K-spectral Sets and Functions of Nonnormal Matrices

Trevor Caldwell

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Anne Greenbaum, Chair

Bernard Deconinck

Randall LeVeque

Program Authorized to Offer Degree:
Applied Mathematics

University of Washington

Abstract

K-spectral Sets and Functions of Nonnormal Matrices

Trevor Caldwell

Chair of the Supervisory Committee:
Professor Anne Greenbaum
Applied Mathematics

In this thesis, we study K-spectral sets and use them to bound norms of functions of nonnormal matrices. For a fixed constant $K > 0$, the set Ω is said to be a K-spectral set for a matrix A if the spectrum $\Lambda(A)$ is contained in Ω and the inequality $\|f(A)\| \leq K \sup_{z\in\Omega} |f(z)|$ holds for all rational functions defined on Ω with poles outside of Ω. K-spectral sets are useful for studying nonnormal matrices, for which the asymptotic behavior of $\|f(A)\|$ suggested by the eigenvalues may not agree well with the short-time or transient behavior.

We extend a result of Crouzeix and Palencia [17], who show that the numerical range $W(A) = \{\langle v^*Av\rangle : v \in \mathbb{C}^n, \|v\| = 1\}$ is a $(1 + \sqrt{2})$-spectral set for any $n \times n$ matrix A, to sets in the complex plane which do not necessarily contain $W(A)$. This allows us to study more general K-spectral sets of interest, such as disks or half-planes. We also find some special cases in which the constant $(1 + \sqrt{2})$ for $W(A)$ can be replaced by 2, which is the value conjectured by Crouzeix. Additionally, we also provide details of our numerical studies related to Crouzeix's conjecture and K-spectral sets. This includes the construction of functions using numerical conformal mapping and optimization procedures to find bounds for K-spectral sets, as well as other bounds found via our extension of the Crouzeix-Palencia arguments or the numerical construction of near-normal dilations of matrices. We compare different analytical and numerical bounds found using these methods, and illustrate how these can be used for potential applications of interest.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I am indebted to my advisor, Anne Greenbaum. She has introduced me to some of the most challenging areas of research in applied mathematics and shown a genuine excitement for working on difficult problems. She has supported me through difficult times, and for that I am eternally grateful. I would also like to thank the rest of my supervisory committee: Bernard Deconinck, Ioana Dumitriu, and Randy LeVeque, for your time and effort.

I would also like to thank our collaborators through the years, in particular Kenan Li, Nick Trefethen, and Michel Crouzeix. Working with Kenan has certainly made the research experience less lonely. Nick's passion for mathematics is unparalleled, and seeing that first-hand has encouraged me through my research. Michel is one of the sharpest minds in the field, and his close ties with Anne have been invaluable for our work. Also, I would like to thank Mark Embree and Michael Overton for organizing conferences and workshops where I could discuss our work.

Finally, I would like to thank my friends and the rest of the Applied Math department. In particular, thanks to Ben Segal and Don Rim for being close friends throughout my time here, and thank you to all the other members of my cohort for their help and friendship. Outside the department, thank you to Kiley Sobel for being such a loving and understanding companion, and to my friends Sean Laguna and Russell Transue for reminding me that it is okay to laugh and have fun every once in a while.

DEDICATION

This is dedicated to my family: my mother, my father, Austen, Darby, Dylan, and Ariel. Your love and support has encouraged me to always keep pursuing my dreams.

NOTATION AND ABBREVIATIONS

$\mathbb{C}$	:	The complex plane
$\mathbb{D}$	:	The open unit disk
$\|A\|$	:	The induced 2-norm of a matrix A
$\|f\|_\Omega$	:	The supremum of the absolute value of f in a set $\Omega \subset \mathbb{C}$
$W(A)$	:	The numerical range (field of values) of a matrix A
$\Lambda(A)$	:	The spectrum of a matrix A
σ_k	:	The kth largest singular value of a given matrix
$C(\Omega)$	:	The set of continuous functions on a set Ω
$\mathcal{H}(\Omega)$	:	The set of analytic functions in Ω
$\mathcal{A}(\Omega)$	:	The set of functions $\mathcal{H}(\Omega) \cap C(\overline{\Omega})$
φ	:	A bijective conformal mapping from a set Ω to $\mathbb{D}$
$R(z, A)$	:	The resolvent of a matrix $(zI - A)^{-1}$
$B(z)$	:	A finite Blaschke product
$c_\Omega(A)$	:	The minimal constant such that for all $f \in \mathcal{A}(\Omega)$, $\|f(A)\| \leq c_\Omega \|f\|_\Omega$
$\overline{\Omega}$	:	The closure of a set Ω
$\overline{z}, \overline{f(z)}$	:	The complex conjugate of a value z or $f(z)$
$\omega(A)$	:	The numerical abscissa: $\max_{z \in W(A)} \mathrm{Re}(z)$
$\alpha(A)$	:	The spectral abscissa: $\max_{z \in \Lambda(A)} \mathrm{Re}(z)$
$r(A)$	:	The numerical radius: $\max_{z \in W(A)} \|z\|$
$\rho(A)$	:	The spectral radius: $\max_{z \in \Lambda(A)} \|z\|$

Chapter 1

INTRODUCTION

1.1 Eigenvalues and nonnormality

Many problems in applied mathematics involve estimating $\|f(A)\|$, where A is a matrix, f is a complex-valued function defined on a subset of the complex plane containing the spectrum (eigenvalues) of A, and $\|\cdot\|$ denotes the induced 2-norm of a matrix. For instance, the stability of a system of differential equations $x'(t) = Ax(t)$ is determined by $\|e^{tA}\|$, and the stability of the finite difference scheme $x_k = Ax_{k-1}$ depends on $\|A^k\|$. In practice, the spectrum is the most common tool to answer these stability questions, and is indeed very useful in asymptotic regimes. If the spectrum $\Lambda(A)$ lies in the left half-plane (for e^{tA}) or in the open unit disk (for A^k), then the solution must eventually decay as $t, k \to \infty$. However, one finds that for nonnormal matrices, the asymptotic behavior of these functions suggested by the eigenvalues may not agree well with the short-time or transient behavior. For example, if A is an $n \times n$ Jordan block with eigenvalue 0,

$$A = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ & & & 0 \end{pmatrix},$$

then $\|A^k\| = 1$ for $1 \leq k < n$, but $\Lambda(A) = \{0\}$.

If A is a normal matrix, then it is unitarily diagonalizable, i.e. $A = Q\Lambda Q^*$ where Λ is a diagonal matrix and Q is a unitary matrix. More generally, if A is a diagonalizable matrix

so that $A = V\Lambda V^{-1}$, then for the dynamical system $x'(t) = Ax(t)$ we have

$$\begin{aligned}\|x(t)\| &= \|e^{tA}x(0)\| \\ &\leq \|e^{tA}\|\ \|x(0)\| \\ &= \|Ve^{t\Lambda}V^{-1}\|\ \|x(0)\| \\ &\leq \|V\|\ \|V^{-1}\| \max_{\lambda\in\Lambda(A)} e^{\lambda t}\|x(0)\|.\end{aligned}$$

Notice that if A were normal, then V could be taken to be unitary so that $\|V\| = \|V^{-1}\| = 1$, and we have a tight upper bound on the behavior of $\|x(t)\|$ based only on the spectrum.

For a function f analytic in a neighborhood of $\Lambda(A)$ and diagonalizable $A = V\Lambda V^{-1}$, we have a similar bound based on the decomposition $f(A) = Vf(\Lambda)V^{-1}$, namely

$$\|f(A)\| \leq \kappa(V) \max_{\lambda\in\Lambda(A)} |f(\lambda)|. \tag{1.1}$$

When A is highly nonnormal, in the sense that the condition number $\kappa(V) = \|V\|\|V^{-1}\| \gg 1$, then this bound becomes weaker, as large condition numbers may significantly overestimate $\|f(A)\|$. Thus, alternative means of estimating bounds for $\|f(A)\|$ are necessary for nonnormal matrices.

1.2 Functions of matrices

As we wish to study norms of functions of matrices, let us recall the various definitions of a function of a matrix $f(A)$; see, for example, [29]. As before, assume that f is analytic in a neighborhood of the spectrum of A. Recall that any $A \in \mathbb{C}^{n\times n}$ can be expressed in the Jordan canonical form $A = PJP^{-1}$, where J is a block diagonal matrix of Jordan blocks $J = \mathrm{diag}(J_1, J_2, \ldots, J_p)$ of the form

$$J_k = \begin{pmatrix} \lambda_k & 1 & & \\ & \lambda_k & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_k \end{pmatrix} \in \mathbb{C}^{m_k\times m_k},$$

where $m_1 + m_2 + \cdots + m_p = n$. We can define the matrix function $f(A)$ using the Jordan blocks as follows:

$$f(A) = Pf(J)P^{-1} = P\text{diag}(f(J_k))P^{-1},$$

where

$$f(J_k) = \begin{pmatrix} f(\lambda_k) & f'(\lambda_k) & \dots & \frac{f^{(m_k-1)}(\lambda_k)}{(m_k-1)!} \\ & f(\lambda_k) & \ddots & \vdots \\ & & \ddots & f'(\lambda_k) \\ & & & f(\lambda_k) \end{pmatrix}.$$

Another useful definition uses a generalization of the Cauchy integral theorem,

$$f(A) = \frac{1}{2\pi i}\int_\Gamma f(z)(zI - A)^{-1}dz, \tag{1.2}$$

where f is analytic on and inside a closed contour Γ that encloses $\Lambda(A)$. Both of these approaches are helpful ways to view matrix functions, as the Jordan normal form provides a convenient algebraic form for function evaluation, while the resolvent-based definition is useful for bounding functions on different domains. Additionally, the resolvent form provides a natural extension to infinite dimensional operators on Hilbert or Banach spaces.

In the finite-dimensional case, where both definitions apply, they are equivalent. Let $f(z)$ be a given complex-valued analytic function on a simply connected open set $\Omega \subset \mathbb{C}$, and let $\Gamma \subset \Omega$ be any simple closed rectifiable curve that strictly encloses all of the eigenvalues of A. Additionally, suppose that A is diagonalizable. Then, $A = V\Lambda V^{-1}$ with $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, and the Cauchy integral theorem gives

$$\begin{aligned} \frac{1}{2\pi i}\int_\Gamma f(z)(zI - A)^{-1}dz &= \frac{1}{2\pi i}\int_\Gamma f(z)(zI - V\Lambda V^{-1})^{-1}dz \\ &= V\ \text{diag}\left[\frac{1}{2\pi i}\int_\Gamma f(z)(z-\lambda_1)^{-1}dz, \dots, \frac{1}{2\pi i}\int f(z)(t-\lambda_n)^{-1}dz\right]V^{-1} \\ &= V\ \text{diag}(f(\lambda_1), \dots, f(\lambda_n))V^{-1} \\ &= f(A), \end{aligned}$$

which gives the desired result for diagonalizable matrices. This can be extended to the general case by taking appropriate limits - see e.g. Theorem 6.2.28 in [30] for details.

1.3 K-spectral sets

Having defined suitable interpretations of $f(A)$, we define K-spectral sets and determine some of their basic properties. While we mostly consider only finite-dimensional matrices corresponding to the Hilbert space $\mathbb{C}^n$ with the standard inner product throughout this paper, we define K-spectral sets in the general setting of a complex Hilbert space H.

Suppose that $A \in \mathcal{L}(H)$ is a bounded linear operator on H. Let Ω be a closed set in the complex plane. For a constant $K > 0$, the set Ω is a K-spectral set for A if the spectrum $\Lambda(A)$ is in Ω and we have

$$\|f(A)\| \leq K\|f\|_\Omega, \tag{1.3}$$

for every $f \in \mathcal{R}(\Omega)$, the set of complex-valued rational functions defined on Ω with poles outside of Ω. Here, $\|\cdot\|_\Omega$ denotes the ∞-norm of f on Ω. In particular, Ω is a spectral set for A if it is a K-spectral set with $K = 1$. Spectral sets were first introduced by John von Neumann [44], who proved the inequality

$$\|f(A)\| \leq \|f\|_{\mathbb{D}}, \tag{1.4}$$

where f is a polynomial (more generally, a rational function with poles outside the closed unit disk $\overline{\mathbb{D}}$, or any function that is analytic in $\mathbb{D}$ and continuous on the boundary) and A is a Hilbert space contraction, meaning $\|A\| \leq 1$. In other words, the closed unit disk is a spectral set for contraction operators. If A is a normal matrix, then the spectrum $\Lambda(A)$ is a minimal spectral set, i.e. a spectral set having no proper closed subset which is spectral.

1.3.1 Numerical range

The numerical range (also called the field of values) of a matrix $A \in \mathbb{C}^{n\times n}$ is the set

$$W(A) = \{\langle Av, v\rangle : v \in \mathbb{C}^n, \|v\| = 1\}. \tag{1.5}$$

Clearly $\Lambda(A) \subset W(A)$, as we can simply take v to be a normalized eigenvector corresponding to eigenvalue λ, and $v^*Av = v^*v\lambda = \lambda$. The numerical range is a closed, convex subset of

$\mathbb{C}$ [30] that contains the convex hull of the spectrum $\Lambda(A)$, and if A is normal, then this containment is an identity. For highly nonnormal matrices, the numerical range may be considerably larger than the minimal convex set containing the spectrum. Then it may give extra information about the matrix that can be used for estimating norms of matrix functions.

We first note some basic properties of the numerical range; see, for example, [27] and [43]. Note that if $z \in W(A)$, then

$$\text{Re } z = \frac{z + \bar{z}}{2} = v^* \left(\frac{A + A^*}{2} \right) v,$$

for some unit vector v. It follows that the real part of the numerical range is

$$\text{Re } (W(A)) = \left[\lambda_{min} \left(\frac{A + A^*}{2} \right), \lambda_{max} \left(\frac{A + A^*}{2} \right) \right].$$

The rightmost part of the numerical range is an important quantity in applications, so we define the numerical abscissa as

$$\omega(A) = \max_{z \in W(A)} \text{Re } z = \lambda_{\max} \left(\frac{A + A^*}{2} \right). \tag{1.6}$$

This idea can also be used to determine the intersection of $W(A)$ with any line in the complex plane, forming the basis of the standard algorithm used to compute points on the boundary of the numerical range [30].

The numerical range is useful in the context of estimating norms of functions of matrices, especially for initial short-time behavior. For example, we can show that the rightmost point in $W(A)$ determines the initial slope of $\|e^{tA}\|$ at $t = 0$. Let x_0 be any vector such that $\|x_0\| = 1$. We have

$$\begin{aligned} \frac{d}{dt} \|e^{tA} x_0\| \Big|_{t=0} &= \frac{d}{dt} \left(x_0^* e^{tA^*} e^{tA} x_0 \right)^{1/2} \\ &= \frac{d}{dt} \left(x_0^* (I + tA^*)(I + tA) x_0 \right)^{1/2} \\ &= x_0^* \left(\frac{A + A^*}{2} \right) x_0. \end{aligned}$$

In fact, for all $t \geq 0$, it can be shown that

$$\|e^{tA}\| \leq e^{t\omega(A)}.$$

Note that $\|e^{tA}\| \leq 1$ for all $t \geq 0$ if and only if $\omega(A) \leq 0$.

We define the numerical radius as

$$r(A) = \sup_{z \in W(A)} |z|. \tag{1.7}$$

This satisfies the following power inequality

$$\|A^k\| \leq 2(r(A))^k,$$

a result due to Berger [8]. While the downside of the numerical range is that it may be much larger than the spectrum for nonnormal matrices and can give significant overestimates after initial or transient behavior, it does have the advantage of being robust to perturbations, due to the fact [43]:

$$W(A + E) \subset W(A) + \overline{\mathbb{D}}_{\|E\|},$$

where $\overline{\mathbb{D}}_{\|E\|}$ denotes the closed disk about 0 of radius $\|E\|$ and the set sum denotes the set of all points $z + e$ where $z \in W(A)$ and $e \in \overline{\mathbb{D}}_{\|E\|}$. Finally, we mention that Michel Crouzeix has conjectured [12, 13] that the numerical range is a 2-spectral set. Though this conjecture is simple to state and all numerical experiments thus far suggest that it is true, it has successfully been proven only in a few very simple special cases. This conjecture and the subsequent work has been a major driving force in recent years for the study of functions of nonnormal matrices, so Chapter 2 is devoted to its discussion. In particular, Crouzeix and Palencia proved that the numerical range is a $(1 + \sqrt{2})$-spectral set in [17].

1.3.2 The Kreiss matrix theorem

The Kreiss matrix theorem [33] concerns the chacterization of matrices and families of matrices which are power bounded. Define

$$\mathcal{P}(A) = \sup_{k \geq 0} \|A^k\|.$$

The matrix A is said to be power bounded if $\mathcal{P}(A) < \infty$. One may be concerned with bounding an entire family of matrices when studying the stability of discretizations of differential equations, since one needs to consider a family of matrices as the mesh width h and/or time step Δt tends to 0. Kreiss related the power bound for A with a constant now known as the Kreiss constant (with respect to matrix powers), which is defined as the smallest C for which

$$\|(z - A)^{-1}\| \leq \frac{C}{|z| - 1} \qquad \forall z \in \mathbb{C}, |z| > 1,$$

or equivalently,

$$\mathcal{K}(A) = \sup_{|z|>1} (|z| - 1)\|(z - A)^{-1}\|. \tag{1.8}$$

This constant provides a measure of how fast the resolvent norm blows up as z approaches $\mathbb{D}$, and gives bounds on powers of any n by n matrix A as follows:

$$\mathcal{K}(A) \leq \mathcal{P}(A) \leq en\mathcal{K}(A). \tag{1.9}$$

The proof is based on a complicated argument based on the resolvent that can be found in [43]. In particular, this provides a bound on the transient growth that powers of nonnormal matrices may exhibit. The constant in (1.8) has been reduced over the course of three decades due to the cumulative efforts of several authors, finally resulting in the constant en [40] in (1.9). The bound in (1.9) shows that for a fixed matrix A, all powers remain bounded so long as its Kreiss constant $\mathcal{K}(A)$ is finite, which is true if and only if $\rho(A) \leq 1$ with no defective eigenvalues on the unit circle. It can also be used to show that a family of matrices is *uniformly* power bounded if and only if there is a uniform bound on the Kreiss constants $\mathcal{K}(A)$ of all matrices in the family, so long as the family is of the same fixed dimension.

We can describe analogues of (1.8) and (1.9) for matrix exponentials. Define the Kreiss constant of A with respect to the left half-plane by

$$\mathcal{K}_e(A) = \sup_{\mathrm{Re}(z)>0} (\mathrm{Re}(z))\|(z - A)\|^{-1}. \tag{1.10}$$

We have, for any $n \times n$ matrix A,

$$\mathcal{K}_e(A) \leq \sup_{t\geq 0} \|e^{tA}\| \leq en\mathcal{K}_e(A). \tag{1.11}$$

1.3.3 ϵ-pseudospectrum

Another K-spectral set commonly used to study matrices and operators is the ϵ-pseudospectrum, which was introduced and popularized by Nick Trefethen in the 1990s - see [43] for a comprehensive review on the subject. There are several equivalent definitions of the ϵ-pseudospectrum $\Lambda_\epsilon(A)$, for a given $\epsilon > 0$ and matrix $A \in \mathbb{C}^{n\times n}$:

$$\begin{aligned}\Lambda_\epsilon(A) &= \{z \in \mathbb{C} : z \in \Lambda(A+E), \text{ for some } E \text{ with } \|E\| < \epsilon\} \\ &= \{z \in \mathbb{C} : \|(zI-A)^{-1}\| > 1/\epsilon\} \\ &= \{z \in \mathbb{C} : \|Av - zv\| < \epsilon \text{ for some unit vector } v\}.\end{aligned}$$

These different definitions can be useful in different contexts, but the second definition based on the resolvent norm is the most useful definition for the purpose of investigating the behavior of functions of matrices.

To illustrate the use of pseudospectra as an alternative to the traditional spectrum for bounding functions of matrices, we list some basic definitions and results below. The ϵ-pseudospectral abscissa of A measures the rightmost extent of $\Lambda_\epsilon(A)$:

$$\alpha_\epsilon(A) = \sup_{z\in\Lambda_\epsilon(A)} \text{Re } z, \tag{1.12}$$

and the ϵ-pseudospectral radius of A measures the maximum magnitude in $\Lambda_\epsilon(A)$:

$$\rho_\epsilon(A) = \sup_{z\in\Lambda_\epsilon(A)} |z|. \tag{1.13}$$

Also, $\alpha(A)$ and $\rho(A)$ denote the traditional spectral abscissa and spectral radius respectively. For any function f analytic in a neighborhood of $\Lambda(A)$, we have

$$\|f(A)\| \geq \max_{\lambda\in\Lambda(A)} |f(\lambda)|,$$

and equality holds when A is normal. Thus:

$$\|e^{tA}\| \geq e^{t\alpha(A)}, \quad \|A^k\| \geq \rho(A)^k.$$

For a simple upper bound using pseudospectra, we have the following theorem. Let Γ_ϵ be the boundary of the ϵ-psuedospectrum for some $\epsilon > 0$, and suppose f is an analytic function on Γ_ϵ and its interior. Then,

$$\|f(A)\| \leq \frac{L_\epsilon}{2\pi\epsilon} \max_{z\in\Gamma_\epsilon} |f(z)|, \tag{1.14}$$

where L_ϵ denotes the arc length of Γ_ϵ - this can be shown using the Cauchy integral form of $f(A)$ and the resolvent definition of pseudospectra. Note that this implies that the closure of the ϵ-pseudospectrum is a K-spectral set for $K = L_\epsilon/(2\pi\epsilon)$. For the special cases considered above we have

$$\|e^{tA}\| \leq \frac{L_\epsilon}{2\pi\epsilon} e^{t\alpha_\epsilon(A)}, \quad \|A^k\| \leq \frac{L_\epsilon}{2\pi\epsilon}\rho_\epsilon(A)^k.$$

Typically, larger values of ϵ yield tighter bounds for small t and k, while smaller values are better for larger t and k. We can think of pseudospectra as interpolating between the spectrum (which gives asymptotic behavior and corresponds to $\epsilon \to 0$) and the numerical range (which gives a better indication of initial behavior, and corresponds to larger values of ϵ).

We can also express the Kreiss constants (1.8) and (1.10) in terms of pseudospectra as

$$\mathcal{K}(A) = \sup_{\epsilon>0} \frac{\rho_\epsilon(A) - 1}{\epsilon}, \qquad \mathcal{K}_e(A) = \sup_{\epsilon>0} \frac{\alpha_\epsilon(A)}{\epsilon}.$$

Thus, the Kreiss constants can be viewed as a measure of how far the pseudospectra of A extend outside the unit circle or into the right half-plane.

1.3.4 Other K-spectral sets

There are still other K-spectral sets of interest beyond those discussed thus far. For example, one may ask when the left half-plane is a K-spectral set for a certain matrix A in order to bound the norm of the matrix exponential $\|e^{tA}\|$, or similarly ask when the unit disk is a K-spectral set to bound norms of matrix powers $\|A^k\|$. These sets are of particular interest due to the possibilities of applications, so we focus on them later starting in Chapter 3.

Finally, we close this section by mentioning other candidate K-spectral sets which have been studied in recent years, many of which can be found in Badea and Beckermann's reference article on K-spectral sets [2]. For matrices with special structure, there are some simple examples of spectral sets. An $n \times n$ matrix is normal if and only if its spectrum $\Lambda(A)$ is a spectral set for A; the unit circle $\mathbb{T}$ is a spectral set for A if and only if A is unitary; and the real axis $\mathbb{R}$ is a spectral set for A if and only if A is Hermitian (self-adjoint).

Beckermann, Crouzeix, and Delyon determined many K-spectral sets containing the numerical range (i.e. $W(A) \subset \Omega$); see [4] and [6]. For example, they showed that a convex sector or strip containing $W(A)$ is $\left(2 + \frac{2}{\sqrt{3}}\right)$-spectral. If the boundary of Ω is a parabola or hyperbola, then Ω is also $\left(2 + \frac{2}{\sqrt{3}}\right)$-spectral. Similar bounds were found for ellipses and regular polygons containing $W(A)$. However, all of these results have since been made obsolete due to a phenomenal recent result by Crouzeix and Palencia, who show that the numerical range is a $(1 + \sqrt{2})$-spectral set. We give an outline of their proof in Chapter 2. In Chapter 3, we extend the Crouzeix-Palencia result to sets which do not necessarily contain the numerical range. In particular, we focus on disks and half-planes, and use our extension of the Crouzeix-Palencia result to give both analytical bounds on $\|f(A)\|$ in Chapter 3 and numerical bounds on $\|f(A)\|$ in Chapters 4 and 5.

Chapter 2

CROUZEIX'S CONJECTURE

2.1 Background and history

In [12] and [13], M. Crouzeix made the following conjecture:

Conjecture 2.1.1. *For any matrix $A \in \mathbb{C}^{n,n}$ and any polynomial p, we have*

$$\|p(A)\| \leq 2\|p\|_{W(A)}. \tag{2.1}$$

In other words, the numerical range of a matrix is conjectured to be a 2-spectral set. The groundwork for this conjecture had been laid in earlier papers, beginning as early as von Neumann's work on spectral sets (in particular, disks and half-planes) in [44]. The modern development of the study of the numerical range as a K-spectral set can be attributed to B. and D. Delyon in [18], who showed that the numerical range is K-spectral with $K = 3 + (2\pi \text{ diameter}(\Omega)^2/ \text{ area}(\Omega))^3$, where Ω is any compact convex set containing $W(A)$. The key ingredient in their proof was the use of an integral representation formula for operators based on the Cauchy transform and a double layer potential. In the two decades following this work, Crouzeix and others have tackled the problem of refining the constant K for the general case, as well as for various special classes of matrices.

We already have bounds of the form (1.1) based on the spectrum of a diagonalizable matrix $A = V\Lambda V^{-1}$, which imply that $W(A)$ is 2-spectral if V satisfies $\kappa(V) \leq 2$. In particular, normal matrices satisfy the bound (2.1), in which case $W(A)$ is the convex hull of $\Lambda(A)$ and is also a spectral set. Below, we list some of the functions f and matrix classes A for which it is known that $\|f(A)\| \leq 2\|f\|_{W(A)}$:

- $f(z) = (z - z_0)^n$. This is originally due to a power inequality proved by Berger [8].

- $W(A)$ *is a circular disk.* This result [35] is due to Okubo and Ando.

- A *is a* 2×2 *matrix.* This proof uses a similarity transformation and the explicit formula for the conformal map from the numerical range (here an elliptical disk) onto the unit disk [12].

- A *is a quadratic matrix.* This means that $(A - z_1 I)(A - z_2 I) = 0$ for some $z_1, z_2 \in \mathbb{C}$. Then, A is unitarily similar to a direct sum of 1×1 and 2×2 matrices, the numerical range is an ellipse, and (2.1) follows by reduction to the 2×2 case [15].

- A *is a* 3×3 *matrix with elliptical numerical range centered at an eigenvalue.* This means that A is of the form

$$\begin{pmatrix} a & b_1 & 0 \\ c_1 & a & b_2 \\ 0 & c_2 & a \end{pmatrix}, \qquad a, b_k, c_k \in \mathbb{C}.$$

  This proof [22] uses similar techniques as in [12] for the 2×2 case, but uses several different similarities for different parameter regimes.

- A *is of the form* $aI + DP$ *or* $aI + PD$*, where* $a \in \mathbb{C}$*,* D *is a diagonal matrix, and* P *is a permutation matrix.* The inequality (2.1) was first shown to hold [26] for perturbed Jordan blocks of the form

$$J_\nu = \begin{pmatrix} \lambda & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ \nu & & & \lambda \end{pmatrix}, \qquad \lambda, \nu \in \mathbb{C}.$$

  In this case, $W(A)$ has n-fold symmetry, and the image of A by a conformal mapping from $W(A)$ onto the unit disk $\mathbb{D}$ has the form cA for some constant c. This was generalized to a larger class of matrices in [10], and to matrices of the form $aI + DP$ in [22].

Remark: In all of these cases, while the conformal mapping φ taking $W(A)$ onto the unit disk may be a complicated function, $\varphi(A)$ is just a linear function of A: $\varphi(A) = \alpha A + \beta I$. Recall that, for A diagonalizable, $\varphi(A)$ is completely determined by the values of φ at the eigenvalues of A so that $\varphi(A)$ may have a very simple formula even when the general formula for $\varphi(z)$ is complicated.

We examine some of these proofs in greater detail to illustrate the techniques involved, starting with Crouzeix's proof of the 2×2 case [12]. This proof uses a different set of tools from the integral representation formula used in [18], as most relevant quantities can be explicitly calculated for 2×2 matrices. We provide a sketch of the proof given in [12] below.

We define

$$\psi_\Omega(A) := \sup\{\|f(A)\| : f \in \mathcal{H}^\infty(\Omega), \|f\|_{L^\infty(\Omega)} \leq 1\}, \tag{2.2}$$

where $\mathcal{H}^\infty(\Omega)$ denotes the Hardy space

$$\mathcal{H}^\infty(\Omega) := \{f : f \text{ holomorphic and bounded in } \Omega\}.$$

Note that $\psi_\Omega(A)$ has the following properties. If φ is a holomorphic isomorphism from Ω onto the unit disk $\mathbb{D}$, then $\psi_\Omega(A) = \psi_\mathbb{D}(\varphi(A))$, since for any $f \in \mathcal{H}^\infty(\Omega)$ we can write $f(A) = g(\varphi(A))$ with $g := f \circ \varphi^{-1}$, and f and g have the same infinity norm. Also, if $A = X^{-1}TX$, then we have $f(A) = X^{-1}f(T)X$, and so $\psi_\Omega(A) = \psi_\Omega(T)$ when X is a unitary matrix. Due to the Schur decomposition, we need only consider upper triangular matrices with eigenvalues of A along the diagonal.

To prove the 2×2 case, we begin by considering $\Omega = \mathbb{D}$. From the above, we need only consider matrices of the form

$$T = \begin{pmatrix} \lambda_1 & \gamma \\ 0 & \lambda_2 \end{pmatrix}, \qquad \lambda_1, \lambda_2 \in \mathbb{D}, \gamma \in \mathbb{C}. \tag{2.3}$$

Furthermore, we can replace γ by $|\gamma|$ via a unitary similarity. Also, we can always find an automorphism ϕ such that $\phi(\lambda_1) + \phi(\lambda_2) = 0$ and $\phi(\lambda_1) \in (0, 1)$, meaning it is sufficient

to consider the case where T is of the form

$$T = \begin{pmatrix} \lambda & 2\delta \\ 0 & -\lambda \end{pmatrix}, \qquad \lambda \in (0,1), \quad \delta \geq 0. \tag{2.4}$$

We can easily compute $\|T\| = \delta + \sqrt{\lambda^2 + \delta^2}$. We claim that

$$\psi_{\mathbb{D}}(T) = \max(1, \|T\|). \tag{2.5}$$

If $\|T\| \leq 1$, then von Neumann's inequality asserts that $\phi_{\mathbb{D}}(T) = 1$, so we need only consider when $\|T\| > 1$. Clearly we have $\psi_{\mathbb{D}}(T) \leq \|T\|$, as we can take $f(z) = z$ in the definition of $\psi_{\mathbb{D}}(T)$. The opposite inequality is shown via a similarity transformation of the form

$$X = \begin{pmatrix} 1 & \mu \\ 0 & \beta \end{pmatrix}, \quad \beta = \frac{1+\lambda^2}{2\sqrt{\lambda^2+\delta^2}}, \quad \mu = \frac{1-\lambda^2-2\beta\delta}{2\lambda}. \tag{2.6}$$

We compute

$$B := X^{-1}TX = \begin{pmatrix} \lambda & 1-\lambda^2 \\ 0 & -\lambda \end{pmatrix},$$

and note that $\|B\| = 1$, so $\psi_{\mathbb{D}}(B) = 1$ and $\psi_{\mathbb{D}}(T) \leq \|X\|\psi_{\mathbb{D}}(B)\|X^{-1}\| = \kappa(X)$, where $\kappa(X)$ denotes the condition number of X. After more algebra, one finds that $\kappa(X) = \|T\|$, which yields the result.

Let Ω be a convex set containing $W(A)$, let φ denote a holomorphic bijection from Ω onto the unit disk $\mathbb{D}$, and recall that $\psi_\Omega(T) = \psi_{\mathbb{D}}(\varphi(T))$. We invoke the previous result with the mapped matrix

$$\varphi(T) = \begin{pmatrix} \varphi(\lambda_1) & \gamma\varphi[\lambda_1,\lambda_2] \\ 0 & \varphi(\lambda_2) \end{pmatrix}, \qquad \varphi[\lambda_1,\lambda_2] := \frac{\varphi(\lambda_1)-\varphi(\lambda_2)}{\lambda_1-\lambda_2}. \tag{2.7}$$

Finally, we explicitly compute the value of the bound $\psi_\Omega(A)$ with $\Omega = W(A)$. If A is normal, then $\psi_\Omega(A) = 1$, so we need only consider nonnormal matrices. In this case it is known that $W(A)$ is always an elliptical disk. Due to the invariance of ψ_Ω under unitary similarity transformations, scaling, and translations, we can assume that $A = \begin{pmatrix} 1 & \gamma \\ 0 & -1 \end{pmatrix}$ and

that $\gamma > 0$. Then, $W(A)$ is an ellipse with foci ± 1 and minor axis γ. Setting $\gamma = \rho - 1/\rho$, we have that the major axis is $\rho + 1/\rho$, and the conformal map from $W(A)$ onto $\overline{\mathbb{D}}$ is known explicitly:

$$\varphi(z) = \frac{2z}{\rho} \exp\left\{ -\sum_{n \geq 1} \frac{(-1)^{n+1}}{n} \frac{2t_{2n}(z)}{1+\rho^{4n}} \right\},$$

where t_n denotes the nth Chebyshev polynomial. Inserting this into the expression for $\|\varphi(A)\|$ gives $\psi(A) = \psi_\Omega(A) = \rho\varphi(1)$. For 2 by 2 matrices, the disk case follows by continuity, so that $\psi(A) = 2$ if $W(A)$ is a disk.

The techniques used here involve the careful determination of a suitable similarity transformation, which is not obvious a priori, as well as the use of the explicit formula for the conformal map from the ellipse to the disk. This sort of approach may be best suited for simple low-dimensional cases, or matrices with very special structure. For example, Glader, Lindström, and Kirula use similar techniques in [22] in order to prove the conjecture for 3×3 matrices with elliptic numerical range centered at an eigenvalue, which have the form

$$\begin{pmatrix} a & b_1 & 0 \\ c_1 & a & b_2 \\ 0 & c_2 & a \end{pmatrix}, \qquad a, b_k, c_k \in \mathbb{C}. \tag{2.8}$$

They use similar techniques as Crouzeix does in [12], as the conformal map from the ellipse to the disk is known. However, they must go to much greater lengths to construct various similarity transformations X such that the condition number $\kappa(X) \leq 2$. First, they simplify matters by parameterizing the family of matrices they study to those of the form

$$\begin{pmatrix} 1 & q/r & r^2 - 1/r^2 \\ 0 & 0 & qr \\ 0 & 0 & -1 \end{pmatrix}, \qquad q > 0, 0 < r \leq 1. \tag{2.9}$$

Their analysis is much more involved, as they must construct different similarity transformations for different parameter regimes, as well as obtain tighter estimates on the values of the conformal map. This entails numerical optimization to provide structural clues for the

optimal similarity transformations in tandem with elementary (but lengthy) optimization by hand. Despite the daunting amount of algebra and careful analysis done here, their proof still cannot cover all 3×3 matrices with elliptical numerical range, let alone all 3×3 matrices in general. One of the issues with extending to matrices with elliptic numerical range that is *not* centered at an eigenvalue is that we no longer have the simple relation $\varphi(A) = cA$ for some $c < 1$. Crouzeix uses interesting methods in [14] in an attempt to prove the conjecture for 3×3 nilpotent matrices, including asymptotic expansions of the conformal mapping and numerical optimization, but is not able to obtain a fully mathematical proof.

Aside from some reductions to the 2×2 case, the only other major class of matrices for which Crouzeix's conjecture has been proven is for certain perturbed Jordan blocks. The study of these types of matrices begins with [26], in which Greenbaum and Choi prove the conjecture for matrices of the form

$$J_\nu = \begin{pmatrix} \lambda & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ \nu & & & \lambda \end{pmatrix}, \qquad \lambda, \nu \in \mathbb{C}. \tag{2.10}$$

Again, the primary technique is to show that if ϕ is a bijective conformal map from $W(A)$ to $\mathbb{D}$, then $\varphi(A)$ is similar to a contraction via a similarity transformation X with condition number $\kappa(X) \leq 2$. The key insight here is that even though the numerical range of J_ν may satisfy a complicated equation and the conformal mapping φ from $W(J_\nu)$ to $\mathbb{D}$ may be a complicated function, the action on the eigenvalues is linear due to the n-fold symmetry of $W(J_\nu)$, and $\varphi(J_\nu) = c(J_\nu - \lambda I)$. This mapped matrix is diagonally similar to $J_{c^n\nu} - \lambda I$, which is a contraction for $|\nu| \leq c^{-n}$, and for $1 \leq c \leq 2^{1/(n-1)}$, the condition number of the similarity transformation is at most 2. Thus, Greenbaum and Choi reduce the problem to bounding the constant c in the conformal mapping in order to maintain both properties.

For $1 > |\nu| > 2^{-n/(n-1)}$, the eigenvector matrix of J_ν has condition number less than 2, so they can assume $\nu < 2^{-n/(n-1)}$. Bounds for various values of n and ν in this regime are obtained via estimates using disks inside the numerical range, as well as approximate

mappings which take regions slightly inside $W(J_\nu)$ to regions containing $\mathbb{D}$. The results here were generalized further by D. Choi in [10] to matrices of the form

$$J_\alpha = \begin{pmatrix} \lambda & \alpha_1 & & \\ & \ddots & \ddots & \\ & & \ddots & \alpha_{n-1} \\ \alpha_n & & & \lambda \end{pmatrix}, \qquad \lambda \in \mathbb{C}, \alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{C}^n. \tag{2.11}$$

This proof is slightly more involved, but has the same key feature of n-fold symmetry of the numerical range and uses similar methods as [26].

Though some of these partial results are impressive and promising, this method of constructing similarity transformations taking $\varphi(A)$ to a contraction is likely difficult to generalize. In all of the proven cases thus far, a key property was the simple action of ϕ on the eigenvalues of A, i.e. $\varphi(A) = cA$. Additionally, enough information was known about the conformal maps (an explicit formula for the ellipse cases, as well as useful symmetry properties and approximations for the perturbed Jordan blocks) so that it was possible to bound the constant c. This approach could still be useful for proving the conjecture for matrices satisfying $\varphi(A) = cA$, but different ideas are likely necessary to tackle more general cases.

Until recent developments, the best estimate for the general case was from Crouzeix in [13], which drew upon integral representation formulas used in the work of B. and F. Delyon in [18] in order to show that the numerical range is a K-spectral set with $2 \leq K \leq 11.08$. The integral estimates in this work are quite involved geometrically, and Crouzeix himself was convinced that a new approach would be needed to improve this bound, as noted in the following remark:

> There is no doubt that refinements are possible which would decrease this bound. We are convinced that our estimate is very pessimistic, but to improve it drastically, it is clear that we have to find a completely different method.

Despite this pessimistic outlook, this work still contained useful results and strategies. One

interesting result is that not only did Crouzeix prove the inequality

$$\|p(A)\| \leq 11.08 \sup_{z \in W(A)} |p(z)|, \tag{2.12}$$

for any square matrix A and polynomial p, but he also showed that a stronger version of this inequality holds. A set $\Omega \subset \mathbb{C}$ is a *complete* K-spectral set for A if

$$\|P(A)\| \leq K \sup_{z \in \Omega} \|P(z)\|, \tag{2.13}$$

for all matrix-valued polynomials P. By this, we mean that $P(z)$ is an $\ell \times m$ matrix whose (i,j)-entry is $p_{ij}(z)$ for some polynomial p_{ij}, and $P(A)$ is an $\ell \times m$ array of operators whose (i,j)-block is $p_{ij}(A)$. In [13], Crouzeix shows that the numerical range $W(A)$ is not only a K-spectral set, but a complete K-spectral set:

$$\|P(A)\| \leq K \sup_{z \in W(A)} \|P(z)\|, \tag{2.14}$$

for some K with $2 \leq K \leq 11.08$, for all $\ell \times m$ matrix-valued polynomials P, and all positive integers ℓ and m. This is sometimes referred to as the completely bounded version of Crouzeix's theorem. While the majority of this dissertation is primarily concerned with the standard form of Crouzeix's conjecture and K-spectral sets, we will return to the idea of complete K-spectral sets when discussing numerical experiments on matrix dilations.

Another key element in [13] is the use of the following double layer potential as a part of the integral representation formula for polynomials defined on $\overline{\Omega}$:

$$\mu(\sigma, z) = \frac{1}{\pi} \frac{d}{ds} \left(\arg(\sigma - z) \right) = \frac{1}{2\pi i} \left(\frac{e^{i\theta}}{\sigma - z} - \frac{e^{-i\theta}}{\overline{\sigma} - \overline{z}} \right). \tag{2.15}$$

Here, Ω is a bounded convex domain in $\mathbb{C}$ such that $W(A) \subset \Omega$, $\sigma(s)$ parameterizes the smooth boundary $\partial\Omega$ in terms of arclength s with $\theta = \arg\left(\frac{d\sigma}{ds}\right)$, and $z \in \overline{\Omega}$. If $z = \sigma$, then $\mu(\sigma, z)$ is the curvature of $\partial\Omega$ at σ. A key property of this potential is that its operator-valued form,

$$\mu(\sigma, A) = \frac{1}{2\pi i} \left(e^{i\theta} (\sigma - A)^{-1} - e^{-i\theta} (\overline{\sigma} - A^*)^{-1} \right), \tag{2.16}$$

is well-defined and positive semi-definite for $\sigma \in \partial\Omega$ when $W(A) \subset \Omega$. This property would again be useful in refinements made by Crouzeix and Palencia in [17], to be discussed in the following section.

2.2 Crouzeix-Palencia result

The optimal constant in Crouzeix's conjecture remained 11.08 for about a decade after the work done in [13]. However, a major breakthrough occurred in late 2016, when Palencia announced that he had reduced the constant from 11.08 to $1 + \sqrt{2}$ during a conference in Crete on evolution equations [36]. Crouzeix simplified Palencia's arguments, leading to a concise and beautiful joint paper [17]. We provide a sketch of the main ideas used in their proof, as extensions and applications of this work constitute a significant portion of this thesis.

Let $\Omega \subset \mathbb{C}$ be a smooth, bounded, convex domain which contains the numerical range $W(A)$. Using a sequence of smooth convex domains $\Omega_n \supset W(A)$, they show that

$$\|f(A)\| \leq (1 + \sqrt{2}) \sup_{z \in W(A)} |f(z)|, \tag{2.17}$$

which we refer to as the Crouzeix-Palencia result. Similar to earlier work in [18] and [13], the approach here uses the Cauchy transform and an integral representation formula based on the double layer potential. From the Cauchy integral formula, we have

$$f(A) = \frac{1}{2\pi i} \int_{\partial\Omega} (\sigma I - A)^{-1} f(\sigma) d\sigma. \tag{2.18}$$

If we parameterize $\partial\Omega$ by arclength s running from 0 to L, this becomes

$$f(A) = \int_0^L \frac{\sigma'(s)}{2\pi i} R(\sigma(s), A) f(\sigma(s)) ds, \tag{2.19}$$

where $R(\sigma, A)$ is the resolvent $(\sigma I - A)^{-1}$.

A key idea in [17] is the study of the Cauchy transform g of the conjugate of f,

$$g(A) = \int_0^L \frac{\sigma'(s)}{2\pi i} R(\sigma(s), A) \overline{f(\sigma(s))} ds. \tag{2.20}$$

Note that $\overline{f(\sigma(s))}$ is not analytic, so the Cauchy integral formula cannot be directly applied here. Crouzeix and Palencia analyzed the operator

$$S := f(A) + g(A)^* = \int_0^L \mu(\sigma(s), A) f(\sigma(s)) ds, \tag{2.21}$$

where the Hermitian operator $\mu(\sigma, A)$ is

$$\mu(\sigma(s), A) = \frac{\sigma'(s)}{2\pi i} R(\sigma(s), A) + \left[\frac{\sigma'(s)}{2\pi i} R(\sigma(s), A) \right]^*. \tag{2.22}$$

Note that this is the same double layer potential used in [18] and [13]. They argued that if Ω is a convex region containing $W(A)$, then $\mu(\sigma, A)$ is positive semidefinite for $\sigma \in \partial\Omega$. Using this fact, they show that for $f \in \mathcal{A}(\Omega)$ with $\|f\|_\Omega = 1$, there holds $\|S\| = \|f(A)+g(A)^*\| \leq 2$. Recall that $\mathcal{A}(\Omega)$ refers to the space of functions analytic inside Ω and continuous in $\overline{\Omega}$.

The remainder of [17] is devoted to relating $\|f(A) + g(A)^*\|$ to $\|f(A)\|$. Crouzeix and Palencia show that if $g(z)$ is defined in Ω by

$$g(z) = \frac{1}{2\pi i} \int_{\partial\Omega} \frac{\overline{f(\sigma)}}{\sigma - z} d\sigma, \tag{2.23}$$

then $g \in \mathcal{A}(\Omega)$ when extended continuously to the boundary $\partial\Omega$, and $\|g\|_\Omega \leq \|f\|_\Omega$. Using this fact in tandem with the previous results, they show that for $\Omega \supset W(A)$ and for all $f \in \mathcal{A}(\Omega)$,

$$\|f(A)\| \leq C_\Omega \sup_{z \in \Omega} |f(z)|, \quad C_\Omega \leq 1 + \sqrt{2}. \tag{2.24}$$

Taking limits as Ω tends to $W(A)$ gives the desired result regarding the numerical range.

Chapter 3

EXTENSIONS OF THE CROUZEIX-PALENCIA RESULT

We demonstrate how to extend the arguments of Crouzeix and Palencia in [17] to show that regions Ω in the complex plane that do *not* necessarily contain the numerical range $W(A)$ are K-spectral sets. In particular, we consider various disks and half-planes containing the spectrum of A. This chapter describes the work done in [9], as well as some additional applications using this extension.

The chapter is organized as follows. We begin by proving some basic theorems extending the results of Crouzeix and Palencia in [17] to regions Ω containing the spectrum of A but not necessarily all of $W(A)$. The techniques used are similar to those in [17]. See also [7] which also uses similar ideas to our extension but instead involves Faber polynomials. In the following section, we discuss the form of the function f that maximizes the ratio $\|f(A)\|/\|f\|_\Omega$. If A is an n by n matrix, then for any nonempty simply connected open set Ω containing the spectrum of A (but not all of $\mathbb{C}$), there exists a function f that attains $\max_{\hat{f}\in\mathcal{A}(\Omega)} \|\hat{f}(A)\|/\|\hat{f}\|_\Omega$, and the form of f is known to be [13, 21, 23]

$$f(z) = B \circ \varphi(z), \tag{3.1}$$

where φ is a bijective conformal mapping from Ω to the unit disk $\mathbb{D}$ and B is a Blaschke product of degree at most $n-1$,

$$B(z) = e^{i\theta} \prod_{j=1}^{n-1} \frac{z-\alpha_j}{1-\overline{\alpha}_j z}, \quad |\alpha_j| \le 1. \tag{3.2}$$

We prove some properties of this optimal B. In particular, we show that if $\|B(\varphi(A))\| > 1$ and if v_1 is a right singular vector of $B(\varphi(A))$ corresponding to the largest singular value σ_1, then $\langle Bv_1, v_1\rangle = 0$. Using this result, we can replace the bound $1+\sqrt{2}$ in (2.24) with 2 in some

special cases (where the bound has already been shown to hold, but by different means), which is the value conjectured by Crouzeix. We also use these results to derive bounds for regions containing the spectrum of A but not necessarily containing $W(A)$. Numerical studies of these properties and bounds for more general cases are discussed in Chapter 4.

3.1 Main results

As in the previous section, let A be a square matrix, and let Ω be a region with smooth boundary containing the spectrum of A in its interior. In [17], Crouzeix and Palencia show that the numerical range $W(A)$ is a $(1+\sqrt{2})$-spectral set for A, i.e. for any f analytic in the interior of $W(A)$ and continuous on its boundary,

$$\|f(A)\| \leq (1+\sqrt{2})\|f\|_{W(A)} = (1+\sqrt{2}) \sup_{z \in W(A)} |f(z)|. \tag{3.3}$$

Recall that for any function $f \in \mathcal{A}(\Omega) := \mathcal{H}(\Omega) \cap C(\overline{\Omega})$, we have

$$f(A) = \frac{1}{2\pi i} \int_{\partial\Omega} (\sigma I - A)^{-1} f(\sigma) d\sigma, \tag{3.4}$$

via the Cauchy integral formula. If we parameterize $\partial\Omega$ by arc length s from 0 to L, this becomes

$$f(A) = \int_0^L \frac{\sigma'(s)}{2\pi i} R(\sigma(s), A) f(\sigma(s)) ds, \tag{3.5}$$

where $R(\sigma, A)$ is the resolvent, $(\sigma I - A)^{-1}$. Crouzeix and Palencia also looked at

$$g(A) = \int_0^L \frac{\sigma'(s)}{2\pi i} R(\sigma(s), A) \overline{f(\sigma(s))} ds, \tag{3.6}$$

and the operator

$$S := f(A) + g(A)^* = \int_0^L \mu(\sigma(s), A) f(\sigma(s)) ds, \tag{3.7}$$

where the Hermitian operator $\mu(\sigma(s), A)$ is

$$\mu(\sigma(s), A) = \frac{\sigma'(s)}{2\pi i} R(\sigma(s), A) + \left[\frac{\sigma'(s)}{2\pi i} R(\sigma(s), A) \right]^*. \tag{3.8}$$

They argued that if Ω is a convex region containing $W(A)$, then $\mu(\sigma, A)$ is positive semidefinite for $\sigma \in \partial\Omega$. In order to extend their arguments to regions Ω that do not necessarily contain $W(A)$, we define

$$M(\sigma, A) := \mu(\sigma, A) - \lambda_{\min}(\mu(\sigma, A))I, \tag{3.9}$$

where $\lambda_{\min}(\mu(\sigma, A))$ is the minimum of the spectrum of $\mu(\sigma, A)$ at the point $\sigma \in \partial\Omega$. Thus, $M(\sigma, A)$ is positive semidefinite on $\partial\Omega$.

Using the same method of proof as in [17], we establish the following:

Lemma 3.1.1. *Let Ω be a region with smooth boundary containing the spectrum of A in its interior. For $f \in \mathcal{A}(\Omega)$ with $\|f\|_\Omega = 1$, let*

$$S = f(A) + g(A)^* + \gamma I, \qquad \gamma := -\int_0^L \lambda_{\min}(\mu(\sigma(s), A))f(\sigma(s))ds. \tag{3.10}$$

Then $\|S\| \leq 2 + \delta$, where

$$\delta = -\int_0^L \lambda_{\min}(\mu(\sigma(s), A))ds. \tag{3.11}$$

Proof. Let u and v be any two unit vectors. For convenience, write $M(s)$ for $M(\sigma(s), A)$ and $\lambda_{\min}(s)$ for $\lambda_{\min}(\mu(\sigma(s), A))$. Then,

$$\begin{aligned}
|\langle Sv, u\rangle| &= \left|\int_0^L \langle M(s)v, u\rangle f(\sigma(s))ds\right| \\
&\leq \int_0^L |\langle M(s)v, u\rangle| ds \quad (\text{since } \|f\|_\Omega = 1) \\
&\leq \int_0^L \langle M(s)u, u\rangle^{1/2} \cdot \langle M(s)v, v\rangle^{1/2} ds \quad (\text{Cauchy-Schwarz, since } M(s) \text{ is PSD}) \\
&\leq \left(\int_0^L \langle M(s)u, u\rangle ds\right)^{1/2} \left(\int_0^L M(s)v, v\rangle ds\right)^{1/2} \quad (\text{Bunyakovskii's inequality}) \\
&= \left\langle \left(\int_0^L M(s)ds\right) u, u \right\rangle^{1/2} \left\langle \left(\int_0^L M(s)ds\right) v, v \right\rangle^{1/2} \\
&= \left\langle \left(2 - \int_0^L \lambda_{\min}(s)ds\right) u, u \right\rangle^{1/2} \left\langle \left(2 - \int_0^L \lambda_{\min}(s)ds\right) v, v \right\rangle^{1/2} \quad \left(\int_0^L \mu(\sigma(s), A)ds = 2I\right) \\
&= 2 + \delta,
\end{aligned}$$

as desired. □

Remark 1: Note that δ can be positive or negative (but necessarily cannot be less than -2). It is 0 if $\Omega = W(A)$, positive if Ω is a subset of $\text{int}(W(A))$, and negative if Ω is convex and $W(A)$ is a subset of the interior of Ω. This follows from the fact shown in [17] that if τ lies on the tangent line to $W(A)$ at a point $\sigma \in \partial W(A)$, the infimum of the spectrum of the Hermitian part of $(\sigma'(s)/(\pi i))R(\tau, A)$ is 0, while on the side of this line that does not contain $W(A)$ it is positive and on the side that does contain $W(A)$ it is negative.

Remark 2: The region Ω in Lemma 3.1.1 need not be simply connected. For example, it could consist of a union of smooth regions (e.g. disks), each of which encloses a part of the spectrum.

The remainder of [17] is aimed at relating $\|f(A) + g(A)^*\|$ to $\|f(A)\|$. We assume that Ω is a bounded *convex* domain with smooth boundary. It is shown in [17] that if $g(z)$ is defined in Ω by

$$g(z) = \frac{1}{2\pi i} \int_{\partial\Omega} \frac{\overline{f(\sigma)}}{\sigma - z} d\sigma, \tag{3.12}$$

then $g \in \mathcal{A}(\Omega)$ (when g is extended continuously to $\partial\Omega$), $g(A)$ satisfies (3.6), and

$$\|g\|_\Omega \leq \|f\|_\Omega. \tag{3.13}$$

Further, it is shown that $g(\partial\Omega) := \{g(\sigma) : \sigma \in \partial\Omega\}$ is contained in the convex hull of the complex conjugate of the set $f(\partial\Omega) := \{f(\sigma) : \sigma \in \partial\Omega\}$.

For any bounded set Ω containing the spectrum of A in its interior, there is a minimal constant $c_\Omega(A)$ (which we write as c_Ω for convenience) such that for all $f \in \mathcal{A}(\Omega)$,

$$\|f(A)\| \leq c_\Omega \|f\|_\Omega. \tag{3.14}$$

One such constant can be derived from the Cauchy integral formula:

$$\|f(A)\| \leq \frac{1}{2\pi} \left(\int_{\partial\Omega} \|(\sigma I - A)^{-1}\| |d\sigma| \right) \|f\|_\Omega, \tag{3.15}$$

but this is usually not optimal. The following theorem uses Lemma 3.1.1 and inequality (3.13) to obtain a new upper bound on c_Ω.

Theorem 3.1.2. *Let Ω be a convex domain with smooth boundary containing the spectrum of A in its interior. Then*

$$c_\Omega \leq 1 + \frac{\delta}{2} + \sqrt{2 + \delta + \delta^2/4 + \hat{\gamma}}, \tag{3.16}$$

where

$$\delta = -\int_0^L \lambda_{\min}(\mu(\sigma(s), A))ds, \qquad \hat{\gamma} = \int_0^L |\lambda_{\min}(\mu(\sigma(s), A))|ds. \tag{3.17}$$

Proof. Let $f \in \mathcal{A}(\Omega)$ satisfy $\|f\|_\Omega = 1$. From (3.10), we can write

$$f(A)^* = S^* - (g(A) + \overline{\gamma}I).$$

Multiply by $f(A)^*f(A)$ on the left and by $f(A)$ on the right to obtain

$$[f(A)^*f(A)]^2 = f(A)^*f(A)S^*f(A) - f(A)^*f(A)(g(A) + \overline{\gamma}I)f(A).$$

Now take norms on each side and use the fact the the norm of any function of A is less than or equal to c_Ω times the supremum of that function on Ω to find

$$\begin{aligned} \|f(A)\|^4 &\leq c_\Omega^3\|S^*\| + c_\Omega\|h(A)\|, \qquad h(z) := f(z)(g(z) + \overline{\gamma})f(z), \\ &\leq c_\Omega^3(2 + \delta) + c_\Omega^2(1 + \hat{\gamma}), \qquad (\text{since } \|S^*\| \leq 2 + \delta \text{ and } \|h\|_\Omega \leq 1 + \hat{\gamma}). \end{aligned}$$

Since this holds for all $f \in \mathcal{A}(\Omega)$ with $\|f\|_\Omega = 1$, it follows that

$$c_\Omega^4 \leq c_\Omega^3(2 + \delta) + c_\Omega^2(1 + \hat{\gamma}),$$

and solving the quadratic inequality $c_\Omega^2 - (2 + \delta)c_\Omega - (1 + \hat{\gamma}) \leq 0$ for c_Ω gives the desired result. □

In all of our numerical experiments (to be discussed in detail in Chapter 4), it has always been the case that $\|f(A)\| \leq \|f(A) + g(A)^*\|$, when f is the function with $\|f\|_{W(A)} = 1$ that maximizes $\|p(A)\|$ over all $p \in \mathcal{A}(W(A))$ with $\|p\|_{W(A)} = 1$. If this could be proved, then it would establish Crouzeix's conjecture from the results in [17]. While we do not know of a proof of this in general, we can establish this for some special cases of matrices and provide simple proofs of some old results discussed in Chapter 2.

3.2 Optimal Blaschke products

As before, assume that A is a general n by n matrix. If Ω is any simply connected proper open subset of $\mathbb{C}$ containing the spectrum of A, then there is a function f such that $\|f\|_\Omega = 1$ and $\|f(A)\| = c_\Omega$. This function f is known to be of the form $B \circ \varphi$ [13, 21, 23] where φ is any bijective conformal mapping from Ω to the unit disk $\mathbb{D}$, and B is a Blaschke product of degree at most $n-1$. Recall that a Blaschke product of degree at most $n-1$ has the form

$$B(z) = e^{i\theta} \prod_{j=1}^{n-1} \frac{z - \alpha_j}{1 - \overline{\alpha}_j z}, \quad |\alpha_j| \leq 1, \tag{3.18}$$

and maps the unit disk to itself. We have allowed $|\alpha_j| = 1$ in this definition so that the degree of $B(z)$ can be less than $n-1$, since factors with $|\alpha_j| = 1$ are just unit scalars.

Now, suppose we have a matrix Ψ whose spectrum lies inside the unit disk $\mathbb{D}$, and assume that Ψ is diagonalizable [1]. If p is a function that is analytic in $\mathbb{D}$, and if $\Psi = V\Lambda V^{-1}$ is an eigendecomposition of Ψ, then $p(\Psi) = Vp(\Lambda)V^{-1}$. It is shown in [23] that of all functions f analytic in $\mathbb{D}$ and satisfying $f(\Lambda) = p(\Lambda)$ (and hence $f(\Psi) = p(\Psi)$), the unique one with smallest $\mathcal{H}^\infty$ norm on $\mathbb{D}$ is a scalar multiple of a Blaschke product of degree at most $n-1$. This means that for any function p analytic in $\mathbb{D}$, if p is not a scalar multiple of such a Blaschke product, then the ratio $\|p(\Psi)\|/\|p\|_\mathbb{D}$ can be increased by replacing p with μB (or just B), where B is a Blaschke product of degree at most $n-1$ and $\mu B(\Lambda) = p(\Lambda)$. Thus we can write

$$\sup_{p \in \mathcal{H}^\infty} \frac{\|p(\Psi)\|}{\|p\|_\mathbb{D}} = \sup_{B \text{ of the form (3.18)}} \|B(\Psi)\|. \tag{3.19}$$

In general, we do not know of an analytic formula for the roots $\alpha_j, j = 1, \ldots, n-1$, of this optimal Blaschke product. However, the following property of the optimal B was observed by us numerically, and a proof was provided by Crouzeix:

Theorem 3.2.1. *Let Ψ be an n by n matrix whose spectrum is inside the unit disk $\mathbb{D}$ and let B be a Blaschke product of degree at most $n-1$ that maximizes $\|\hat{B}(\Psi)\|$ over all Blaschke*

[1] The following results extend, by taking appropriate limits, to the case where Ψ is not diagonalizable.

products $\hat{B}$ of degree at most $n-1$. Assume that $\|B(\Psi)\| > 1$. If v_1 is a right singular vector of $B(\Psi)$ corresponding to the largest singular value σ_1, then

$$\langle B(\Psi)v_1, v_1\rangle = 0. \tag{3.20}$$

Proof. Let $M = B(\Psi)$, where B is the Blaschke product of the form (3.18) for which $\|B(\Psi)\|$ is maximal. No matrix of the form

$$(M - \alpha I)(I - \overline{\alpha}M)^{-1}, \quad |\alpha| < 1,$$

can have larger norm than M since this is also a Blaschke product in Ψ. Let v_1 be a unit right singular vector of M corresponding to the largest singular value σ_1, and define $w = (I - \overline{\alpha}M)v_1$. Then

$$\|(M - \alpha I)v_1\| = \|(M - \alpha I)(I - \overline{\alpha}M)^{-1}w\| \leq \sigma_1\|(I - \overline{\alpha}M)v_1\|.$$

Squaring both sides, this becomes

$$\langle Mv_1, Mv_1\rangle - 2\mathrm{Re}(\overline{\alpha}\langle Mv_1, v_1\rangle) + |\alpha|^2 \leq \sigma_1^2[1 - 2\mathrm{Re}(\overline{\alpha}\langle Mv_1, v_1\rangle) + |\alpha|^2\langle Mv_1, Mv_1\rangle],$$

and since $\langle Mv_1, Mv_1\rangle = \sigma_1^2$,

$$2(\sigma_1^2 - 1)\mathrm{Re}(\overline{\alpha}\langle Mv_1, v_1\rangle) \leq |\alpha|^2(\sigma_1^4 - 1).$$

Choosing α so that $\overline{\alpha}\langle Mv_1, v_1\rangle = |\alpha||\langle Mv_1, v_1\rangle|$, we have

$$2|\alpha||\langle Mv_1, v_1\rangle| \leq |\alpha|^2(\sigma_1^2 + 1),$$

and letting $|\alpha| \to 0$, this implies that $|\langle Mv_1, v_1\rangle| = 0$. □

We use this result in tandem with the Crouzeix-Palencia result and its extension in the following section, where we prove some old results using these new techniques, as well as provide new bounds for regions which do not necessarily contain $W(A)$. However, it is likely that there is much more to say about the optimal Blaschke product, both in general and for special cases. An interesting open problem is to determine other properties of the optimal B.

3.3 Bounds for special cases

In this section, we use some of the tools discussed in previous sections to find bounds for various special classes of matrices and candidate K-spectral sets. First, we use the Crouzeix-Palencia result in tandem with the optimal Blaschke product condition to provide simpler proofs of some known results for special classes of matrices and sets Ω. We derive some bounds using sets which do not necessarily contain $W(A)$ by utilizing our extension of the Crouzeix-Palencia result.

3.3.1 Ω is a disk

First, we consider the case where Ω is a closed disk, which may or may not contain $W(A)$.

Theorem 3.3.1. *If the spectrum of A is contained in a closed disk Ω, then Ω is a $\max\{1, 2+\delta\}$-spectral set for A, where δ is defined in (3.17).*

Proof. Suppose Ω is a closed disk with center c. We have (see pg. 205 of [38], for example)

$$g(z) = \frac{1}{2\pi i}\int_{\partial\Omega} \frac{\overline{f(\sigma)}}{\sigma - z} d\sigma = \overline{f(c)}, \quad z \in \Omega.$$

Hence, $g(A) = \overline{f(c)}I$. Now, suppose that $f = B \circ \varphi$ is a function that maximzes $\|f(A)\|$ over all functions with $\|f\|_\Omega = 1$. Furthermore, assume $\|f(A)\| > 1$ and let v_1 denote the unit right singular vector of $f(A)$ corresponding to the largest singular value σ_1 and let $u_1 = f(A)v_1/\|f(A)v_1\|$ denote the corresponding unit left singular vector. Then,

$$u_1^*[f(A) + g(A)^* + \gamma I]v_1 = u_1^* f(A)v_1 = \|f(A)\|,$$

since $u_1^* v_1 = 0$ by Theorem 3.2.1. It follows that $\|f(A)\| \leq \max\{1, \|S\|\}$ in 3.10, as desired. □

In particular, this provides a new proof of the following statement due to Okuno and Ando [35]:

If $W(A)$ is contained inside a closed disk Ω, then Ω is a 2-spectral set for A. (3.21)

This follows from the above, because in this case $\delta \leq 0$. Furthermore, if $W(A)$ is a proper subset of Ω, then $c_\Omega < 2$, and even if Ω does not contain all of $W(A)$, the estimate $\|f(A)\| \leq \max\{1, 2+\delta\} \cdot \|f\|_\Omega$ still holds.

3.3.2 A case in which the bound is sharp

From the previous section, we know that if Ω is a disk containing the spectrum of A in its interior, then Ω is a $\max\{1, 2+\delta\}$-set, where δ is defined as in (3.17). We show that this bound is sharp when A is a 3 by 3 Jordan block and Ω is a disk of radius less than or equal to 1 centered at the eigenvalue of A, which we take to be 0 for convenience.

Theorem 3.3.2. *Let A be a 3 by 3 Jordan block with eigenvalue 0 and let Ω be any disk about the origin with radius $r \leq 1$. Then*

$$\max_{f \in \mathcal{A}(\Omega)} \frac{\|f(A)\|}{\|f\|_\Omega} = 2 + \delta, \tag{3.22}$$

where δ is defined in (3.17).

Proof. The function f that achieves the maximum in (3.22) is of the form $B \circ \varphi$, where B is a Blaschke product of degree at most 2 and $\varphi(z) = z/r$ maps Ω to the unit disk $\mathbb{D}$. Since $\varphi(A)$ is a scalar multiple of a Jordan block, it is easy to see that the optimal B is $B(z) = z^2$ for $r \leq 1$. Hence, the left-hand side of (3.22) is $1/r^2$.

Now we evaluate the right-hand side of (3.22). Since $\sigma(s) = re^{is/r}$ on $\partial\Omega$, we can write

$$\frac{\sigma'(s)}{2\pi i} R(\sigma(s), A) = \frac{e^{is/r}}{2\pi}(re^{is/r} I - A)^{-1} = \frac{1}{2\pi r} \begin{pmatrix} 1 & \frac{e^{-is/r}}{r} & \frac{e^{-2is/r}}{r^2} \\ 0 & 1 & \frac{e^{-is/r}}{r} \\ 0 & 0 & 1 \end{pmatrix},$$

and

$$\mu(\sigma(s), A) = \frac{1}{2\pi r} \begin{pmatrix} 2 & \frac{e^{-is/r}}{r} & \frac{e^{-2is/r}}{r^2} \\ \frac{e^{is/r}}{r} & 2 & \frac{e^{-is/r}}{r} \\ \frac{e^{2is/r}}{r^2} & \frac{e^{is/r}}{r} & 2 \end{pmatrix}.$$

For $r \leq 1$, the smallest eigenvalue of this matrix is

$$\lambda_{\min}(\mu(\sigma(s), A)) = \frac{2r^2 - 1}{2\pi r^3},$$

independent of s. Routine integration yields

$$\delta = -\int_{\partial\Omega} \frac{1 - 2r^2}{2\pi r^3} ds = 2\pi r \left(\frac{1 - 2r^2}{2\pi r^3} \right) = \frac{1}{r^2} - 2.$$

Thus, $2 + \delta = r^2$, and the bound is achieved. □

3.3.3 A different bound on $c_{W(A)}$

Since a disk containing the spectrum of a matrix Ψ is a $\max\{1, 2+\delta\}$-spectral set for Ψ, we can use this to obtain a different bound on $c_{W(A)}$. Let φ be a bijective conformal mapping from $W(A)$ to the unit disk $\mathbb{D}$. Then $\mathbb{D}$ is a K-spectral set for $\varphi(A)$, where $K = \max\{1, 2+\delta_{\varphi(A)}\}$, and

$$\delta_{\varphi(A)} = -\int_0^{2\pi} \lambda_{\min}(\mu(\sigma(s), \varphi(A)))ds, \tag{3.23}$$

where $\sigma(s) = e^{is}$. It follows that $W(A)$ is a K-spectral set for A, with the same value of K, since for any $f \in \mathcal{A}(W(A))$, we have

$$\|f(A)\| = \|f \circ \varphi^{-1}(\varphi(A))\| \leq K\|f \circ \varphi^{-1}\|_{\mathbb{D}} = K\|f\|_{W(A)}.$$

Thus far, bounds of this type have only been useful for determining numerically better bounds on $c_{W(A)}$ (or more generally, c_Ω with the spectrum of A contained in Ω) for some test problems, to be discussed in Chapter 4. It is an open question whether such bounds can be determined theoretically. See [16] for some early work in this direction.

3.3.4 Conditions for other special cases

We have seen that we can use the techniques of the preceding sections to show that Crouzeix's conjecture holds for matrices whose numerical range is a circular disk. We know that the conjecture has been proven to hold for other classes of matrices listed at the beginning of

Chapter 2, and that in all of these cases $\varphi(A)$ is a linear function of A: $\varphi(A) = \alpha A + \beta I$, where φ is the conformal mapping taking $W(A)$ onto $\mathbb{D}$. From our numerical experiments, it appears that in all of these cases the function g corresponding to the optimal f has the form

$$g(A)^* = c_0 I + c_1 (f(A)^*)^{-1}. \tag{3.24}$$

Assuming $\|f(A)\| > 1$, we have

$$|u_1^*[f(A) + g(A)^* + \gamma I]v_1| = \left| \|f(A)\| + \frac{c_1}{\|f(A)\|} \right|. \tag{3.25}$$

If $\mathrm{Re}(c_1) \geq 0$, or more generally, $\mathrm{Re}(c_1) \geq |c_1|^2/(2\|f(A)\|^2)$, then this is greater than or equal to $\|f(A)\|$, whence $\|f(A)\| \leq \|S\|$.

Since the optimal f has magnitude 1 on $\partial\Omega$, $\overline{f(\sigma)}$ in (3.12) can be replaced by $1/f(\sigma)$ and formula (3.12) can be expanded using the residue theorem:

$$g(z) = \frac{1}{2\pi i} \int_{\partial\Omega} \frac{1}{(\sigma - z) f(\sigma)} d\sigma = \frac{1}{f(z)} + \sum_j \mathrm{Res}\left(\frac{1}{(\sigma - z) f(\sigma)}, \sigma = \beta_j \right),$$

where each β_j is a distinct root of f. For example, suppose the optimal Blaschke product B has degree 1, which is always true for 2 by 2 matrices and holds for some other special cases. Then, the formula for $g(z)$ becomes

$$g(z) = \frac{1}{f(z)} - \frac{1}{(z - \beta) f'(\beta)},$$

where $\beta \in W(A)$ is the point that is mapped to 0 by f. If it turns out that $f(A) = c(A - \beta I)$ for some constant c, then

$$g(A) = \left(1 - \frac{c}{f'(\beta)}\right) f(A)^{-1}, \tag{3.26}$$

and Crouzeix's conjecture can be proven to hold for this matrix A if it can be shown that the constant in parentheses has positive real part.

3.3.5 A is a 2 by 2 matrix

We know that Crouzeix's conjecture holds for 2 by 2 matrices [12], but we provide a new proof by showing here that if A is diagonalizable, then $g(A)$ is a positive multiple of $f(A)^{-1}$

and the result follows from [17] and (3.24 - 3.25). If the 2 by 2 matrix A is not diagonalizable, then it is unitarily similar to a scalar multiple of a Jordan block, and since the numerical range is a disk in this case, it follows from the previous argument that $g(A)$ is a multiple of the identity.

In order to make use of the results of the previous section, we need the form of the optimal Blaschke product. For the particular case of 2 by 2 matrices, we can use Theorem 3.2.1 from the previous section in order to provide new proofs of the following results, which can be found in slightly different forms in [12].

Lemma 3.3.3. *Let A be a 2 by 2 matrix whose spectrum lies inside the unit disk, and let $M = B(A)$, where B is the Blaschke product of degree 1 for which $\|B(A)\|$ is maximal. Then, a necessary condition for $\|B(A)\|$ to be maximal is that $tr(B(A)) = 0$.*

Proof. Assume that $\|B(A)\| > 1$. Write a singular value decomposition of M as $M = U\Sigma V^*$, where $U = (u_1, u_2)$ and $V = (v_1, v_2)$. Since $Mv_1 = u_1\sigma_1$ is orthogonal to v_1, the left singular vector u_1 satisfies $u_1 = v_2 e^{i\theta_1}$ for some $\theta_1 \in [0, 2\pi)$. Similarly, since u_2 is orthogonal to u_1, we have $u_2 = v_1 e^{i\theta_2}$ for some $\theta_2 \in [0, 2\pi)$. Thus we can write

$$M(v_1, v_2) = (v_2, v_1) \begin{pmatrix} e^{i\theta_1} & 0 \\ 0 & e^{i\theta_2} \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} = (v_1, v_2) \begin{pmatrix} 0 & \sigma_1 e^{i\theta_1} \\ \sigma_2 e^{i\theta_2} & 0 \end{pmatrix}.$$

It follows that M is unitarily similar to the matrix

$$\begin{pmatrix} 0 & \sigma_1 e^{i\theta_1} \\ \sigma_2 e^{i\theta_2} & 0 \end{pmatrix},$$

whose eigenvalues are $\pm\sqrt{\sigma_1\sigma_2}e^{i(\theta_1+\theta_2)/2}$. Thus, a necessary condition for $\|B(A)\|$ to be maximal is that $\mathrm{tr}(B(A)) = 0$. □

Lemma 3.3.4. *Let A be a 2×2 matrix with eigenvalues λ_1, λ_2 inside the unit disk. Let*

$$B(z) = \frac{z - \alpha}{1 - \overline{\alpha}z}, \quad |\alpha| < 1, \tag{3.27}$$

and assume that $\|B(A)\| > 1$ *for some* α*. A necessary condition for* α *to maximize* $\|B(A)\|$ *is*

$$\frac{\lambda_1 - \alpha}{1 - \overline{\alpha}\lambda_1} = -\frac{\lambda_2 - \alpha}{1 - \overline{\alpha}\lambda_2}. \tag{3.28}$$

If $\lambda_1 = -\lambda_2$*, then* $\alpha = 0$ *is the unique maximizer of* $\|B(A)\|$*. Otherwise, a maximizer* α *of* $\|B(A)\|$ *must satisfy* $|\alpha| < 1$ *and*

$$1 + |\alpha|^2 = \frac{2}{\mathrm{tr}(A)}(\alpha + \overline{\alpha}\det(A)). \tag{3.29}$$

Proof. Condition 3.28 follows from the fact that the eigenvalues of $B(A)$ are $(\lambda_k - \alpha)/(1 - \overline{\alpha}\lambda_k), k = 1, 2$. After some algebra, this becomes

$$\lambda_1 + \lambda_2 - 2\alpha - 2\overline{\lambda_1\lambda_2} + |\alpha|^2(\lambda_1 + \lambda_2) = 0, \tag{3.30}$$

and if $\lambda_2 = -\lambda_1$, this becomes

$$\alpha = \overline{\alpha}\lambda_1^2. \tag{3.31}$$

If $\alpha \neq 0$, this would imply $|\lambda_1| = 1$, contradicting the assumption that $|\lambda_1| < 1$. Hence, $\alpha = 0$ must be the unique maximizer of $\|B(A)\|$. Otherwise, if $\lambda_1 \neq -\lambda_2$, then dividing by $\lambda_1 + \lambda_2$ yields

$$1 + |\alpha|^2 = \frac{2}{\lambda_1 + \lambda_2}(\alpha + \overline{\alpha}\lambda_1\lambda_2), \tag{3.32}$$

which is equivalent to 3.29. □

With these lemmas in hand, we return to showing Crouzeix's conjecture holds for 2 by 2 matrices. Due to the invariance of the relation (3.14) under unitary similarity transformations and scalings and translations of A, we can assume that A is of the form

$$\begin{pmatrix} 1 & a \\ 0 & -1 \end{pmatrix}, \quad a > 0.$$

Then $W(A)$ is an elliptical disk with foci ± 1, major axis of length $\sqrt{4+a^2}$ on the real axis, and minor axis of length a on the imaginary axis. The conformal mapping from $W(A)$ onto $\mathbb{D}$ can be written as (see p. 373-374 of [28], for example)

$$\varphi(z) = \frac{2z}{\rho} \exp\left\{ \sum_{n=1}^{\infty} \frac{2(-1)^n T_{2n}(z)}{n(1+\rho^{4n})} \right\},$$

where T_{2n} is the $2n$th Chebyshev polynomial of the first kind and $\rho = (a + \sqrt{a^2+4})/2$.

Note that φ maps the eigenvalues ± 1 to $\pm\varphi(1) \in \mathbb{D}$. Hence, from Theorem 3.3.4, we know that the root of the optimal Blaschke product is 0 for $\Psi = \phi(A)$, since the eigenvalues of Ψ are equal in magnitude and of opposite sign. This implies that if f maximizes $\|f(A)\|$ with $\|f(A)\|_{W(A)} = 1$, then f is of the form $f(z) = (B \circ \varphi)(z) = \varphi(z)$, and we also have $f(A) = \varphi(1)A$. It follows from (3.26) with $\beta = 0$ that

$$g(A) = \left(1 - \frac{\varphi(1)}{\varphi'(0)}\right) f(A)^{-1}.$$

Note that

$$\varphi(1) = \frac{2}{\rho} \exp\left\{ \sum_{n=1}^{\infty} \frac{2(-1)^n}{n(1+\rho^{4n}} \right\}, \quad \varphi'(0) = \frac{2}{\rho} \exp\left\{ \sum_{n=1}^{\infty} \frac{2}{n(1+\rho^{4n})} \right\},$$

from which it is clear that the coefficient of $f(A)^{-1}$ is real and positive. If $f(A) = U\Sigma V^*$ is an SVD of $f(A)$, then $g(A) = \eta V \Sigma^{-1} U^*$, where $\eta = 1 - \varphi(1)/\varphi'(0) \geq 0$, and $g(A)^* = \eta U \Sigma^{-1} V^*$. Thus,

$$u_1^*[f(A) + g(A)^*]v_1 = \|f(A)\| + \frac{\eta}{\|f(A)\|} \geq \|f(A)\|,$$

implying that $\|f(A)\| \leq \|f(A) + g(A)^*\| \leq 2$, as expected.

Chapter 4

NUMERICAL EXPERIMENTS

The previous chapters cover many of the recent advancements in the study of K-spectral sets, in particular the numerical range. However, there remain many difficult open questions. It is generally difficult to analytically determine a value of K for which a given set is K-spectral, so numerical computations can be useful for studying these sets. We can numerically compute the function f of the form (3.1) that attains $\sup_{\hat{f}\in\mathcal{A}(\Omega)} \|\hat{f}(A)\|/\|\hat{f}\|_\Omega$ by first determining a conformal mapping from Ω to the unit disk $\mathbb{D}$ numerically, and then determining the roots of the optimal Blaschke product via an optimization procedure.

In this chapter, we detail some of the numerical experiments done to study K-spectral sets. We begin by detailing the procedure for constructing numerical conformal maps from a given region Ω to the unit disk. While we are interested primarily in studying norms of functions of matrices, some of the techniques described here are currently being used in more general contexts and are, in fact, incorporated into the software package Chebfun [20]. We focus on the case of a domain Ω with a smooth boundary, and use the Kerzmann-Stein integral equation [31, 32] to compute the conformal mapping. Smooth boundaries can be represented very accurately and concisely with Chebyshev or trigonometric series, so we make use of the open-source software package Chebfun [20] for our implementation.

We show how one can attempt to compute the optimal f of the form $B \circ \varphi$ for a given A and region Ω containing the spectrum of A by numerically determining the roots of the optimal Blaschke product (3.2) in order to maximize $\|f(A)\| = \|B(\varphi(A))\|$. We have no guarantee that we will find the optimal roots, but this allows us to give an approximation of c_Ω in (3.14). We believe this is close to the true value of c_Ω, and it is at least a lower bound up to roundoff error. For diagonalizable matrices $A = V\Lambda V^{-1}$, we can evaluate $f(A)$

by using the value of f at the eigenvalues and setting $f(A) = Vf(\Lambda)V^{-1}$. In cases where A is not diagonalizable or the eigenvectors are extremely ill-conditioned, the Cauchy integral formula can be used to evaluate $f(A)$.

The remainder of the chapter is devoted to numerical experiments which make use of these techniques. In particular, we compute some numerical bounds using δ and $\lambda_{\min}$ in (3.11) for various test problems, and compare with other numerical bounds found via optimization methods or computing Cauchy integrals as in (3.15). We also do some numerical studies of $\lambda_{\min}(\mu(\sigma, A))$ to try to gain insight on its behavior inside or outside of $W(A)$, as well as detail some experiments with computing matrix dilations. If A is a linear operator on a Hilbert space H (e.g. a square matrix in $\mathbb{C}^{n\times n}$), then a dilation of A is a linear operator M on a larger space $K \supset H$ such that $A = P_H M|_H$, where P_H is the orthogonal projection onto H. These dilations are potentially useful objects in the study of $\|f(A)\|$, since they may have nice properties that A does not have. Even if A is highly nonnormal, one can construct well-behaved (near normal) dilations that could be used instead to bound $\|f(A)\|$. We show how to construct these numerically, and we study the behavior of functions of these dilations and how it compares to that of the original operator.

4.1 Numerical conformal mapping

A standard method of constructing conformal maps is to use Schwarz-Christoffel mappings, which provide transformations of the upper half-plane or the unit disk onto the interior of a simple polygon. By approximating the numerical range $W(A)$ with a polygon and using a numerical package such as the SC toolbox [19], developed by Driscoll and Trefethen, one can obtain an approximation to the desired mapping $\varphi(A)$. This method works best when the sets to be mapped are polygonal, such as the numerical range of a diagonal matrix. However, in the context of nonnormal matrices, we are mostly interested in sets with smooth boundaries. The boundary of the numerical range $\partial W(A)$ may be some combination of smooth parts and straight line segments, and for most generic matrices the boundary is analytic. Other means of constructing these conformal maps based on integral equations

take advantage of this smoothness and yield a more efficient numerical procedure.

In Crouzeix's numerical experiments for 3×3 matrices [15], he employs a modification of Symm's method [41], which involves solving an integral equation of the first kind to obtain the boundary correspondence function between $\partial W(A)$ and the unit circle. This is done by writing $\varphi(z) = z \exp(u + iv)$, where $u(z)$ and $v(z)$ are harmonic real-valued functions and noting that $u(z) = -\log|z|$ on $\partial W(A)$. Parameterizing the boundary as $\sigma(\theta)$ with $\theta \in [0, 2\pi]$ and noting that there exists a real-valued density function q such that

$$(u + iv)(z) = \int_0^{2\pi} q(\theta) \log(\sigma(\theta) - z) d\theta, \quad \int_0^{2\pi} q(\theta) d\theta = -1, \tag{4.1}$$

Crouzeix solves the boundary equation $u = -\log|z|$ by approximating $q(\theta)$ by a degree n trigonometric polynomial and using the trapezoidal rule with $2n + 1$ collocation points. As long as the boundary is analytic, this method is very efficient, with super-algebraic convergence due to the trapezoidal rule, though there are some difficulties dealing with non-generic geometries with corners and cusps.

4.1.1 Conformal mapping in Chebfun

For our numerical experiments dealing with nonnormal matrices, we use the software package Chebfun [20] in order to conformally map domains with smooth boundaries to the unit disk. Such domains can be represented very accurately and concisely with Chebyshev or trigonometric series, as is done in Chebfun, and that representation of boundary curves can very easily be differentiated, integrated, or otherwise manipulated. This means that the boundary correspondence function, defining the image of points on the boundary of such a domain under the conformal map, can be computed to an accuracy near the limits of machine precision and the conditioning of the problem. This boundary correspondence function can again be represented very accurately as a Chebyshev or trigonometric series and used with the Cauchy integral formula to find the images of interior points in the domain. We use the same technique to compute the image of a square matrix A whose eigenvalues lie inside the region under the mapping. The inverse boundary correspondence function can be

determined using bisection or other methods and once this is computed it too can be used with the Cauchy integral formula to compute inverse images of interior points or of matrices.

For comparison, we consider using a simple polygonal approximation to the domain and the SC package [19] for conformally mapping the polygon. Although this package is accurate and easy to use, it cannot, of course, take advantage of the smooth boundary, and so cannot achieve the level of accuracy that we achieve in Chebfun.

To do the conformal mapping of such a domain Ω to the unit disk $\mathbb{D}$, we use the Kerzman-Stein integral equation [31, 32]. This method is based upon the relationship between the Riemann mapping function of a smooth, bounded, simply connected domain Ω and the Szegö kernel $\mathcal{S}$ of Ω. The Szegö kernel satisfies the relation

$$\varphi'(z) = \frac{2\pi}{\mathcal{S}(a,a)}\mathcal{S}^2(z,a), \qquad z \in \Omega,$$

where $\varphi : \Omega \to \mathbb{D}$ is the desired conformal mapping sending a to 0. One can find $\varphi(z)$ without any integration via

$$\varphi(z) = \frac{T(z)}{i}\frac{\varphi'(z)}{|\varphi'(z)|}, \qquad z \in \partial\Omega,$$

where $T(z)$ denotes the unit tangent to $\partial\Omega$ at z. In order to compute $\mathcal{S}$, we use the method in [32], which computes $\mathcal{S}$ as the solution of an integral equation of the second kind,

$$f(z) + \int_{w\in\partial\Omega} A(z,w)f(w)d\sigma = h(z), \qquad z \in \partial\Omega, \tag{4.2}$$

where $f(z) \equiv \mathcal{S}(z,a)$, $d\sigma$ denotes an element of arc length on $\partial\Omega$, the right-hand side is

$$h(z) = \frac{1}{2\pi i}\frac{\overline{T(z)}}{a-z}, \qquad z \in \partial\Omega,$$

and the kernel function is

$$A(w,z) = \left[\frac{1}{2\pi i}\frac{T'(z)}{w-z}\right]^* - \frac{1}{2\pi i}\frac{T'(z)}{z-w}, \qquad w \neq z \text{ (0 if } w = z).$$

This has the advantage of a smooth, skew-Hermitian kernel, which gives rise to a well-conditioned linear system of equations once discretized. We parameterize the boundary in

terms of arclength with a Chebyshev interpolant, and employ equally spaced collocation points so that the trapezoidal rule yields geometric convergence. After solving for $\mathcal{S}$, we use its relationship with the Riemann mapping function to compute the boundary correspondence between $\partial\Omega$ and the unit circle $\mathbb{T}$. Finally, we compute the map at interior points by way of the Cauchy integral formula.

Now we demonstrate the level of accuracy we can achieve with this approach. Let $\Omega = E_\rho$ denote the interior of the Bernstein ellipse with foci at ± 1 and with semiminor and semimajor axis lengths summing to $\rho > 1$. The formula for the conformal map from E_ρ to $\mathbb{D}$ is explicitly known in terms of Jacobi elliptic functions, so we can compare our numerical results with that of the SC package against the known mapping. For this numerical experiment, we calculate the boundary map using the Kerzman-Stein equation with n points, as well as the SC map using the same n points for the approximating polygon. We map an interior ellipse $E_\rho/2$ inside the unit disk and evaluate the sup-norm of the difference between each approximation φ_n and the known solution φ. In Tables 4.1 and 4.2, we show the sup-norms $\|\varphi(E_\rho/2) - \varphi_n(E_\rho/2)\|_\infty$ for varying number of points n and different values for ρ. Note that $\rho = 1.7$ roughly corresponds to an axis ratio of $2:1$, while $\rho = 1.2$ corresponds to a ratio of about $5:1$, so significant crowding begins to occur for such an elongated ellipse. Despite this, the Chebfun implementation of the Kerman-Stein method gives very accurate results for the interior mapping for all cases, while the SC package struggles mapping elongated ellipses with significant crowding, and generally cannot match the same level of accuracy. For example, when $\rho = 1.2$, the SC Toolbox gives a warning signifying instability due to crowding, and the error actually begins to increase with at least $n = 128$ points. We see this in Tables 4.1 and 4.2, with the first showing errors using the Kerzman-Stein method implemented in Chebfun, and the second showing errors using the SC package.

The boundary mapping is very fast to compute even with a large number of sample points, so it is easy to obtain a highly accurate boundary correspondence function. As can be seen in the following tables, the interior map retains the accuracy of the boundary mapping, even with the additional Cauchy integral. This interior mapping is also accurate very close to the

Table 4.1: Error norms $\|\varphi(E_\rho/2) - \varphi_n(E_\rho/2)\|_\infty$ for mapping n points in the interior of a Bernstein ellipse E_ρ using the Kerzman-Stein integral equation. As the number of discretization points n increases, the error rapidly decreases, eventually reaching the limits of machine precision and the conditioning of the problem. The problem is more ill-conditioned as ρ gets smaller due to crowding.

n	$\rho = 1.7$	$\rho = 1.4$	$\rho = 1.2$
16	$1.3 \cdot 10^{-4}$	$2.1 \cdot 10^{-3}$	$2.9 \cdot 10^{-2}$
32	$3.0 \cdot 10^{-8}$	$7.6 \cdot 10^{-6}$	$8.6 \cdot 10^{-4}$
64	$5.0 \cdot 10^{-13}$	$1.0 \cdot 10^{-8}$	$5.6 \cdot 10^{-6}$
128	$6.2 \cdot 10^{-15}$	$1.3 \cdot 10^{-12}$	$1.6 \cdot 10^{-7}$
256	$3.6 \cdot 10^{-15}$	$8.0 \cdot 10^{-15}$	$3.4 \cdot 10^{-10}$

Table 4.2: Error norms $\|\varphi(E_\rho/2) - \varphi_n(E_\rho/2)\|_\infty$ for mapping n points in the interior of a Bernstein ellipse E_ρ using the SC package. In this case, the error begins to stagnate as n gets larger and ρ gets smaller, as this method suffers worse from crowding instabilities.

n	$\rho = 1.7$	$\rho = 1.4$	$\rho = 1.2$
16	$5.4 \cdot 10^{-3}$	$3.8 \cdot 10^{-3}$	$3.5 \cdot 10^{-3}$
32	$1.3 \cdot 10^{-3}$	$9.5 \cdot 10^{-4}$	$8.1 \cdot 10^{-4}$
64	$3.2 \cdot 10^{-4}$	$2.4 \cdot 10^{-4}$	$2.0 \cdot 10^{-4}$
128	$8.0 \cdot 10^{-5}$	$5.9 \cdot 10^{-5}$	$6.7 \cdot 10^{-3}$
256	$2.0 \cdot 10^{-5}$	$1.5 \cdot 10^{-5}$	$3.1 \cdot 10^{-1}$

boundary, even though the Cauchy integrals become nearly singular. Additionally, we can write the inverse interior map from $\mathbb{D}$ to Ω in terms of the boundary correspondence function, so we can obtain similar levels of accuracy for the inverse map. If one desires the inverse boundary correspondence function, then a simple bisection or Newton iteration scheme can be used to find inverse images pointwise, which can be fit with a trigonometric interpolant. The following plots show some examples in which disks and rays in the unit disk (upper-left corner) are conformally mapped to various smooth domains:

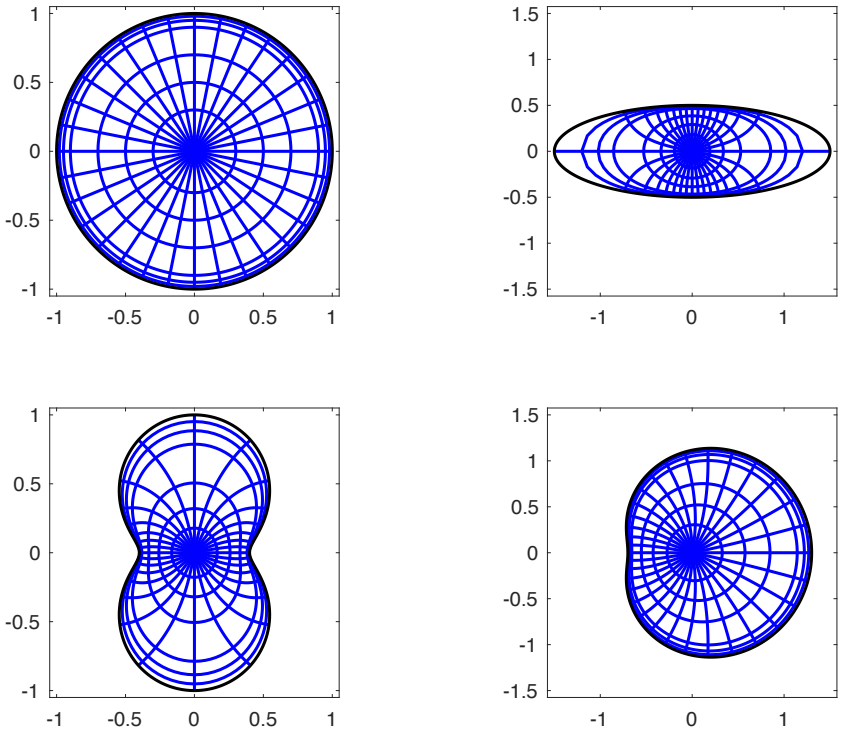

Figure 4.1: Mapping disks and rays in the unit disk to various smooth domains

4.1.2 Some improvements using rational approximants

Recently, drastic speedups were achieved by using rational functions to approximate the conformal map. The AAA ("aggressive Antoulas-Anderson") algorithm [34] implemented in Matlab makes use of the following key ideas for approximation by rational functions on a set in the complex plane:

1. Representation of the rational approximant in barycentric form with interpolation at certain support points selected from a set provided by the user.

2. Grow the approximation degree one by one, selecting support points in a systematic greedy fashion to avoid exponential instabilities.

3. Solve linear algebra problems at each step using linearized least-squares fitting.

This can be used in tandem with our Chebfun implementation of the Kerzman-Stein integral equation for conformal mapping. One method is to use the Kerzman-Stein integral equation to solve for the images along the unit circle of points along the boundary of the original domain Ω, and use the AAA algorithm to compute a rational function interpolating this data, i.e. an approximation to the conformal map φ mapping Ω to $\mathbb{D}$ [24]. For more accuracy in the approximation, one can compute mappings of some interior points using Cauchy integrals and the boundary correspondence function, and supply this as extra data for interpolation.

This is most useful when an approximation to the conformal map φ is desired at many points, because the rational function supplied by the AAA algorithm is lightning fast to evaluate. There are some stability issues due the appearance of spurious poles at some steps of the algorithm, so one cannot generally achieve the same level accuracy as careful Cauchy integrals for mapping individual points. However, one can still generally achieve a relative accuracy near 10^{-8} for smooth domains that don't have long and thin regions, and can map hundreds of thousands of points in the time frame on the order of a second. For example, the plot below shows the numerical range of a random 4 by 4 matrix and 10,000 interior points mapped to the unit disk in under a tenth of a second on a dual-core i5 laptop:

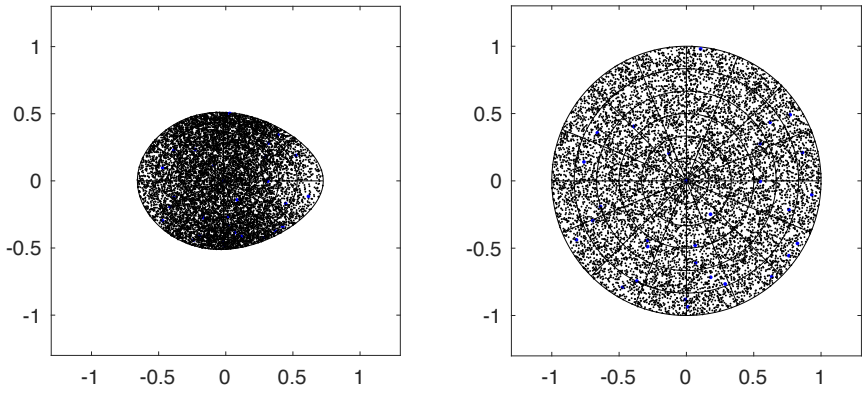

Figure 4.2: Mapping many points to the unit disk using the AAA algorithm for rational approximation

4.2 Computing bounds on $\|f(A)\|$ numerically

Using the methods described in Chapter 3 and the previous section, we show how to compute bounds on norms of functions of matrices numerically. For these experiments, we will take σ to lie on a circle enclosing the spectrum of a given matrix A but not necessarily enclosing the numerical range $W(A)$. More precisely, we first determine the center c and radius r of the smallest circle enclosing the spectrum of A, and then consider circles about c with radius $R > r$, i.e.

$$\sigma(s) = c + Re^{is/R}, \qquad \sigma'(s) = ie^{is/R},$$

where s parameterizes the circle $\partial\Omega$ by arclength from 0 to $2\pi R$. We compute $\lambda_{\min}(\mu(\sigma, A))$ at points σ on these circles, and use this to bound c_Ω in (3.14) for each disk Ω. For a first example, let us consider a 3×3 perturbed Jordan block:

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0.1 & 0 & 0 \end{pmatrix}. \tag{4.3}$$

In Figure 4.3, we plot the eigenvalues and numerical range of A, as well as the disks on which we compute $\lambda_{\min}$.

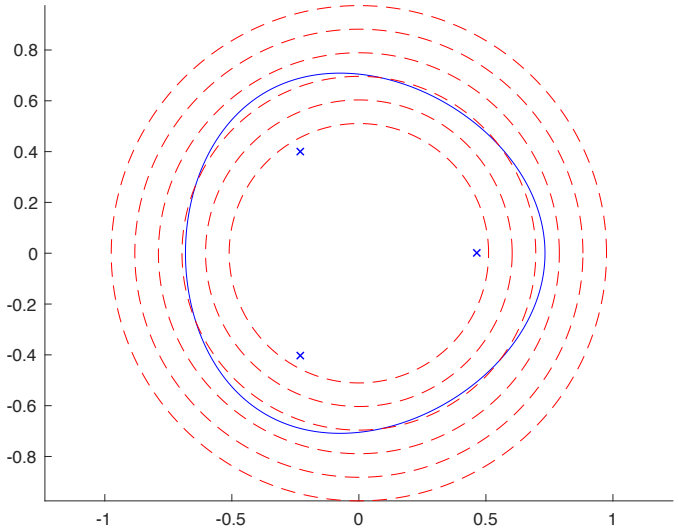

Figure 4.3: Eigenvalues (blue x's) and numerical range (solid blue curve) of the 3 by 3 perturbed Jordan block (4.3). The red dashed circles are the ones on which $\lambda_{\min}(\mu(\sigma, A))$ was computed.

We provide a plot of $\lambda_{\min}(s)$ for each of the circles as functions of the arclength s along each curve in Figure 4.4. In this plot, the bottom curve corresponds to the innermost circle, and the curves move up as the circles become larger. We see from the figure that $\lambda_{\min}$ decreases rapidly as σ moves inside $W(A)$ towards the spectrum, but it grows very slowly as σ moves outside $W(A)$.

Table 4.3 shows the values of δ and $\hat{\gamma}$ in (3.17) and the upper bound bound on c_Ω in (3.16) (labeled K_δ) for each each of the disks, starting with the smallest disk about the spectrum. Because the regions we are considering are disks, we can actually employ the improved bound $c_\Omega \leq 2 + \delta$ in Theorem 3.3.1. We compare these bounds on c_Ω with the upper bound found using the Cauchy integral formula and the resolvent norm in (3.15) (labeled K_{Cauchy}), and with the numerical computation of $\|f(A)\| = \|B \circ \varphi\| = c_\Omega$. The latter computation is done by first determining a conformal map φ from Ω to the unit disk $\mathbb{D}$, and then using an optimization routine to determine the roots of the Blaschke product B which maximize

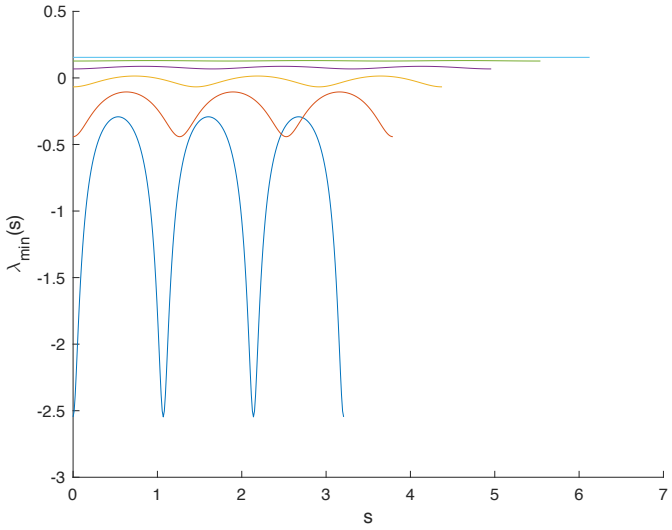

Figure 4.4: Plot of $\lambda_{\min}(\mu(\sigma(s), A))$ vs. arclength s on each of the dashed circles in Figure 4.3. The bottom curve corresponds to the innermost circle surrounding the spectrum, and the curves move up as the circles become larger.

$\|f(A)\| = \|B \circ \varphi\|$. We believe that this is, within numerical error, the true value of c_Ω, but it at least provides a lower bound for c_Ω. Notice that in all cases, $2 + \delta < K_\delta < K_{\text{Cauchy}}$, and that in some cases, the improved disk bound $c_\Omega \leq 2 + \delta$ is very close to the largest value returned by our optimization code for $\|B \circ \varphi(A)\|$. Recall from (3.22) that this bound is actually made sharp for the 3×3 Jordan block with no perturbation.

Let us now consider a random complex upper triangular matrix A of dimension $n = 12$ and run through the same computations. In Figure 4.5, we again plot the eigenvalues and numerical range of A, together with disks on which we compute $\lambda_{\min}$. We plot $\lambda_{\min}$ as a function of arclength in 4.6, and compute the same numerical bounds as before in Table 4.4. Even for this random matrix, we get similar results: $2 + \delta < K_\delta < K_{\text{Cauchy}}$ in all cases.

As mentioned in Chapter 3, we can use the quantity $\delta_{\varphi(A)}$ in (3.23) in order to obtain numerically better bounds on $c_{W(A)}$ for the test problems considered in this section. For the 3×3 perturbed Jordan block, $\delta_{\varphi(A)} = -0.0013$, implying that $W(A)$ is a 1.9987-spectral set

δ	$\hat{\gamma}$	K_δ	$2+\delta$	K_{Cauchy}	$\|B(\varphi(A))\|$
2.6706	2.6706	5.3559	4.6706	6.1273	3.8360
0.8854	0.8854	3.4344	2.8854	4.1915	2.7465
0.0943	0.1221	2.5367	2.0943	3.3667	2.0629
-0.3869	0.3869	2.2339	1.6131	2.8760	1.6061
-0.7131	0.7131	2.1019	1.2869	2.5472	1.2858
-0.9475	0.9475	2.0177	1.0525	2.3118	1.0525

Table 4.3: Values of δ and γ in (3.17) and upper bounds (K_δ and $2+\delta$) on c_Ω in (3.16), upper bound K_{Cauchy} on c_Ω in (3.15), and lower bound $\|B(\varphi(A))\|$ found by numerical optimization of B for disks in Figure 4.3.

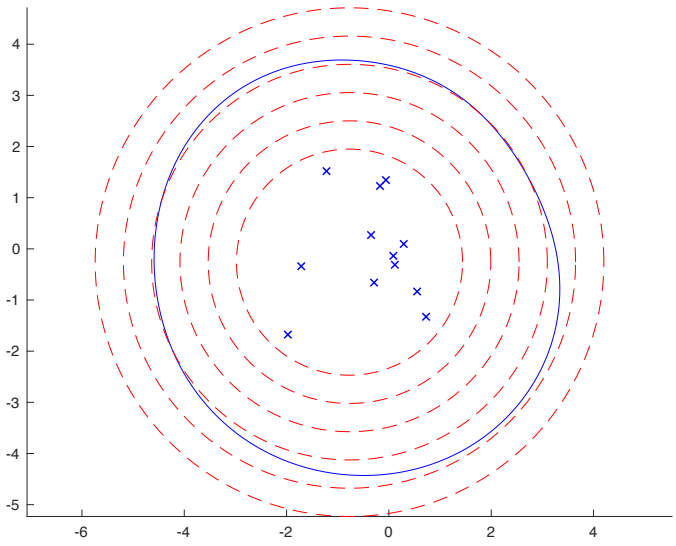

Figure 4.5: Eigenvalues (blue x's) and numerical range (solid blue curve) of a random complex upper triangular matrix A of dimension $n = 12$. The red dashed circles are the ones on which $\lambda_{\min}(\mu(\sigma, A))$ was computed.

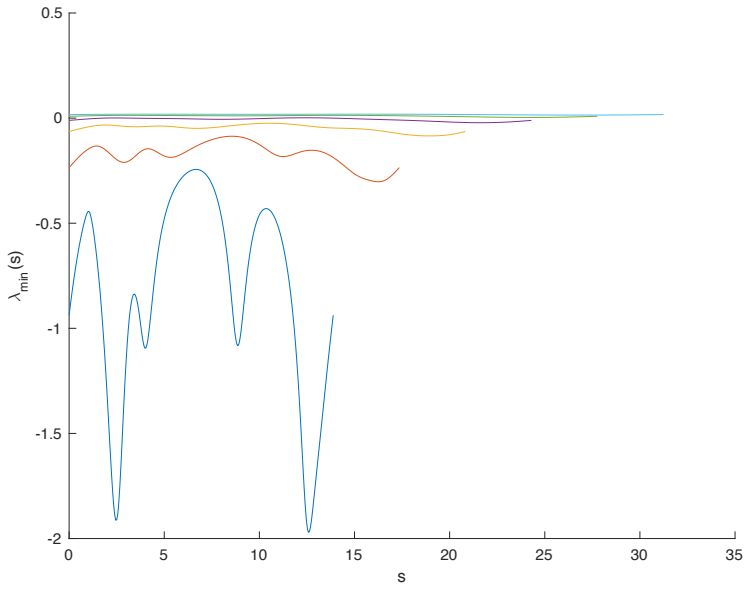

Figure 4.6: Plot of $\lambda_{\min}(\mu(\sigma(s), A))$ vs. arclength s on each of the dashed circles in Figure 4.5. The bottom curve corresponds to the innermost circle surrounding the spectrum, and the curves move up as the circles become larger.

δ	$\hat{\gamma}$	K_δ	$2+\delta$	K_{Cauchy}	$\|B(\varphi(A))\|$
11.3176	11.3176	14.1859	13.3176	16.8946	5.9760
3.0361	3.0361	5.7393	5.0361	7.4288	2.8609
1.0012	1.0012	3.5629	3.0012	4.8246	2.1673
0.1629	0.1699	2.6110	2.1629	3.6602	1.8147
-0.2735	0.2735	2.2841	1.7265	3.0160	1.5789
-0.5389	0.5389	2.1702	1.4611	2.6127	1.4010

Table 4.4: Values of δ and γ in (3.17) and upper bounds (K_δ and $2+\delta$) on c_Ω in (3.16), upper bound K_{Cauchy} on c_Ω in (3.15), and lower bound $\|B(\varphi(A))\|$ found by numerical optimization of B for disks in Figure 4.5.

for A, since $\max\{1, 2+\delta_{\varphi(A)}\} = 1.9987$. Similarly, for the 12 by 12 random upper triangular matrix, $\delta_{\varphi(A)} = 0.0112$, implying that $W(A)$ is a 2.0112-spectral set for A. This method provides better numerical upper bounds than previously known bounds, and it is an open question whether such bounds can be determined theoretically. However, the numerical result for the random upper triangular matrix suggests that this will not be a way to prove Crouzeix's conjecture, since the bound is larger than the conjectured value of 2.

4.3 Experiments involving matrix dilations

In this section, we describe work in our paper [25], which details our numerical studies of matrix dilations and their applications to the study of nonnormal matrices. We begin by recalling the definition of a matrix dilation and suggest why studying them may be useful for bounding norms of functions of nonnormal matrices. Let A be a square matrix or a linear operator on a Hilbert space H. A *dilation* of A is a linear operator Z on a larger space $\mathcal{K} \supset H$ such that $A = P_H Z|_H$, where P_H is an orthogonal projection onto H. In the case where A is an n by n matrix, we can identify $H = \mathbb{C}^n$ with the subset of $\mathcal{K} = \mathbb{C}^N, N > n$, consisting of vectors whose last $N - n$ components are 0. Then, any N by N matrix of the form

$$Z = \begin{pmatrix} A & * \\ * & * \end{pmatrix},$$

is a dilation of A. If we also require that Z^m be a dilation of A^m for all positive integers m, then Z is said to be a *power dilation* of A.

Dilations of an operator A may have nice properties that A does not have, which can prove useful for bounding norms of functions of nonnormal matrices. We know from the previous chapters that if A is highly nonnormal, then the spectrum $\Lambda(A)$ may give little useful information regarding the norms of functions of A. However, if Z is a (near) normal power dilation of A, then $\|p(A)\| \leq \|p(Z)\|$ for any polynomial p since $p(A)$ is a block of $p(Z)$, and $\|p(Z)\|$ is (approximately) determined by $\Lambda(Z)$. More concretely, if Z is similar to a normal operator N via a similarity transformation S with moderate condition number

$\kappa(S)$, then $\|p(A)\| \leq \|p(Z)\| \leq \kappa(S) \sup_{z \in \Lambda(Z)} |p(z)|$.

4.3.1 Constructing finite unitary dilations

In our approach, we construct near normal dilations of nonnormal matrices, with the spectrum of the dilated operator Z around the boundary of the numerical range of the matrix A that it dilates. To do this, we make use of a theorem by Sz.-Nagy [42], which states that every contraction C ($\|C\| \leq 1$) has a unitary power dilation U. In matrix form, this dilation is a doubly infinite block triangular matrix known as the Schäffer matrix [39]:

$$U = \begin{pmatrix} \ddots & \ddots & & & & & \\ & 0 & I & & & & \\ & & 0 & D_C & -C^* & & \\ & & & C & D_{C^*} & & \\ & & & & 0 & I & \\ & & & & & 0 & \ddots \\ & & & & & & \ddots \end{pmatrix}, \quad D_C = (I - C^*C)^{1/2}. \tag{4.4}$$

Unfortunately, this is necessarily an infinite dilation and if it is truncated to obtain a finite matrix, then the truncated matrix is no longer unitary. However, one can construct finite unitary dilations with a similar structure as follows (see, for example, p. 59 of [30]). Begin with any unitary dilation of C, e.g.,

$$\begin{pmatrix} C & D_{C^*} \\ D_C & -C^* \end{pmatrix}. \tag{4.5}$$

Form a larger matrix U_k by adding identity operators in the following fashion:

$$U_k = \begin{pmatrix} C & D_{C^*} & & & \\ & & I & & \\ & & & \ddots & \\ & & & & I \\ D_C & -C^* & & & \end{pmatrix}. \tag{4.6}$$

This finite matrix is unitary, and U_k^m is a dilation of C^m for $m = 1, \ldots, k-1$. If an operator A is similar to a contraction C via a well-conditioned similarity transformation, i.e. $A = XCX^{-1}$ with $\kappa(X) \leq Q$, then the unitary dilations (4.4) and (4.6) become near normal dilations of A once appropriate similarity transformations are applied. Define $S =$ block diag$(\ldots, I, X, I, \ldots)$ and $S_k =$ block diag$(X, I, \ldots, I)$. Then SUS^{-1} and $S_k U_k S_k^{-1}$ are dilations of A. If X is normalized so that $\|X\| = 1$, then $\kappa(S) = \kappa(S_k) = \kappa(X)$.

Recall the definition (2.13) of a complete K-spectral set. It was shown by Paulsen [37] that if the unit disk $\mathbb{D}$ is a complete K-spectral set for an operator A, then A is similar to a contraction via a similarity transformation with condition number $K : A = XCX^{-1}, \kappa(X) = K$. Now, assume that the interior of the numerical range $W(A)$ is nonempty and not all of $\mathbb{C}$. Let φ be a bijective conformal mapping from the interior of $W(A)$ to the unit disk $\mathbb{D}$, extended continuously to the boundary. It follows from the Crouzeix-Palencia result (2.17) that the unit disk $\mathbb{D}$ is a complete K-spectral set for $\varphi(A)$ for some $K \leq 1 + \sqrt{2}$. Hence, from Paulsen's theorem we have $\varphi(A) = XCX^{-1}$, where C is a contraction and $\kappa(X) = K$.

Sz.-Nagy's theorem implies that C has a unitary power dilation U, and $\varphi^{-1}(U)$ is a normal power dilation of $\varphi^{-1}(C)$ with spectrum on $\partial W(A)$. Let $S =$ block diag$(\ldots, I, X, I, \ldots)$. Then $S\varphi^{-1}(U)S^{-1}$ is a power dilation of A with spectrum on $\partial W(A)$ that is similar to the normal operator $\varphi^{-1}(U)$ via a similarity transformation S with conditon number K. In other words, every matrix A whose numerical range has nonempty interior has a near normal power dilation with spectrum on $\partial W(A)$.

We are interested in the construction of finite dilations using finite matrices. If C is a finite matrix and one considers a dilation U_k of the form (4.6), then one can construct an mth degree polynomial approximation f_m to φ^{-1} that satisfies $f_m(C) = \varphi^{-1}(C)$. Assuming $k \geq m + 1$, then $f_m(U_k)$ will be a normal dilation of $\varphi^{-1}(C)$ with spectrum approximately on $\partial W(A)$ and satisfying that $[f_m(U_k)]^\ell$ is a dilation of $[\varphi^{-1}(C)]^\ell$ for $\ell = 1, \ldots, k-m$. One can take the degree of f_m large enough to approximate φ^{-1} on $\overline{\mathbb{D}}$ to any desired level of accuracy and make powers $[f_m(U_k)]^\ell$ be dilations of $[\varphi^{-1}(C)]^\ell$ for any finite number of powers ℓ. Finally, defining $S_k =$ block diag$(X, I, \ldots, I)$, we have that $S_k f_m(U_k) S_k^{-1}$ is a dilation of

A with spectrum approximately on $\partial W(A)$ that is similar to the normal operator $f_m(U_k)$ via a similarity transformation S_k whose condition number is K.

4.3.2 Numerical construction of near normal dilations

Using the ideas in the previous sections, we can numerically construct a finite near normal dilation Z of a given n by n matrix A with spectrum approximately on $\partial W(A)$ as follows:

1. Compute the numerical range $W(A)$. We do this using the `fov` command in Chebfun [20], which uses a standard algorithm [30] to compute points along the boundary of $W(A)$ and fits a Chebyshev series to a level of accuracy near machine precision.

2. Compute a bijective conformal mapping from $W(A)$ to the unit disk $\mathbb{D}$. We do this by using the Chebfun implementation of the Kerzman-Stein integral equation described in Section 4.1.1.

3. Evaluate $\varphi(A)$, where φ is the conformal map from $W(A)$ to $\mathbb{D}$. For diagonalizable matrices, this can be done by evaluating φ at the eigenvalues of A, and for the general case, this can be done by discretizing the Cauchy integral formula using the trapezoidal rule.

4. Find a similarity transformation X such that $C = X^{-1}\varphi(A)X$ is a contraction and the condition number of X is minimized. We do this by attempting to solve constrained optimization problems using `fmincon` in MATLAB.

5. Form a finite unitary dilation of the form (4.6), adjusting the number of blocks k as necessary to ensure that we can approximate functions of interest using polynomials of degree at most $k-1$.

6. Evaluate $\varphi^{-1}(U_k)$ and a polynomial approximation $f_m(U_k)$, such that the degree $m \ll k-1$, $f_m(C) = \varphi^{-1}(C)$, and $f_m(U_k) \approx \varphi^{-1}(U_k)$. This is often a computational stum-

bling block, as the degree of polynomial needed to approximate the inverse map φ^{-1} may be quite large, especially when trying to map long thin regions to the unit disk.

7. After computing $f_m(U_k)$, multiply its first block row on the left by X and its first block column on the right by X^{-1} to obtain the dilation Z of A.

We have carried out this procedure successfully only for very small matrices A, as it quickly becomes computationally difficult as the dimension increases. As an alternative, we can instead require Z to have its spectrum on a circle enclosing $W(A)$ rather than on $\partial W(A)$, which is a substantially simpler computational problem. We first compute $W(A)$ and a circle enclosing it, and map this circle to the unit disk by shifting and scaling. If the circle about $W(A)$ has center c and radius r, then $\varphi(z) = (z-c)/r$. The numerical range of $\varphi(A)$ is $W(\varphi(A)) = (W(A)-c)/r$ and is a subset of $\mathbb{D}$. We can use a fixed point iteration due to Choi and Greenbaum [11] and based on work by Okubo and Ando [1, 35] in order to find a matrix X with $\kappa(X) \leq 2$ such that $C = X^{-1}\varphi(A)X$ is a contraction. Because the numerical radius of $\varphi(A)$ is less than or equal to 1, it is guaranteed to converge. Additionally, after constructing C and forming the unitary dilation U_k as in (4.6), we have no issues with evaluating $\varphi^{-1}(U_k)$ since $\varphi^{-1}(z) = rz + c$ is a first degree polynomial. We form the matrix $Z = S_k\varphi^{-1}(U_k)S_k^{-1}$, where $S_k = \text{block diag}(X, I, \ldots, I)$. Z is a dilation of A with spectrum on the disk enclosing $W(A)$ and with eigenvectors having condition number at most 2, and its powers Z^m will be dilations of the corresponding powers A^m for $m = 1, \ldots, k-1$.

4.3.3 Examples

To illustrate this approach, consider a matrix of the form

$$A = \begin{pmatrix} \lambda & a \\ 0 & \lambda \end{pmatrix}, \qquad a > 0.$$

The numerical range of an n by n Jordan block with eigenvalue λ is a disk about λ of radius $r = \cos(\pi/(n+1))$, so in this case $W(A)$ is a disk about λ of radius $r = a/2$. Applying the

conformal map $\varphi(Z) = (z - \lambda)/r$ gives

$$\varphi(A) = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix},$$

and applying the algorithm given in [11] yields a matrix X such that $C = X^{-1}\varphi(A)X$ is a contraction, with

$$X = \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix}, \qquad C = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

We construct a unitary dilation U of the form (4.6) using this contraction C, and map back to $W(A)$ with $\varphi^{-1}(U) = \lambda I + \frac{a}{2}U$. Finally, apply the similarity transformation to $\varphi^{-1}(U)$ by left-multiplying the first two rows by X and right-multiplying the first two columns by X^{-1} to obtain the following matrix Z:

$$Z = \begin{pmatrix} \lambda & a & 0 & 0 & & & & \\ 0 & \lambda & 0 & a/4 & & & & \\ & & \lambda & 0 & a/2 & 0 & & \\ & & 0 & \lambda & 0 & a/2 & & \\ & & & & \ddots & & \ddots & \\ & & & & & \ddots & & \ddots \\ a/2 & 0 & 0 & 0 & & & \lambda & 0 \\ 0 & 0 & -a/2 & 0 & & & 0 & \lambda \end{pmatrix}. \tag{4.7}$$

The matrix Z is a near-normal dilation of A with spectrum on $\partial W(A)$.

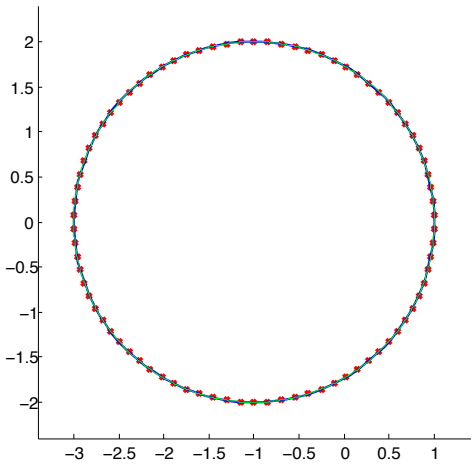

Figure 4.7: $W(A)$ together with eigenvalues of Z for $\lambda = -1, a = 4$.

We would also like to understand the behavior of these dilations in relation to the original operators being dilated. As a starting point, we investigate evolution processes, i.e. solutions to systems of differential equations. For example, consider the system of differential equations $y' = Ay$ with

$$A = \begin{pmatrix} -1 & 4 \\ 0 & -1 \end{pmatrix}, \quad y(0) = y_0,$$

which has the solution $y(t) = e^{tA}y_0$. We would like to compare $\|e^{tA}\|$ with $\|e^{tZ}\|$.

Suppose that $\hat{y}' = Z\hat{y}$. Then, $\hat{u}(t) = e^{tZ}\hat{y}(0)$ and $\|e^{tZ}\| \leq 2e^{t\omega(A)} = 2e^t$, since Z has an eigenvector matrix with condition number 2 and the spectral abscissa of Z is at most $\alpha(A)$, which is the numerical abscissa of A. Since e^{tZ} is a dilation of e^{tA}, it follows that $\|e^{tA}\| \leq \|e^{tZ}\| \leq 2e^t$. This is slightly weaker than the known bound $\|e^{tA}\| \leq e^{t\omega(A)}$ (see pg. 138 of [43]), but this same bounding technique can be used for other functions as well.

We show the behavior of $y(t) = e^{tA}y(0)$ and $\hat{y}(t) = e^{tZ}\hat{y}(0)$ in the plots below, where $y(0) = [u_0, v_0]^T$ is a random initial vector and $\hat{y}(0) = [u_0, v_0, 0, \ldots, 0]^T$:

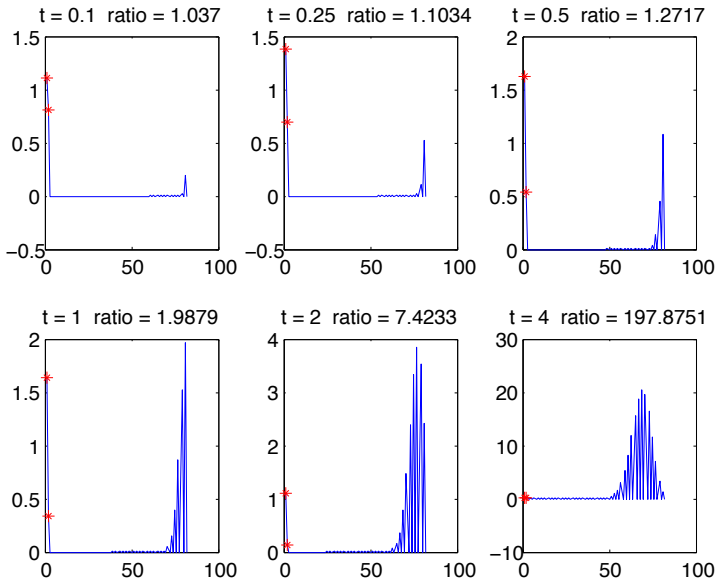

Figure 4.8: Behavior of $y(t)$ (asterisks) and $\hat{y}(t)$ (solid lines).

The asterisks are the components of $y(t)$, and the solid line goes through the components of $\hat{y}(t)$. The horizontal axis indicates the index of the components - there are 80 total components in this numerical experiment since we used 38 additional identity blocks in the construction of the finite dilation, yielding an 80 by 80 near normal dilation Z. We can see that for early time ($t = 0.1, 0.25$), the first two components dominate the behavior, as a small disturbance begins to appear in the rightmost components. Later at times $t = 0.5$ and $t = 1$, this disturbance has grown to be about the same size as the first two components corresponding to the original matrix A. Later at $t = 2$, the first two components begin to decay, while the rightmost components continue to grow. Finally, at $t = 4$, the rightmost components introduced by the dilation have completely dwarfed the decaying components on the left, and have aggregated to form a leftward advecting wave, completely dwarfing the decaying components on the left corresponding to the original operator. We can also see that the norm ratios $\|e^{tZ}\|/\|e^{tA}\|$ rapidly increase, so that by $t = 4$, it is clear that bounding $\|e^{tA}\|$ with $\|e^{tZ}\|$ would no longer be of any practical use.

We can understand the difference in the behavior of these operators in the following sense. The original equation $y' = Ay$ is a pair of coupled ordinary differential equations, and we may think of the dilation equation $\hat{y}' = Z\hat{y}$ as spatially differenced approximation to a pair of coupled partial differential equations. Define

$$\hat{y}(t) = \begin{pmatrix} u_1(t) \\ v_1(t) \\ \vdots \\ u_k(t) \\ v_k(t) \end{pmatrix},$$

where $u_j(t)$ and $v_j(t)$ are approximations to the multivariate functions $u(x,t)$ and $v(x,t)$ at $x = j$. The middle blocks of the dilated equation $\hat{y}' = Zy$ are

$$\begin{aligned} \frac{du_j}{dt} &= -u_j + 2u_{j+1} = u_j + 2(u_{j+1} - u_j) \approx u_j + 2\frac{\partial u}{\partial x}\Big|_{x=j}, \\ \frac{dv_j}{dt} &= -v_j + 2v_{j+1} = v_j + 2(v_{j+1} - v_j) \approx v_j + 2\frac{\partial v}{\partial x}\Big|_{x=j}. \end{aligned} \tag{4.8}$$

We can interpret this as a method of lines approach with forward differencing in the spatial domain for the advection equations $u_t = u + 2u_x$ and $v_t = v + 2v_x$, with boundary conditions coupling u and v at the ends of the domain. Recall that $w_t - cw_x = aw$ has the solution $w(t) = e^{at}w_0(x + ct)$, which is a wave traveling to the left with speed c and growing or decaying like e^{at}.

The equations for $u_j(t)$ and $v_j(t), j = 1, \ldots, k$, can actually be solved exactly, assuming that $u_2(0) = v_2(0) = \cdots = u_k(0) = v_k(0) = 0$. From (4.7), we have

$$v_k' = -v_k - 2u_2 \quad \Longrightarrow \quad v_k(t) = e^{-t}v_k(0) - 2\int_0^t e^{-(t-s)}u_2(s)ds,$$

which remains zero as long as $u_2(s) = 0$ for $s \le t$. We also have, for $j = k-1, \ldots, 2$,

$$v_j' = -v_j - 2v_{j+1} \quad \Longrightarrow \quad v_j(t) = e^{-t}v_j(0) - 2\int_0^t e^{-(t-s)}v_{j+1}(s)ds,$$

which is zero as long as $v_{j+1}(s) = 0$ for $s \le t$. This implies that $v_2, \ldots, v_k$ are all 0 until the wave on the right in Figure 4.8 reaches block 2. However, at this point e^{tZ} is no longer a

dilation of e^{tA}, so we need to include more blocks in the dilation Z in order to study times this large.

We can proceed similarly to evaluate each u_j. For u_k, we have

$$u_k' = -u_k + 2u_1 \quad \Longrightarrow \quad v_k(t) = e^{-t}u_k(0) + 2\int_0^t e^{-(t-s)}u_1(s)ds.$$

Note that $u_1(t)$ and $v_1(t)$ are the same as the solutions for the original problem, so that $u_1(t) = e^{-t}u_0 + 4te^{-t}v_0$ and $v_1(t) = e^{-t}v_0$. Thus, we can substitute the expression for $u_1(s)$ to obtain

$$u_k(t) = 2\int_0^t e^{-(t-s)}[e^{-s}u_0 + 4se^{-s}v_0]ds = 2te^{-t}u_0 + 4t^2e^{-t}v_0.$$

The equation for u_{k-1} is

$$u_{k-1}' = -u_{k-1} + 2u_k \quad \Longrightarrow \quad u_{k-1}(t) = 2\int_0^t e^{-(t-s)}u_k(s)ds,$$

and substituting the expression for $u_k(s)$ yields

$$u_{k-1}(t) = e^{-t}\left[2t^2u_0 + \frac{8}{3}t^3v_0\right].$$

Continuing in this fashion, we iteratively compute $u_{k-2}, u_{k-3}, \ldots$ to find that

$$u_{k-j}(t) = e^{-t}\left[\frac{2^{j+1}}{(j+1)!}t^{j+1}u_0 + \frac{2^{j+3}}{(j+2)!}t^{j+2}v_0\right], \quad j = 0, \ldots, k-2. \tag{4.9}$$

While we had originally hoped that the behavior of $\|e^{tZ}\|$ could be useful for understanding the behavior of $\|e^{tA}\|$ for at least small times t, even this simple example shows that the dilated operator takes on a life of its own. It is mainly the middle blocks of (4.7) that determine the behavior of $\|e^{tZ}\|$ and $e^{tZ}\hat{y}(0)$. Generating the near-normal dilation by way of finding a contraction C to which $\varphi(A)$ is 2-similar eventually leads to the equations in (4.8). These dictate the behavior of the dilated operator, which generates a wave that grows like e^t and moves left at speed 2, but these do not directly involve the matrix A at all and are instead determined by the mapping φ^{-1} from $\mathbb{D}$ to $W(A)$.

Chapter 5

APPLICATIONS AND OPEN PROBLEMS

In this chapter, we discuss some applications of the analytical and numerical methods used in Chapters 3 and 4 to bound norms of functions of matrices. We apply the methods of these chapters for specific classes of functions of matrices which are useful for applications, such as matrix powers and exponentials. We conclude by summarizing the research done in this dissertation and mentioning some interesting open problems and other potential application areas not discussed here.

5.1 Applications

Examples of some functions of matrices which may be seen in practice include the following:

- The solution of a linear system $(A - tI)x = b$ is given in terms of the resolvent: $x(t) = (A - tI)^{-1}b$.

- The solution of the dynamical system $u'(t) + Au(t) = 0$ is given in terms of the matrix exponential: $u(t) = \exp(-tA)u(0)$.

- The solution of the second order differential equation $u''(t) + Au(t) = 0$ is given in terms of trigonometric matrix functions and matrix roots: $u(t) = \cos(\sqrt{A}t)u(0) + (\sqrt{A})^{-1}\sin(\sqrt{A}t)u'(0)$.

- The solution of the difference equation $u_k = Au_{k-1}$ is given in terms of matrix powers: $u_k = A^k u_0$. The convergence behavior of iterative methods like Gauss-Seidel or SOR also involves norms of matrix powers.

- The convergence behavior of some Krylov subspace iterations, such as GMRES, depend on norms of polynomials of matrices.

- Decay bounds for singular values of solutions to certain Lyapunov equations with low-rank right hand sides depend on rational functions of matrices [5]

- The behavior of nonlinear systems of the form $u'(t) = Au(t) + f(u(t))$, whose solution is determined by the linear problem. Nonnormal effects may lead to nonlinear instabilities, even when the linear problem is asymptotically stable.

We will focus primarily on matrix powers A^k and matrix exponentials e^{tA}, for which interesting sets to investigate are disks and half-planes, respectively. In particular, we would like to determine when the unit disk $\mathbb{D}$ is a K-spectral set for a given matrix A with spectrum inside $\mathbb{D}$ and for some moderate value of K, even if the numerical range extends far outside $\mathbb{D}$. This would be helpful for bounding norms of matrix powers when A is nonnormal. Similarly, we would like to do the same for the left half-plane Π_0 and matrix exponentials e^{tA}.

5.1.1 Matrix powers

For powers of matrices A^k, we are most often interested in investigating the unit disk as a possible K-spectral set. Even if the numerical range $W(A)$ extends well past the unit disk, we can still determine upper bounds on $\|A^k\|$ using $\mathbb{D}$ via the methods of Chapter 3 and 4. As a simple first example, let us consider a matrix for which we can calculate some bounds explicitly. Let A be a scaled Jordan block of the form

$$A = \begin{pmatrix} \lambda & a \\ 0 & \lambda \end{pmatrix}, \qquad -1 < \lambda < 1, a > 0.$$

The spectrum of this matrix consists of a double eigenvalue $\lambda \in (-1, 1)$, but the numerical range is a disk centered at λ with radius $a/2$, so it may extend outside of the unit disk. We

can explicitly calculate the operator $\mu(\sigma(s), A)$ in (3.8) with this particular matrix A:

$$\begin{aligned}\mu(\sigma(s), A) &= \frac{\sigma'(s)}{2\pi i} R(\sigma(s), A) + \left[\frac{\sigma'(s)}{2\pi i} R(\sigma(s), A)\right]^* \\ &= \frac{1}{2\pi} \begin{pmatrix} \frac{2-2\lambda\cos(s)}{1+\lambda^2-2\lambda\cos(s)} & \frac{ae^{is}}{(1-\lambda e^{is})^2} \\ \frac{ae^{-is}}{(1-\lambda e^{-is})^2} & \frac{2-2\lambda\cos(s)}{1+\lambda^2-2\lambda\cos(s)} \end{pmatrix},\end{aligned}$$

which has minimum eigenvalue

$$\lambda_{\min}(s) = \frac{1}{2\pi}\left(\frac{2-a-2\lambda\cos(s)}{1+\lambda^2-2\lambda\cos(s)}\right).$$

Integrating, we find that

$$\delta = -\int_0^{2\pi} \lambda_{\min}(s) ds = -2 + \frac{a}{1-\lambda^2},$$

implying that the unit disk is a K-spectral set for A with $K \leq \max\{1, 2+\delta\} = \max\{1, a/(1-\lambda^2)\}$. In particular, when $a \leq 1-\lambda^2$, we have $K = 1$ and $\|A\| \leq 1$, and we recover von Neumann's inequality for the disk, which states that $\mathbb{D}$ is a spectral set if and only if $\|A\| \leq 1$. When $a > 1-\lambda^2$, we get the bound $K \leq a/(1-\lambda^2)$. For instance, with the matrix

$$A = \begin{pmatrix} 0.8 & 0.54 \\ 0 & 0.8 \end{pmatrix},$$

we have $K \leq 0.8/(1-(0.54)^2) = 1.5$, meaning that powers A^k can never grow to be larger than 1.5 in norm. We know that a disk about the numerical range is always a 2-spectral set for any matrix A. In this case, the numerical range of A extends slightly outside of $\mathbb{D}$, and yet $\mathbb{D}$ is actually at most a 1.5-spectral set for A. In fact, we can determine that $\mathbb{D}$ is exactly a 1.5-spectral set, since we can explicitly compute the optimal Blasckhe product to be $(A-0.8I)(I-0.8A)^{-1}$, which has norm 1.5.

Generally, the formulas for bounds found this way will not be as simple. Let us consider another 2×2 matrix

$$A = \begin{pmatrix} \lambda & a \\ 0 & -\lambda \end{pmatrix}, \quad 0 < \lambda < 1, a \in \mathbb{R}.$$

The numerical range $W(A)$ is now an ellipse centered at the origin with minor axis a and major axis $\sqrt{4\lambda^2 + a^2}$. Running through the same calculations, we find that

$$\begin{aligned} K &\leq 2 + \delta \\ &= \frac{2}{\pi(1-\lambda^2)\sqrt{a^2+4\lambda^2}} \left([a^2 - (1-\lambda^2)^2] E_K(n) + (1+\lambda^2)^2 E_\pi(n,m) \right), \end{aligned}$$

where $E_K(m)$ denotes the elliptic integral of the first kind:

$$E_K(m) = \int_0^{\pi/2} [1 - m\sin^2(\theta)]^{-1/2} d\theta,$$

and $E_\pi(n,m)$ denotes the elliptic integral of the third kind:

$$E_\pi(n,m) = \int_0^{\pi/2} [1 - n\sin^2(\theta)]^{-1} [1 - m\sin^2(\theta)]^{-1/2} d\theta,$$

with parameters

$$m = \frac{4\lambda^2[(1-\lambda^2)^2 - a^2]}{(1-\lambda^2)(a^2+4\lambda^2)}, \quad n = -\frac{4\lambda^2}{(1-\lambda^2)^2}.$$

As before, we can use this bound to find matrices for which the unit disk is K-spectral with $K \leq 2$, even if $W(A)$ is not contained in $\mathbb{D}$. For instance, if

$$A = \begin{pmatrix} 0.8 & 1.5 \\ 0 & -0.8 \end{pmatrix},$$

the unit disk is approximately a 1.9791-spectral set, even though the numerical radius $r(A) \approx 1.0966$ extends past the unit disk. However, we can see that these bounds quickly become difficult to calculate explicitly, as even the 2 by 2 case contains messy elliptic integrals. Hence, unless one can find effective lower bounds on $\lambda_{\min}$ for particular matrices, these techniques are best suited for numerical bounds.

To see how this sort of bound may be useful in practice, consider the steady-state advection-diffusion problem (see e.g. page 237 of [43])

$$-\nu u''(x) + \gamma u'(x) = f(x), \qquad x \in [0,1], u(0) = \alpha, u(1) = \beta,$$

for constant viscosity $\nu > 0$ and wind speed $\gamma > 0$. In order to approximate solutions to this numerically, we could discretize the interval $[0,1]$ with $n+2$ uniformly spaced grid points

and use centered finite differences to give rise to an n by n coefficient matrix. We can label points from left to right (downwind with $\gamma > 0$), or from right to left (upwind). Here, the terms downwind and upwind arise from the fact that the first ordering follows the direction of the wind γ and the second is against it. The coefficient matrix for the downwind scheme is

$$A = \text{tridiag}\left(-\nu - \frac{\gamma}{2n},\ 2\nu,\ \ -\nu + \frac{\gamma}{2n}\right), \tag{5.1}$$

while the coefficient matrix for the upwind scheme is the transpose of this matrix,

$$A = \text{tridiag}\left(-\nu + \frac{\gamma}{2n},\ 2\nu,\ \ -\nu - \frac{\gamma}{2n}\right). \tag{5.2}$$

Suppose that we use Gauss-Seidel iteration to compute the solution u. In this case, we split the matrix into its diagonal, strictly lower triangular, and strictly upper triangular parts, $A = D + L + U$, and use the iteration matrix $S = -(D+L)^{-1}U$. The upwind and downwind schemes lead to Gauss-Seidel iteration matrices with different nonnormal properties, which affect the rate at which these schemes converge. We can observe this effect by calculating bounds of the form $2 + \delta$ from Theorem 3.3.1 for several disks containing the spectrum.

In Figure 5.1, we see the spectrum and numerical range of the iteration matrix S corresponding to the downwind scheme (5.1). We notice that the numerical range is contained well within the unit disk, so we expect for $\|S^k\|$ to initially decay proportionally to $r(A)^k$, where $r(A) \approx 0.8216$ is the numerical radius. In fact, in Figure 5.2, we see that $\|S^k\|$ quickly approaches the asymptotic decay rate dictated by the spectral radius $\rho(S)^k \approx (0.4335)^k$, and is near machine epsilon using around 50 iterations (note the logarithmic scale on the y-axis). We also have bounds of the form $2 + \delta$ using the disks in Figure 5.1. Notice from Figure 5.2 that while the bounds corresponding to disks of larger radii (i.e. 0.9 and 1.0) are better initially, they are clearly gross overestimates as k gets larger. Because $\|S\| = 0.8578$, we know those disks cannot get better bounds than the simple estimate $\|S^k\| \leq \|S\|^k$. However, since the spectrum is well inside $\mathbb{D}$, we can get bounds using smaller disks that go inside of the numerical range of S. For these disks, we get bounds which are initially large overestimates, but then more closely match the true behavior of $\|S^k\|$ for intermediate and large values of

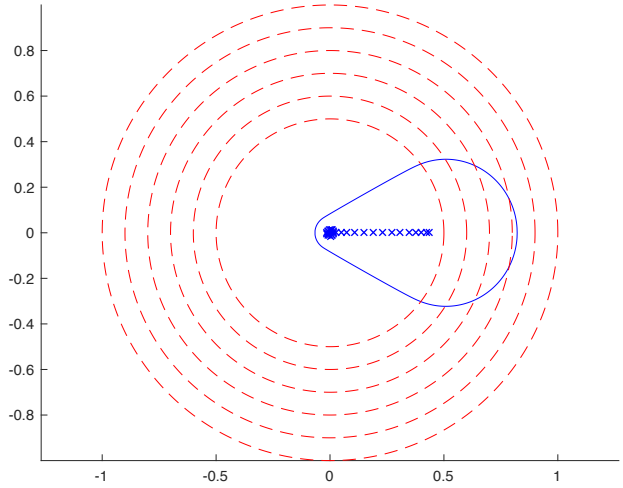

Figure 5.1: Eigenvalues (blue x's) and numerical range (solid blue curve) of the Gauss-Seidel iteration matrix corresponding to the downwind scheme (5.1). The red dashed circles are the ones on which $\lambda_{\min}(\mu(\sigma, A))$ was computed.

k. In a similar fashion to choosing ϵ for calculations involving the ϵ-pseudospectrum of S, we can choose the radius of the disk so that Ω is closer to the boundary of the spectrum, which gives better long-time estimates, or close to the numerical range or unit disk, which gives better short-time estimates.

However, we see that when we switch to the upwind scheme (5.2), the nonnormal effects are more noticeable. From looking at Figure 5.3, we should already be concerned with the enlarged numerical range - in this case, the numerical range is nearly at the boundary of the unit disk, even though the spectrum is the same as that of the downwind iteration matrix. Even though $W(S)$ does not extend outside of the unit disk, we see from Figure 5.4 that the iteration scheme takes much longer to begin approaching the asymptotic decay rate. We can see this by examining the constants $2+\delta$ corresponding to the bounds using disks of different radii. In this case, the unit disk is a spectral set, and the bound from this disk is very close to the simple bound $\|S^k\| \leq \|S\|^k$, since $\|S\| \approx 0.9992$. However, the disk of radius 0.5 is a $2.5604 \cdot 10^{11}$-spectral set, implying that many more iterations are necessary for the upper

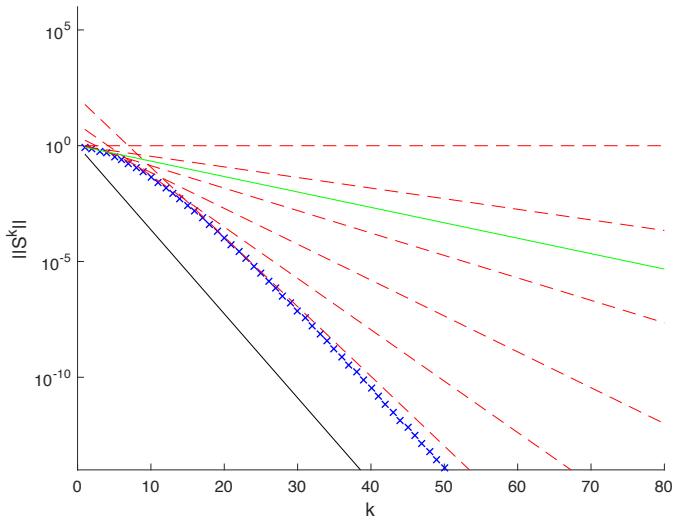

Figure 5.2: $\|S^k\|$ (blue x's) for the downwind iteration matrix S together with several upper and lower bounds. The black line is the lower bound $\rho(S)^k$, the red dashed lines are upper bounds of the form $2+\delta$ from the disks in Figure 5.1, and the green line is the upper bound $\|S\|^k$.

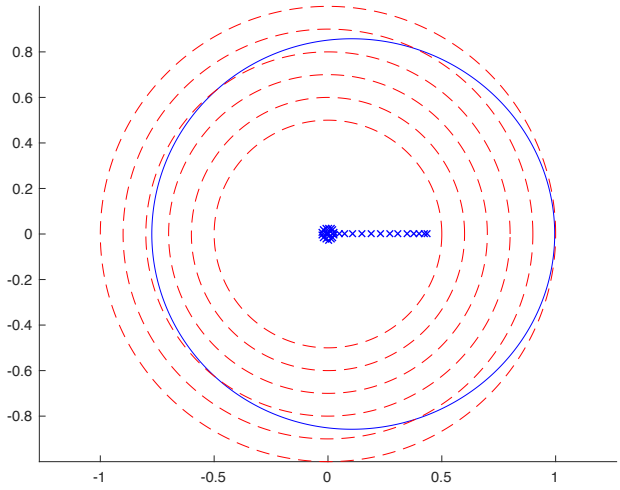

Figure 5.3: Eigenvalues (blue x's) and numerical range (solid blue curve) of the Gauss-Seidel iteration matrix corresponding to the upwind scheme (5.2). The red dashed circles are the ones on which $\lambda_{\min}(\mu(\sigma, A))$ was computed.

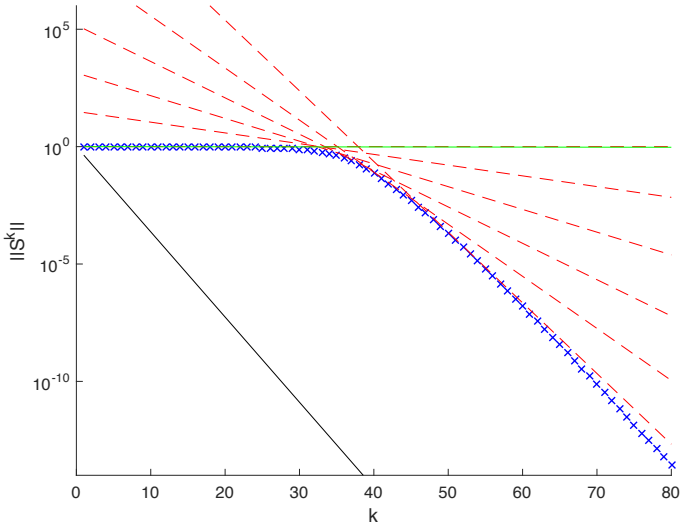

Figure 5.4: $\|S^k\|$ (blue x's) for the upwind iteration matrix together with several upper and lower bounds. The black line is the lower bound $\rho(S)^k$, the red dashed lines are upper bounds of the form $2+\delta$ from the disks in Figure 5.3, and the green line is the upper bound $\|S\|^k$.

bounds from this set to overtake those from the larger disks. This example shows not only does the spectrum fail to tell the full story of the behavior of powers of a nonnormal matrix, but the numerical range may not give sufficient bounds, even when it lies inside the unit disk. Here we used disks of varying radii between the spectrum and numerical range as our $\max\{1, 2+\delta\}$-sets, which can be used to piece together the transient behavior and show the transition from initial decay rates dictated by numerical radius to asymptotic decay rates dictated by the spectrum.

5.1.2 Matrix exponentials

We now seek bounds on norms of matrix exponentials $\|e^{tA}\|$. In this case, we investigate various half-planes $\Pi_c := \{z : \mathrm{Re}(z) < c\}$ as possible K-spectral sets, particularly the left half-plane Π_0. However, we can reduce the half-plane problem to the disk problem and study matrix exponentials in the same way that we studied matrix powers in the previous section.

Let $\varphi(z) = (z+1)/(z-1)$, so that $\varphi(A) = (A+I)(A-I)^{-1}$. This function is a bijective conformal mapping from Π_0 to the unit disk $\mathbb{D}$, which implies that Π_0 is a $\max\{1, 2+\delta_{\varphi(A)}\}$-set, with $\delta_{\varphi(A)}$ defined as in (3.23). As in the previous section, let us first consider a simple 2 by 2 matrix of the form

$$A = \begin{pmatrix} \lambda & a \\ 0 & \lambda \end{pmatrix}, \qquad \lambda < 0,\ a > 0.$$

The mapped matrix is then

$$\varphi(A) = \begin{pmatrix} \frac{\lambda+1}{\lambda-1} & -\frac{2a}{(\lambda-1)^2} \\ 0 & \frac{\lambda+1}{\lambda-1} \end{pmatrix}.$$

From the calculations in the previous section, we find that

$$\delta_{\varphi(A)} = -\int_0^{2\pi} \lambda_{\min}(s, \varphi(A))ds = -2 - \frac{a}{2\lambda},$$

meaning that the left half-plane Π_0 is a K-spectral set with $K = \max\{1, |a/2\lambda|\}$. For instance, with the matrix

$$A = \begin{pmatrix} -2 & 7 \\ 0 & -2 \end{pmatrix}, \tag{5.3}$$

we have $K \leq 1.75$, meaning that e^{tA} can never grow larger than 1.75 in norm. One can also investigate the decay on different time scales by doing the same calculations with different half-planes. We already know an upper bound on initial behavior based on the numerical abscissa, namely that $\|e^{tA}\| \leq e^{t\omega(A)} = e^{(1.5)t}$, as well as a lower bound on the asymptotic behavior based on the spectral abscissa, namely $\|e^{tA}\| \geq e^{t\alpha(A)} = e^{-2t}$. Using these techniques, we can determine other bounds on initial to transient behavior using half-planes contained within the right-half plane (i.e. Π_c with $c < \omega(A) = 1.5$) and bounds for transient to asymptotic behavior using half-planes contained within the left-half plane (i.e. Π_c with $c > \alpha(A) = -2$). We can see some example bounds using various half-planes together with the bounds based on the numerical and spectral abscissas in Figure 5.5.

For another example of a heavily nonnormal matrix we can analyze in the same way, let us consider a 60 by 60 Chebyshev spectral approximation A to an advection-diffusion

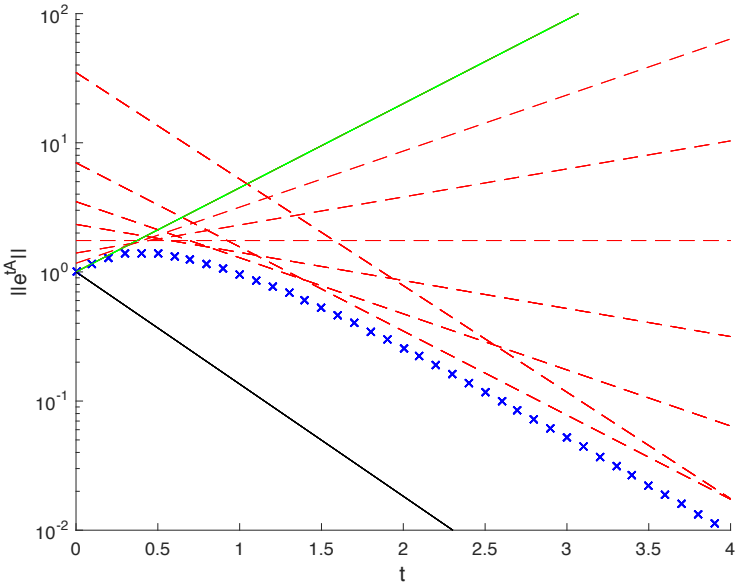

Figure 5.5: $\|e^{tA}\|$ (blue x's) for the matrix A in (5.3) with several upper and lower bounds. The black line is the lower bound $e^{t\alpha(A)}$, the red dashed lines are upper bounds of the form $2+\delta$ from various half-planes Π_c corresponding to lines of slope c on the semi-log plot, and the green line is the upper bound $e^{t\omega(A)}$.

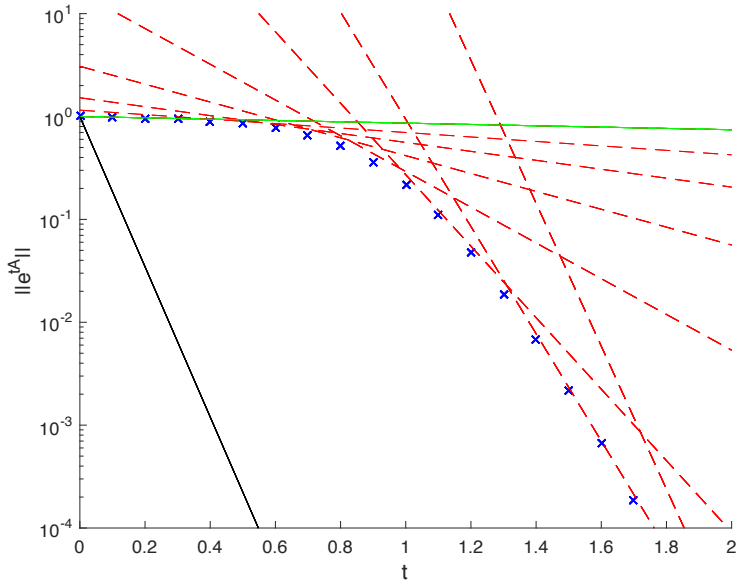

Figure 5.6: $\|e^{tA}\|$ (blue x's) for a Chebyshev spectral approximation A to $Lu = \eta u_{xx} + u_x, u(0) = u(1) = 0, \eta = 0.015$, with several upper and lower bounds. The black line is the lower bound $e^{t\alpha(A)}$, the red dashed lines are upper bounds of the form $2 + \delta$ from various half-planes Π_c corresponding to lines of slope c on the semi-log plot, and the green line is the upper bound $e^{t\omega(A)}$.

operator, $Lu = \eta u_{xx} + u_x, u(0) = u(1) = 0$, where $\eta = 0.015$. This example can be found in pp. 410-11 of [43]. Using various half-planes between the spectral and numerical abscissas as before, we obtain the bounds in Figure 5.6.

5.1.3 Other applications

We end this section by briefly mentioning some other possibilities for candidate K-spectral sets which may be useful for bounding functions of matrices in different applications. In particular, we recall that the region Ω in Lemma (3.1.1) need not be simply connected. It could consist of a union of disks, each of which encloses a part of spectrum. Such a set may be thought of as an alternative to ϵ-pseudospectra, where the boundary of Ω does

not necessarily have constant resolvent norm. Another example which has already proven useful is an annulus, with the spectrum contained inside of the outer disk and outside of the inner disk. In fact, it is proved in an upcoming paper by Crouzeix and Greenbaum [16] that various annular regions are $(1+\sqrt{2})$-spectral sets and that a more general convex region with a circular hole or cutout is a $(3+2\sqrt{3})$-spectral set. The proofs make use of our extension of the Crouzeix-Palencia result together with some estimates on $\lambda_{\min}$. These results are used to give bounds on the convergence rate of the GMRES algorithm for solving linear systems by choosing a region that includes the origin the origin, as well as for rational Krylov subspace methods for approximating $f(A)b$ by using a region which avoids the poles of the rational function approximating f.

5.2 Summary and open problems

In this dissertation, we have studied the use of K-spectral sets for bounding functions of nonnormal matrices. In particular, in [9] we have shown how the arguments in [17] can be extended in order to provide information about regions that contain the spectrum of A but not necessarily all of the numerical range $W(A)$. Some of the most interesting regions for applications are disks about the spectrum, which we have shown to be K-spectral sets for $K = \max\{1, 2+\delta\}$. This not only provides a new proof that if $W(A) \subset \mathbb{D}$, then $\mathbb{D}$ is a 2-spectral set for A, but it also provides useful bounds for norms of matrix powers $\|A^k\|$, as well as for norms of matrix exponentials $\|e^{tA}\|$ by conformally mapping half-planes to $\mathbb{D}$. This extension is being applied in [16] to annular regions, which are ideal K-spectral sets for studying rational functions with poles. Some interesting open problems here involve determining values of K for other K-spectral sets that may be useful for applications, or to prove some of the numerical bounds discussed in this thesis analytically, even if just for special cases.

We derived one property of optimal Blaschke products, as in Blaschke products that maximize $\|B(\Psi)\|$ where Ψ is a given matrix whose spectrum lies in the unit disk. In particular, if $\|B(\Psi)\| > 1$, then the left and right singular vectors of $B(\Psi)$ corresponding to

the largest singular value must be orthogonal to each other. This fact can be used to provide new proofs of Crouzeix's conjecture for special classes of matrices, and an interesting open problem is to determine other properties of this optimal B.

We described some of the numerical methods involved in our computational experiments involving K-spectral sets. Perhaps most useful is the implementation of numerical conformal mapping in Chebfun, which has since been adapted and improved in [24] and is certainly useful beyond the study of K-spectral sets. Another interesting series of numerical experiments involved the construction of near-normal dilations of nonnormal matrices, with the spectrum of the dilated operator around the boundary of the numerical range of the original matrix. While there has been much work aimed at proving the existence of dilations with various special properties, there has been little study of the behavior of functions of these dilations, especially when compared to the original operator. Doing this comparison for some matrix exponentials, we found that the dilated operator takes on a life of its own and eventually dominates the part corresponding to the original operator.

Attempting to prove Crouzeix's conjecture is a notoriously difficult problem, but this is what helped spark the recent flood of interest in K-spectral sets and nonnormal matrices. The Crouzeix-Palencia result is especially powerful, since it provides new mathematical tools which are useful both for the numerical range and for more general K-spectral sets, as shown in this dissertation. The best hope for proving the conjecture may be to show that $\|f(A)\| \leq \|f(A) + g(A)^*\|$ for the optimal f, at least for some certain classes of matrices. This is something we consistently observed numerically, and would prove the conjecture if true in general. For applications, a more interesting problem would be to continue to study properties of $\lambda_{\min}$ for regions such as disks, half-planes, or sets with holes.

BIBLIOGRAPHY

[1] T. Ando. Structure of operators with numerical radius 1. *Acta Sci.Math. (Szeged)*, 34:11–15, 1973.

[2] C. Badea and B. Beckermann. Spectral sets. In Leslie Hogben, editor, *Handbook of Linear Algebra*, chapter 107, pages 1–26. CRC Press, Oxford, 2 edition, 2013.

[3] C. Badea, B. Beckermann, and M. Crouzeix. Intersections of several disks of the Riemann sphere as k-spectral sets. *Com. Pure Appl. Anal.*, 8:1:37–54, 2009.

[4] C. Badea, M. Crouzeix, and B. Delyon. Convex domains and K-spectral sets. *Math. Zeitschrift*, 2:345–365, 2006.

[5] J. Baker, M. Embree, and J. Sabino. Fast singular value decay for Lyapunov solutions with nonnormal coefficients. *SIAM J. on Matrix Anal. and Appl.*, 36(2):656–668, 2014.

[6] B. Beckermann and M. Crouzeix. Operators with numerical range in a conic domain. *Archiv der Mathematik*, 88:547–559, 2007.

[7] B. Beckermann and M. Crouzeix. Faber polynomials of matrices for non-convex sets. *JAEN J. of Approx.*, 6 (2):219–231, 2014.

[8] C.A. Berger. A strange dilation theorem. *Notices Amer. Math. Soc.*, 12:590, 1965.

[9] T. Caldwell, A. Greenbaum, and K. Li. Some extensions of the Crouzeix-Palencia result. *SIAM J. on Matrix Anal. and Appl.*, 39(2):769–780, 2018.

[10] D. Choi. A proof of Crouzeix's conjecture for a class of matrices. *Linear Algebra and its Applications*, 438:8:3247–3257, 2013.

[11] D. Choi and A. Greenbaum. An algorithm for finding a 2-similarity transformation from a numerical contraction to a contraction. *SIAM J. Matrix Anal. and Appl.*, 36:1248–1262, 2015.

[12] M. Crouzeix. Bounds for analytical functions of matrices. *Integral Equations and Operator Theory*, 48:461–477, 2004.

[13] M. Crouzeix. Numerical range and functional calculus in Hilbert space. *Journal of Functional Analysis*, 244:668–690, 2007.

[14] M. Crouzeix. Spectral sets and 3×3 nilpotent matrices. *Topics in Functional and Harmonic Analysis*, 14:27–42, 2013.

[15] M. Crouzeix. Some constants related to numerical range. *SIAM Journal on Matrix Analysis and Applications*, 37:420–442, 2016.

[16] M. Crouzeix and A. Greenbaum. Spectral sets: numerical range and beyond. 2018.

[17] M. Crouzeix and C. Palencia. The numerical range as a spectral set. *SIAM J. Matrix Anal. and Appl.*, 38(2):649–655, 2017.

[18] B. Delyon and F. Delyon. Generalizations of von Neumann's spectral sets and integral representation of operators. *Bull. Soc. Math. France*, 127:25–41, 1999.

[19] T.A. Driscoll. Algorithm 756: A matlab toolbox for schwarz-christoffel mapping. *ACM Trans. Math. Softw.*, 22(2):168–186, 1996.

[20] T.A. Driscoll, N. Hale, and L.N. Trefethen. Chebfun user's guide. *Pafnuty Publications, Oxford*, 2014.

[21] J. Earl. A note on bounded interpolation in the unit disc. *J. London Math. Soc.*, 13:419–423, 1976.

[22] C. Glader, M. Kurula, and M. Lindström. Crouzeix's conjecture holds for tridiagonal 3×3 matrices with elliptic numerical range centered at an eigenvalue. *SIAM J. Matrix Anal. and Appl.*, 39:346–364, 2018.

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

[24] A. Gopal and L.N. Trefethen. Representation of conformal maps by rational functions. *Numer. Math.*, to appear 2018.

[25] A. Greenbaum, T. Caldwell, and K. Li. Near normal dilations of nonnormal matrices and linear operators. *SIAM J. on Matrix Anal. and Appl.*, 37(4):1365–1381, 2016.

[26] A. Greenbaum and D. Choi. Crouzeix's conjecture and perturbed Jordan blocks. *Linear Algebra and its Applications*, 436:2342–2352, 2012.

[27] K.E. Gustafson and D.K.M. Rao. *Numerical Range*. Springer, New York, USA, 1997.

[28] P. Henrici. *Applied and computational complex analysis*, volume 3. Wiley-Interscience, 1985.

[29] N.J. Higham. *Functions of Matrices: Theory and Computation*. Society for Industrial and Applied Mathematics, Philadelphia, USA, 2008.

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

[31] N. Kerzman and E.M. Stein. The Cauchy kernel, the Szegö kernel, and the Riemann mapping function. *Math. Ann.*, 236:85–93, 1978.

[32] N. Kerzman and M. Trummer. Numerical conformal mapping via the Szegö kernel. *Jour. of Comp. Appl. Math*, 14:111–123, 1986.

[33] H.-O. Kreiss. Über die stabilitätsdefinition für differenzengleichungen die partielle differential-gleichungen approximieren. *BIT*, 2:153–181, 1962.

[34] Y. Nakatsukasa, O. S'ete, and L.N. Trefethen. The AAA algorithm for rational approximation. *SIAM J. Sci. Comp.*, to appear 2018.

[35] K. Okubo and T. Ando. Constants related to operators of class c_ρ. *Manuscripta Math.*, 16:4:385–394, 1975.

[36] C. Palencia. Some contributions to M. Crouzeix's conjecture. Numerical Methods for Evolution Equations, 2016.

[37] V. Paulsen. *Completely bounded maps and operator algebras*. Cambridge Univ. Press, 2002.

[38] R. Remmert. *Theory of Complex Functions*. Springer, 1991.

[39] J. Schäffer. On unitary dilations of contractions. *Proc. Amer. Math. Soc.*, 6:322, 1995.

[40] M.N. Spijker. On a conjecture by LeVeque and Trefethen related to the Kreiss matrix theorem. *BIT*, 31:551–555, 1991.

[41] G. Symm. An integral equation method in conformal mapping. *Num. Math.*, 9:250–258, 1966.

[42] B. Sz.-Nagy and C. Foias. *Harmonic Analysis of Operators on Hilbert Space*. Norh-Holland, New York, 1970.

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

[44] J. von Neumann. Eine spektraltheorie für allgemeine operatoren eines unitären raumes. *Math. Nachrichten*, 4:258–281, 1951.