-1$ match those of p at eigenvalues corresponding to a t by t Jordan block. For simplicity of exposition, however, we will assume throughout the remainder of this paper that A is diagonalizable with n distinct eigenvalues.

With this assumption, we look for a set S containing the spectrum of A and positive scalars m and M such that for all $p \in \mathcal{H}^\infty(S)$,

$$m \cdot \inf\{\|f\|_{\mathcal{L}^\infty(S)} : \ f(\Lambda) = p(\Lambda)\} \le \|p(A)\| \le M \cdot \inf\{\|f\|_{\mathcal{L}^\infty(S)} : \ f(\Lambda) = p(\Lambda)\}. \tag{9}$$

When the ratio $\kappa(V)/1$ in (3) is large, we hope to find sets S for which $M/m << \kappa(V)$.

If S is not the unit disk but is a simply connected open set (other than the entire complex plane) then it can be mapped onto the open unit disk $\mathcal{D}$ via a one-to-one analytic function g. In this case, the problem of finding the minimal-norm interpolating function $\tilde{f}$ can be translated to the unit disk, since

$$\inf\{\|f\|_{\mathcal{L}^\infty(S)} : f(\lambda_j) = p(\lambda_j), \ j = 1, \ldots, n\} =$$
$$\inf\{\|(f \circ g^{-1}) \circ g\|_{\mathcal{L}^\infty(S)} : ((f \circ g^{-1}) \circ g)(\lambda_j) = p(\lambda_j), \ j = 1, \ldots, n\} =$$
$$\inf\{\|F\|_{\mathcal{L}^\infty(\mathcal{D})} : F(g(\lambda_j)) = p(\lambda_j), \ j = 1, \ldots, n\}.$$

Hence one can study the minimal-norm interpolating function (but interpolating the values $p(\lambda_j) \equiv (p \circ g^{-1})(g(\lambda_j)))$ for the mapped matrix $g(A)$ on the unit disk using the techniques in [9]. Some results about the minimal-norm interpolating function are also known for multiply-connected sets S. See, for example, [1, 20].

Since multiplying $p(\Lambda)$ by a nonzero scalar maintains the inequalities in (9), one can consider only values $p(\Lambda)$ for which $\max_{j=1,\ldots,n} |p(\lambda_j)| = 1$. Letting $\mathbf{w} = (w_1, \ldots, w_n)^T$ denote the vector of values $w_j = p(\lambda_j)$ and $\mathrm{diag}(\mathbf{w})$ denote the diagonal matrix whose (j,j)-entry is w_j, one can write the largest scalar m and the smallest scalar M for which (9) holds for all $p \in \mathcal{H}^\infty(S)$ in the form

$$m_{opt}(A,S) = \inf_{\|\mathbf{w}\|_\infty = 1} \frac{\|V \mathrm{diag}(\mathbf{w}) V^{-1}\|}{\inf\{\|f\|_{\mathcal{L}^\infty(S)} :\ f(\lambda_j) = w_j,\ j = 1, \ldots, n\}}, \tag{10}$$

$$M_{opt}(A,S) = \sup_{\|\mathbf{w}\|_\infty = 1} \frac{\|V \mathrm{diag}(\mathbf{w}) V^{-1}\|}{\inf\{\|f\|_{\mathcal{L}^\infty(S)} :\ f(\lambda_j) = w_j,\ j = 1, \ldots, n\}}. \tag{11}$$

Let S be the unit disk $\mathcal{D}$, and let C_Λ denote the supremum over all vectors $\mathbf{w}$ with $\|\mathbf{w}\|_\infty = 1$ of the denominator in (10) or (11):

$$C_\Lambda \equiv \sup_{\|\mathbf{w}\|_\infty = 1} \inf\{\|f\|_{\mathcal{L}^\infty(\mathcal{D})} :\ f(\lambda_j) = w_j,\ j = 1, \ldots, n\}. \tag{12}$$

This quantity also has been studied and is known as the *constant of interpolation* [6]. Using (10), one can give upper and lower bounds on $m_{opt}(A, \mathcal{D})$ in terms of C_Λ:

$$\frac{1}{C_\Lambda} \le m_{opt}(A, \mathcal{D}) \le \frac{\kappa(V)}{C_\Lambda}. \tag{13}$$

Define

$$\delta \equiv \min_k \prod_{j \ne k} \left| \frac{\lambda_j - \lambda_k}{1 - \bar{\lambda}_j \lambda_k} \right|. \tag{14}$$

The following theorem can be found in [6, p. 278]:

Theorem. The number δ in (14) and the constant of interpolation C_Λ satisfy the inequalities

$$\frac{1}{\delta} \le C_\Lambda \le \frac{c}{\delta}\left(1 + \log \frac{1}{\delta}\right), \tag{15}$$

where c is some absolute constant.

Tight bounds on the constant c in (15) are known. For example, it is shown in [15, p. 268] that

$$C_\Lambda \leq \frac{2e + 4e\log(1/\delta)}{\delta}. \tag{16}$$

Substituting (15) and (16) into (13) gives

$$\frac{\delta}{2e + 4e\log(1/\delta)} \leq m_{opt}(A, \mathcal{D}) \leq \kappa(V)\delta. \tag{17}$$

2 The Field of Values and Two-by-Two Matrices

If S is the field of values of A,

$$S = W(A) \equiv \{\langle Aq, q\rangle : \ \|q\| = 1\},$$

then Crouzeix has shown that $M_{opt}(A, W(A)) \leq 11.08$ and he conjectures that 11.08 can be replaced by 2. The bound $M_{opt}(A, W(A)) \leq 2$ has been proved in the case when $W(A)$ is a disk or when A is a 2 by 2 matrix [3, 4].

Since the field of values is the polynomial numerical hull of degree 1 [11], and since any function of a 2 by 2 matrix A can be written as a first degree polynomial in A, it is also the case that for any $p \in \mathcal{H}^\infty(W(A))$,

$$\|p(A)\| = \|p_1(A)\| \geq \|p_1\|_{\mathcal{L}^\infty(W(A))} \geq \inf\{\|f\|_{\mathcal{L}^\infty(W(A))} : \ f(A) = p(A)\}.$$

Hence $m_{opt}(A, W(A)) \geq 1$ (and, in fact, $m_{opt}(A, W(A)) = 1$).

Thus, the norm behavior of functions of a 2 by 2 matrix A is determined, to within at most a factor of 2, by the field of values (and the eigenvalues and Jordan structure necessary to determine the functions f for which $f(A) = p(A)$):

$$\inf\{\|f\|_{\mathcal{L}^\infty(W(A))} : \ f(A) = p(A)\} \leq \|p(A)\| \leq 2 \cdot \inf\{\|f\|_{\mathcal{L}^\infty(W(A))} : \ f(A) = p(A)\}.$$

Following is a simple 2 by 2 example:

$$A = \begin{pmatrix} 0 & 1 \\ -.01 & 0 \end{pmatrix}. \tag{18}$$

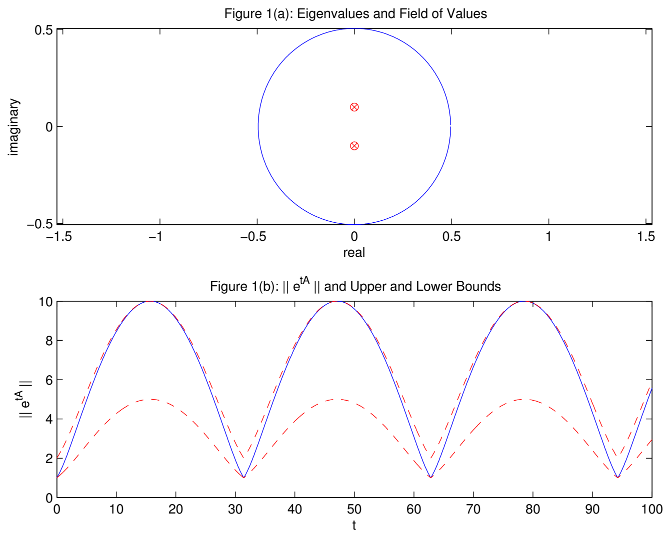

Figure 1: (a) Eigenvalues and field of values of A; (b) $\|e^{tA}\|$ (solid) and upper and lower bounds (dashed).

The eigenvalues of A are $\pm 0.1i$, and the field of values is close to a disk of radius 0.5; more precisely, it is an ellipse centered at the origin with major axis of length 1.01 along the imaginary axis and minor axis of length 0.99 along the real axis. These are pictured in Figure 1a. If we map this ellipse conformally to the unit disk, with the origin mapping to itself, then the eigenvalues are mapped to approximately $\pm 0.2i$. For a given function p, we can then compute the function $\tilde{f}$ of minimal $\mathcal{L}^\infty$-norm on the unit disk that takes the values $p(\pm 0.1i)$ at the mapped eigenvalues. We consider the functions $p(z) = e^{tz}$, for $t \in [0, 100]$. Figure 1b is a plot of $\|e^{tA}\|$ (solid) together with the lower bound $\|\tilde{f}\|_{\mathcal{L}^\infty(\mathcal{D})}$ and the upper bound $2\|\tilde{f}\|_{\mathcal{L}^\infty(\mathcal{D})}$ (dashed).

3 The Unit Disk and Perturbed Jordan Blocks

Let A be an n by n perturbed Jordan block:

$$A = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ \nu & & & 0 \end{pmatrix} \quad |\nu| \in (0,1). \tag{19}$$

The eigenvalues of this matrix are the nth roots of ν; if $\nu = |\nu| e^{i\theta}$, $-\pi < \theta \leq \pi$, and $\nu^{1/n} = |\nu|^{1/n} e^{i\theta/n}$ and $\omega = e^{2\pi i/n}$, then

$$\lambda_j = \nu^{1/n} \omega^{j-1}, \quad j = 1, \ldots, n.$$

The corresponding eigenvectors are

$$\mathbf{v}_j = \begin{pmatrix} 1 \\ \nu^{1/n} \omega^{j-1} \\ \nu^{2/n} \omega^{2(j-1)} \\ \vdots \\ \nu^{(n-1)/n} \omega^{(n-1)(j-1)} \end{pmatrix}, \quad j = 1, \ldots, n.$$

The matrix V of eigenvectors is the *Vandermonde* matrix for the eigenvalues. Its transpose is:

$$V^T = \begin{pmatrix} 1 & \nu^{1/n} & \ldots & \nu^{(n-1)/n} \\ 1 & \nu^{1/n}\omega & \ldots & \nu^{(n-1)/n}\omega^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & \nu^{1/n}\omega^{n-1} & \ldots & \nu^{(n-1)/n}\omega^{(n-1)^2} \end{pmatrix} = \begin{pmatrix} 1 & \lambda_1 & \ldots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \ldots & \lambda_2^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & \lambda_n & \ldots & \lambda_n^{n-1} \end{pmatrix}. \tag{20}$$

Let p be a function analytic in the unit disk $\mathcal{D}$ and let $\mathbf{w}$ be the vector of values of p at the eigenvalues of A. Then $p(A) = V \mathrm{diag}(\mathbf{w}) V^{-1}$. It was noted earlier that the function $\tilde{f}$ which, among all analytic functions f satisfying $f(\lambda_j) = w_j$, $j = 1, \ldots, n$, has minimal infinity norm on $\mathcal{D}$ is of the form (8). The scalars γ_k and μ are determined by the interpolation conditions:

$$\mu \sum_{k=0}^{n-1} \gamma_k \lambda_j^k = w_j \sum_{k=0}^{n-1} \bar{\gamma}_k \lambda_j^{n-1-k}, \quad j = 1, \ldots, n.$$

Writing this in the form of a matrix equation, we have

$$\mu V^T \gamma = \mathrm{diag}(\mathbf{w}) V^T \Pi \bar{\gamma},$$

where $\gamma \equiv (\gamma_0, \ldots, \gamma_{n-1})^T$ is the vector of unknown coefficients, V^T is the same transposed Vandermonde matrix as in (20), and Π is the permutation matrix with 1's along its skew diagonal. Equivalently,

$$V^{-T} \mathrm{diag}(\mathbf{w}) V^T \Pi \bar{\gamma} = \mu \gamma. \tag{21}$$

It is shown in [9] that there is a positive scalar μ for which (21) is satisfied with a certain nonzero vector γ, and that the largest such value μ is the $\mathcal{L}^\infty$-norm of $\tilde{f}$. Equating real and imaginary parts in (21) leads to the eigenvalue problem described in [9]. Letting $C = V^{-T} \mathrm{diag}(\mathbf{w}) V^T \Pi$, this eigenvalue problem can be written in the form

$$\begin{pmatrix} \mathrm{Re}(C) & \mathrm{Im}(C) \\ \mathrm{Im}(C) & -\mathrm{Re}(C) \end{pmatrix} \begin{pmatrix} \mathrm{Re}(\gamma) \\ \mathrm{Im}(\gamma) \end{pmatrix} = \mu \begin{pmatrix} \mathrm{Re}(\gamma) \\ \mathrm{Im}(\gamma) \end{pmatrix}. \tag{22}$$

Equation (21) itself is known as a *coneigenvalue* problem [13, p. 245].

Because the transposed Vandermonde matrix appearing in (21) is the transposed eigenvector matrix of the corresponding perturbed Jordan block, we are able to prove the following:

Theorem 1. Let A be defined by (19). The eigenvalues of the matrix in (22) are plus and minus the singular values of $p(A)$. Thus, for any function $p \in \mathcal{H}^\infty(\mathcal{D})$,

$$\|p(A)\| = \inf\{\|f\|_{\mathcal{L}^\infty(\mathcal{D})} : \ f(A) = p(A)\}, \tag{23}$$

and hence $m_{opt}(A, \mathcal{D}) = M_{opt}(A, \mathcal{D}) = 1$.

Proof: By the previous discussion, equation (21) is the same as

$$(p(A))^T \Pi \bar{\gamma} = \mu \gamma. \tag{24}$$

Note also that for $j = 1, \ldots, n-1$, $(A^T)^j \Pi$ is symmetric: Since $(A^T)^j$ has ν's along its $(n-j)$th superdiagonal and 1's along its jth subdiagonal and zeros elsewhere, when the order of the columns is reversed by multiplying on the right by Π, the resulting matrix has ν's along the $(n-j)$th super-skewdiagonal (running from upper right to lower left) and 1's along the jth sub-skewdiagonal and zeros elsewhere. Since any function $p(A)$ can be written as a polynomial of degree at most $n-1$ in A, it follows that $(p(A))^T \Pi$ is (complex) symmetric and can be written in the form: $(p(A))^T \Pi = X \Sigma X^T$, where X is a unitary matrix and Σ is the diagonal matrix of singular values. Since $p(A) = (\Pi X) \Sigma X^T$, it has the same singular values. Letting $\mathbf{x}_j$ denote the jth column of X, we have the relations $\mathbf{x}_j^* \mathbf{x}_k = \mathbf{x}_k^T \bar{\mathbf{x}}_j = \delta_{jk}$ and hence γ in equation (21) can be taken to be $\mathbf{x}_j$ and then μ will be σ_j. Thus the singular values of $p(A)$ are eigenvalues of the matrix in (22). The other eigenvalues are found by setting $\gamma = i\mathbf{x}_j$ in (21). Then $\bar{\gamma} = -i\bar{\mathbf{x}}_j$ and $X \Sigma X^T \bar{\gamma} = -i\sigma_j \mathbf{x}_j = -\sigma_j \gamma$. Thus the eigenvalues of the matrix in (22) are plus and minus the singular values of $p(A)$, and the largest of these eigenvalues is therefore both $\|p(A)\|$ and, by the argument in [9], $\inf\{\|f\|_{\mathcal{L}^\infty(\mathcal{D})} : \ f(A) = p(A)\}$. $\square$

If we look again at the 2 by 2 example of the previous section with A defined by (18), but now relate $\|p(A)\|$ to $\|\hat{f}\|_{\mathcal{L}^\infty(\mathcal{D})}$ where $\hat{f}$ is the minimal-norm function satisfying $\hat{f}(\pm 0.1i) = p(\pm 0.1i)$, then instead of obtaining upper and lower bounds on $\|p(A)\|$ that differ by a factor of 2, we obtain an exact equality. In particular, when $p(A) = e^{tA}$, the plot of $\|\hat{f}\|_{\mathcal{L}^\infty(\mathcal{D})}$ is identical to the solid curve showing $\|e^{tA}\|$ in Figure 1b.

Of course, equality (23) holds also for any matrix A that is unitarily similar to the one in (19), and if $B = ZAZ^{-1}$ where A is the matrix in (19), then one has the estimates

$$\frac{1}{\kappa(Z)} \cdot \inf\{\|f\|_{\mathcal{L}^\infty(\mathcal{D})} : \ f(B) = p(B)\} \leq \|p(B)\| \leq$$

$$\kappa(Z) \cdot \inf\{\|f\|_{\mathcal{L}^\infty(\mathcal{D})} : \ f(B) = p(B)\},$$

so that $M_{opt}(B, \mathcal{D}) / m_{opt}(B, \mathcal{D}) \leq (\kappa(Z))^2$.

Suppose g is a conformal mapping from $\mathcal{D}$ onto some other set $S \subset \mathbf{C}$. If $B = g(A)$, where A is defined by (19), then $m_{opt}(B, S) = M_{opt}(B, S) = 1$, since for any $p \in \mathcal{H}^\infty(S)$,

$$\|p(B)\| = \|(p \circ g)(A)\| = \inf\{\|f\|_{\mathcal{L}^\infty(\mathcal{D})} : f(A) = (p \circ g)(A)\} =$$

$$\inf\{\|(f \circ g^{-1}) \circ g\|_{\mathcal{L}^\infty(\mathcal{D})} : ((f \circ g^{-1}) \circ g)(A) = (p \circ g)(A)\} =$$

$$\inf\{\|F\|_{\mathcal{L}^\infty(S)} : F(B) = p(B)\}. \tag{25}$$

Since the unit disk can be mapped conformally to any simply connected open set S other than the entire complex plane, for any such set S there is a family of matrices $B \equiv B(\nu, n)$ satisfying (25) for all $p \in \mathcal{H}^\infty(S)$. Since any function $g(A)$ can be written as a polynomial of degree $n-1$ in A, these matrices have the form

$$B = \begin{pmatrix} c_0 & c_1 & \dots & c_{n-1} \\ c_{n-1}\nu & c_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & c_1 \\ c_1\nu & \dots & c_{n-1}\nu & c_0 \end{pmatrix},$$

for certain coefficients $c_0, \dots, c_{n-1}$.

The perturbed Jordan block in (19) is a special case of a *companion matrix.* A general companion matrix with eigenvalues in $\mathcal{D}$ again has an eigenvector matrix that is the Vandermonde matrix for its eigenvalues. Hence equation (24) still holds, and we are able to obtain a lower bound on $\|p(A)\|$ with $m = 1$, but an upper bound may be more problematic.

Corollary 2. Let A be a *companion matrix* with eigenvalues in $\mathcal{D}$:

$$A = \begin{pmatrix} 0 & 1 & & & \\ 0 & 0 & 1 & & \\ \vdots & \vdots & \ddots & \ddots & \\ 0 & 0 & & 0 & 1 \\ c_0 & c_1 & \dots & c_{n-2} & c_{n-1} \end{pmatrix},$$

where all roots of $z^n - \sum_{j=0}^{n-1} c_j z^j$ lie in $\mathcal{D}$. Then for any function $p \in \mathcal{H}^\infty(\mathcal{D})$:

$$\|p(A)\| \geq \inf\{\|f\|_{\mathcal{L}^\infty(\mathcal{D})} : f(A) = p(A)\}, \tag{26}$$

and hence $m_{opt}(A, \mathcal{D}) = 1$.

Proof: It is easy to see that the eigenvector matrix for A is the Vandermonde matrix for its eigenvalues; the eigenvalues are the values λ that satisfy $\lambda^n - \sum_{j=0}^{n-1} c_j \lambda^j = 0$, and the corresponding eigenvectors have the form $(1, \lambda, \lambda^2, \dots, \lambda^{n-1})^T$. Hence equation (24) holds in this case as well. Since $p(A)^T\Pi$ is not necessarily symmetric, we cannot make the argument that if γ is a singular vector of $p(A)^T\Pi$ then μ is the corresponding singular value, but since $p(A)^T\Pi\bar{\gamma} = \mu\gamma$ and $\|\Pi\bar{\gamma}\| = \|\gamma\|$, it follows that $\|p(A)^T\| \geq |\mu|$. Hence $\|p(A)\| \geq |\mu|$ and $m_{opt}(A, \mathcal{D}) = 1$. $\square$

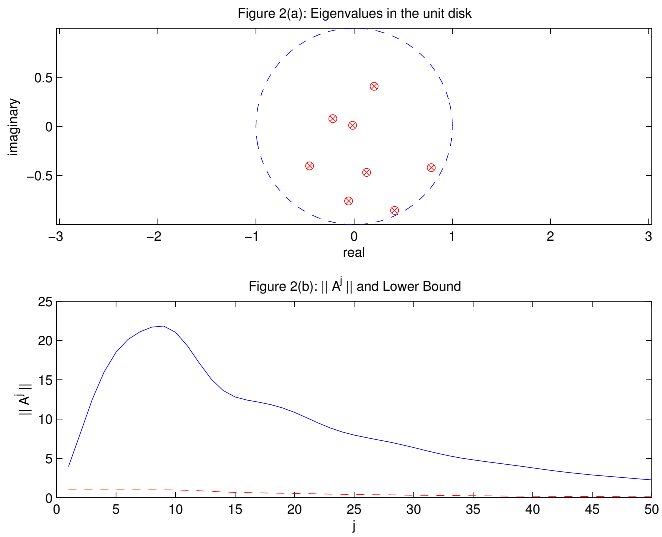

Figure 2: (a) Eigenvalues of A in the unit disk; (b) $\|A^k\|$ (solid) and lower bound (dashed).

The following example illustrates the fact that for a general companion matrix with eigenvalues in $\mathcal{D}$, if one takes S to be the unit disk, then while inequality (5) holds with $m = 1$, this is not necessarily a very sharp lower bound on $\|p(A)\|$, and M in (6) may have to be significantly larger. Here A is an 8 by 8 companion matrix with eigenvalues randomly distributed in $\mathcal{D}$ and we looked at $p(A) = A^k$, $k = 1, \ldots, 50$. The eigenvalues of A are plotted, along with the unit disk, in Figure 2a, and $\|A^k\|$ is plotted (solid) along with the lower bound $\inf\{\|f\|_{\mathcal{L}^\infty(\mathcal{D})} : f(A) = A^k\}$ (dashed) in Figure 2b. Note that the lower bound is always less than or equal to 1 since $|z^k| < 1$ in $\mathcal{D}$, so if the powers of A grow in norm before starting to decrease, as in this example, then the lower bound will not be very sharp.

Equation (21) together with (14) and (17) provides one more interesting piece of information:

Corollary 3. The 2-norm condition number of the n by n Vandermonde matrix for the points $\lambda_1, \ldots, \lambda_n \in \mathcal{D}$ satisfies

$$\kappa(V) \geq \max_k \prod_{j \neq k} \left| \frac{1 - \bar{\lambda}_j \lambda_k}{\lambda_j - \lambda_k} \right|. \tag{27}$$

Proof: If A is the companion matrix whose eigenvalues are $\lambda_1, \ldots, \lambda_n \in \mathcal{D}$ and whose eigenvector matrix is the Vandermonde matrix V for these points, then Corollary 2 states that $m_{opt}(A, \mathcal{D}) = 1$. From (17), $\kappa(V) \geq m_{opt}(A, \mathcal{D})/\delta$, and the desired result follows from substituting expression (14) for δ. $\Box$

A number of estimates have been derived for the condition number of a Vandermonde matrix. See, for example, [12, p. 428] for a summary of results. It is shown in [7, 8], for instance, that

$$\max_k \prod_{j \neq k} \frac{\max(1, |\lambda_j|)}{|\lambda_k - \lambda_j|} \leq \|V^{-1}\|_\infty \leq \max_k \prod_{j \neq k} \frac{1 + |\lambda_j|}{|\lambda_k - \lambda_j|}, \tag{28}$$

where $\| \cdot \|_\infty$ denotes the operator norm corresponding to the ∞-norm for vectors. Since the 2-norm of an n by n matrix is greater than or equal to $n^{-1/2}$ times the ∞-norm of the matrix, and since $\|V\|_2 \geq \|V^T e_1\|_2 = n^{1/2}$, the left-hand side of (28) is also a lower bound for the 2-norm condition number $\kappa(V)$. Sometimes (27) is a larger lower bound and sometimes (28) is, depending on the exact distribution of the points $\lambda_1, \ldots, \lambda_n$.

4 Relation to the Polynomial Numerical Hull

The following problem received considerable attention in the complex analysis literature during the early part of the last century [18, 10]:

> (*) Suppose $f(z) \equiv \sum_{j=0}^{n-1} c_j z^j + z^n h(z)$ where $h \in \mathcal{H}^\infty(\mathcal{D})$, is a function analytic in $\mathcal{D}$ and bounded in modulus by 1 there. In what region must its nth partial sum $\sum_{j=0}^{n-1} c_j z^j$ be bounded in modulus by 1?

One might generalize this question as follows:

> (**) Let χ be a given polynomial of degree n. Suppose $f(z) \equiv \sum_{j=0}^{n-1} c_j z^j + \chi(z) h(z)$ where $h \in \mathcal{H}^\infty(\mathcal{D})$, is a function analytic in $\mathcal{D}$ and bounded in modulus by 1 there. In what region must its nth partial sum $\sum_{j=0}^{n-1} c_j z^j$ be bounded in modulus by 1?

Let A be an n by n companion matrix with eigenvalues in $\mathcal{D}$ and with minimal polynomial χ of degree n, or, more generally, let A be any matrix with eigenvalues in $\mathcal{D}$ and minimal polynomial χ that satisfies inequality (26). Any function $p(A)$ can be written in the form $p_{n-1}(A) + \chi(A)h(A)$, where p_{n-1} is the polynomial of degree at most $n-1$ satisfying $p_{n-1}(A) = p(A)$ and h is any function in $\mathcal{H}^\infty(\mathcal{D})$. By Corollary 2, since the infimum in (26) is attained, we can write

$$\|p(A)\| = \|p_{n-1}(A)\| \geq \min_{h\in\mathcal{H}^\infty(\mathcal{D})} \sup_{z\in\mathcal{D}} |p_{n-1}(z) + \chi(z)h(z)|,$$

or, dividing each side by $\|p_{n-1}(A)\|$,

$$1 \geq \min_{h\in\mathcal{H}^\infty(\mathcal{D})} \sup_{z\in\mathcal{D}} |p_{n-1}(z)/\|p_{n-1}(A)\| + \chi(z)h(z)|.$$

The answer to (**) is a region in which $|p_{n-1}(z)|/\|p_{n-1}(A)\| \leq 1$, or, $|p_{n-1}(z)| \leq \|p_{n-1}(A)\|$ for all $p_{n-1} \in \Pi_{n-1}$; hence it is a subset of the polynomial numerical hull of degree $n-1$ for A. In general, the polynomial numerical hull of degree $n-1$ for an n by n companion matrix A is not known, and the question (**) remains unanswered as well.

In the special case where $\chi(z) = z^n$, as in the original question (*), however, the answer is known. This corresponds to the case where A is an n by n Jordan block with eigenvalue 0. Although Theorem 1 was derived here only for $|\nu| \in (0,1)$, it follows from a theorem of Carathéodory and Fejér [2] and results in [5] that it holds also for $\nu = 0$. Hence, in this case as well, for any $p \in \mathcal{H}^\infty(\mathcal{D})$, if p_{n-1} is the polynomial of degree $n-1$ that satisfies $p_{n-1}(A) = p(A)$, then

$$\|p(A)\| = \|p_{n-1}(A)\| = \min_{h\in\mathcal{H}^\infty(\mathcal{D})} \sup_{z\in\mathcal{D}} |p_{n-1}(z) + z^n h(z)|,$$

or, dividing each side by $\|p_{n-1}(A)\|$,

$$1 = \min_{h\in\mathcal{H}^\infty(\mathcal{D})} \sup_{z\in\mathcal{D}} |p_{n-1}(z)/\|p_{n-1}(A)\| + z^n h(z)|.$$

The answer to (*) is a region in which $|p_{n-1}(z)|/\|p_{n-1}(A)\| \leq 1$, or, $|p_{n-1}(z)| \leq \|p_{n-1}(A)\|$ for all $p_{n-1} \in \Pi_{n-1}$; hence it is a subset of the polynomial numerical hull of degree $n-1$ for A. For any larger region R, (*) implies that there is a polynomial $p_{n-1} \in \Pi_{n-1}$ for which $\min_{h\in\mathcal{H}^\infty(\mathcal{D})} \sup_{z\in\mathcal{D}} |p_{n-1}(z) + z^n h(z)| \leq 1$, but $\sup_{z\in R} |p_{n-1}(z)| > 1$. Hence $\|p_{n-1}(A)\| \leq 1$, but $\sup_{z\in R} |p_{n-1}(z)| > 1$, so R is not a subset of the polynomial numerical hull of degree $n-1$. Thus, question (*) is equivalent to the question: What is the polynomial numerical hull of degree $n-1$ for an n by n Jordan block with eigenvalue 0? The answer is given in [18, 10, 5]: It is a disk about the origin of radius $r_{n-1,n}$ where, for n even, $r_{n-1,n}$ is the positive root of $2r^n + r - 1 = 0$, and for n odd $r_{n-1,n}$ is greater than or equal to the positive root of this equation and is the largest value of r that satisfies

$$1 - r - 2r^n + 2r^n \frac{[1 - \cos(d/(n-1))] + r[1 - \cos(\pi - d)/n)]}{1 + r} \geq 0,$$

for all $d \in \mathbf{R}$; in either case, for large n,

$$r_{n-1,n} = 1 - \frac{\log(2n)}{n} + \frac{\log(\log(2n))}{n} - o\left(\frac{1}{n}\right).$$

5 More General Matrices

While the norm behavior of functions of perturbed Jordan blocks is perfectly described by that of the minimal norm interpolating function on the unit disk, it appears very difficult, perhaps impossible, for general matrices A to find a set S in the complex plane for which (5) and (6) hold with M/m of moderate size. The intuitive reason is that the minimal norm interpolating function may oscillate wildly between the interpolation points; that is, the constant of interpolation C_Λ in (12) may be huge. For example, if there is an interpolation point, say λ_1, at the origin, then δ in (14) is less than or equal to $\prod_{j=2}^{n} |\lambda_j|$, so C_Λ is greater than or equal to $\prod_{j=2}^{n}(1/|\lambda_j|)$. This number grows exponentially with n unless all but a fixed number of the λ_j's approach the boundary of $\mathcal{D}$ as $n \to \infty$.

Suppose the eigenvector matrix of A is ill-conditioned but its condition number is still much much smaller than C_Λ. It follows from (13) that m_{opt} will be very small if we take S to be the unit disk. We can choose a different set S which, when conformally mapped to the unit disk, brings the eigenvalues closer to the boundary, in such a way as to make C_Λ smaller, but then M_{opt} may be very large.

Suppose, for example, that the eigenvalues of A are uniformly distributed about a circle of radius $r < 1$. One can derive an expression for δ in (14) as follows [6, p. 284]: A finite Blaschke product with zeros at $\lambda_1, \ldots, \lambda_n$ is

$$B(z) = \prod_{j=1}^{n} \frac{z - \lambda_j}{1 - \bar{\lambda}_j z} = \frac{z^n - r^n}{1 - r^n z^n}. \tag{29}$$

Since the derivative of $B(z)$ is

$$B'(z) = \sum_{k=1}^{n} \frac{1 - |\lambda_k|^2}{(1 - \bar{\lambda}_k z)^2} \prod_{j \neq k} \frac{z - \lambda_j}{1 - \bar{\lambda}_j z},$$

it follows that

$$B'(\lambda_k) = \frac{1}{1 - r^2} \prod_{j \neq k} \frac{\lambda_k - \lambda_j}{1 - \bar{\lambda}_j \lambda_k},$$

and

$$\delta = (1 - r^2) \min_k |B'(\lambda_k)|.$$

Differentiating the second expression for $B(z)$ in (29) shows that

$$\delta = n r^{n-1} \frac{1 - r^2}{1 - r^{2n}}. \tag{30}$$

Now, we have already seen that if A is a perturbed Jordan block with $\nu = r^n$, then its eigenvalues are uniformly distributed about the circle of radius r and its eigenvector matrix is the Vandermonde matrix for the eigenvalues; in this case, taking $S = \mathcal{D}$, we found $m_{opt} = M_{opt} = 1$. Suppose now that n is even and A is a block diagonal matrix with two $n/2$ by $n/2$ diagonal blocks, one of which is a perturbed Jordan block with $\nu = r^{n/2}$ and the other of which is a perturbed Jordan block with $\nu = -r^{n/2}$:

$$A = \left(\begin{array}{cccc|cccc} 0 & 1 & & & & & & \\ & \ddots & \ddots & & & & & \\ & & 0 & 1 & & & & \\ r^{n/2} & & & 0 & & & & \\ \hline & & & & 0 & 1 & & \\ & & & & & \ddots & \ddots & \\ & & & & & & 0 & 1 \\ & & & & -r^{n/2} & & & 0 \end{array} \right). \tag{31}$$

Then A again has n eigenvalues uniformly distributed about the circle of radius r, but its eigenvector matrix is block diagonal with the blocks being the Vandermonde matrices for the even-numbered eigenvalues and for the odd-numbered eigenvalues (numbering them counterclockwise around the circle), respectively. The transposed diagonal blocks of the eigenvector matrix can each be written as the product of a unitary matrix and a diagonal matrix:

$$\sqrt{\frac{2}{n}}\begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & \omega & \dots & \omega^{n/2-1} \\ \vdots & \vdots & & \vdots \\ 1 & \omega^{n/2-1} & \dots & \omega^{(n/2-1)^2} \end{pmatrix}\begin{pmatrix} 1 & & & \\ & re^{i\theta} & & \\ & & \ddots & \\ & & & r^{n/2-1}e^{(n/2-1)i\theta} \end{pmatrix},$$

where $\theta = 0$ for the first block and $\theta = 2\pi/n$ for the second and $\omega = e^{4\pi i/n}$. The condition number of the eigenvector matrix is therefore $r^{n/2-1}$. It follows from (12), (15), and (30) that $m_{opt}(A, \mathcal{D}) \leq nr^{n/2-1}(1-r^2)/(1-r^{2n})$. For any fixed $r < 1$, this approaches 0 exponentially with n.

In order to get a larger value for m_{opt}, one might consider a smaller disk $\mathcal{D}_s$ of radius $s \in (r, 1)$. Unless $s \equiv s_n \to 1$ as $n \to \infty$ (in which case $m_{opt}(A, \mathcal{D}_s)$ will approach 0), however, the value for $M_{opt}(A, \mathcal{D}_s)$ will approach ∞ as $n \to \infty$. To see this, consider the function $p(A) = A^{n/2-1}$. This matrix has a 1 in the upper right corner of each diagonal block and $\pm r^{n/2}$ on the first subdiagonal of each diagonal block. Its norm is clearly greater than or equal to 1. Yet the maximum absolute value of the minimal norm interpolating function on $\mathcal{D}_s$ is certainly less than or equal to the maximum absolute value of $z^{n/2-1}$ on $\mathcal{D}_s$, which is $s^{n/2-1}$. It follows that $M_{opt}(A, \mathcal{D}_s) \geq 1/s^{n/2-1}$, which, for fixed $s < 1$, grows exponentially with n.

Thus, for n large, if we take S in (9) to be any disk centered at the origin, the ratio M/m for the matrix A in (31) will be huge. Perhaps there is a set on which this ratio is of moderate size, but we have not found such a set. In any case, the values M_{opt} and m_{opt} will likely be extremely sensitive to small changes in the set S, since the constant of interpolation C_Λ is so large.

6 Future Directions

For many matrices A, it seems difficult or impossible to find a set S for which (9) holds with M/m of moderate size. The reason has to do with the possibly highly oscillatory behavior of even minimal-norm interpolating functions. In a sense, as the matrix size increases, one is allowing more and more complicated functions – every function of a 2 by 2 matrix is really an affine linear function in that it can be written as a first degree polynomial in A, and every function of a matrix A of small dimension is really a low degree polynomial in A. Thus for matrices of small dimension, say, $n < 10$, the constant of interpolation (12) is usually of moderate size and the minimal norm interpolating function cannot have extremely large $\mathcal{L}^{\infty}$-norm. But for large n this is not the case; the constant of interpolation is typically huge.

In applications, the functions of interest – A^k, e^{tA}, etc. – usually are not highly oscillatory. Thus a more productive exercise might be to limit the class of functions in some appropriate way and look for a set S for which (9) is satisfied only for this limited class of functions. In studying the polynomial numerical hull of degree k, one limits the class of functions to polynomials of degree k or less, but one also relates $\|p(A)\|$ to the size of the polynomial p at points in $\mathbf{C}$ rather than to the size of the minimal norm interpolating function on some set $S \subset \mathbf{C}$. It might prove interesting to look for sets S where (9) holds for polynomials of some fixed degree or for other appropriate classes of functions.

Another option is to consider two different sets in the complex plane – one to give lower bounds on $\|p(A)\|$ and the other to give upper bounds. Based on Crouzeix's results [3, 4], the field of values is certainly one appropriate set to consider for upper bounds: $\|p(A)\| \leq 11.08\|p\|_{\mathcal{L}^{\infty}(W(A))}$. His conjecture that $\|p(A)\| \leq 2\|p\|_{\mathcal{L}^{\infty}(W(A))}$ is truly fascinating. Note from the arguments in this paper that we need only consider functions p of the form $B \circ g$, where g is a conformal mapping from $W(A)$ to $\mathcal{D}$ and B is a finite Blaschke product (of degree $n-1$ if A is an n by n matrix), since other functions p will satisfy $p(A) = \mu B(g(A))$ for some scalar μ and some such B and g and will have larger ∞-norm on $W(A)$ than $\mu B \circ g$. Thus the conjecture is equivalent to the statement that $\|B(g(A))\| \leq 2$, for all conformal mappings g from $W(A)$ to $\mathcal{D}$ and for all Blaschke products of degree at most $n-1$. We too have attempted to prove this conjecture or to disprove it by finding a counterexample numerically, so far without success in either direction.

References

[1] M. Abrahamse, *The Pick interpolation theorem for finitely connected domains*, Michigan Math. J. 26 (1979), pp. 195–203.

[2] C. Carathéodory and L. Fejér, *Über den Zusammenhang der Extremen von harmonischen Funktionen mit ihren Koeffizienten und über den Picard-Landauschen Satz*, Rend. Circ. Mat. Palermo, 32 (1911), pp. 218–239.

[3] M. Crouzeix, *Numerical range and functional calculus in Hilbert space*, J. Functional Analysis 244 (2007), pp. 668-690.

[4] M. Crouzeix, *Bounds for analytical functions of matrices*, Integr. Equ. Oper. Theory 48 (2004), pp. 461–477.

[5] V. Faber, A. Greenbaum, and D. Marshall, *The polynomial numerical hulls of Jordan blocks and related matrices*, Lin. Alg. Appl. 374 (2003), pp. 231–246.

[6] J. B. Garnett, *Bounded Analytic Functions*, Springer, 2007.

[7] W. Gautschi, *On inverses of Vandermonde and confluent Vandermonde matrices*, Numer. Math. 4 (1962), pp. 117–123.

[8] W. Gautschi, *On inverses of Vandermonde and confluent Vandermonde matrices III*, Numer. Math. 29 (1978), pp. 445–450.

[9] C. Glader and M. Lindström, *Finite Blaschke product interpolation on the closed unit disc*, J. Math. Anal. Appl. 273 (2002), pp. 417–427.

[10] G. M. Goluzin, *Geometric Theory of Functions of a Complex Variable*, v. 26, Translations of Mathematical Monographs, American Mathematical Society, 1969.

[11] A. Greenbaum, *Generalizations of the field of values useful in the study of polynomial functions of a matrix*, Lin. Alg. Appl. 347 (2002), pp. 233–249.

[12] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, SIAM, 1996.

[13] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1990.

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

[15] P. Koosis, *Introduction to H_p Spaces*, 2nd ed., Cambridge Tracts in Mathematics 115, Cambridge University Press, 1998.

[16] O. Nevanlinna, *Convergence of Iterations for Linear Equations*, Birkhäuser, Basel, 1993.

[17] O. Nevanlinna, *Hessenberg matrices in Krylov subspaces and the computation of the spectrum*, Numer. Funct. Anal. and Optimiz. 16 (1995), pp. 443–473.

[18] I. Schur and G. Szegö, *Über die Abschnitte einer im Einheitskreise beschränkten Potenzreihe*, Sitzungsberichte Preuss. Akad. Wiss. Berlin (1925), pp. 545–560.

[19] L. N. Trefethen and M. Embree, *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*, Princeton University Press, 2005.

[20] V. Vinnikov and S. Fedorov, *The Nevanlinna-Pick interpolation problem in multiply connected domains*, Jour. of Math. Sciences 105 (2001), pp. 2109–2126.