MATH 245 Linear Algebra
Bucknell University
5 December 2023
Problem Included

The Foundations of the Crouzeix Conjecture: Polynomial Functions of Matrices and Their Numerical Range

Sofi Jeffrey

Introduction

Just as Newtonian mechanics have been understood for centuries, the oldest and most simple objects in mathematics are generally well-behaved and well-understood. We often hear of new and innovative ways complex mathematics is applied in science and engineering, and we obsessively study and replicate the clever genius of countless deceased mathematicians. Oftentimes, this can lull a naive mathematician into the false comfort that mathematics is static and complete. Fortunately, *or unfortunately depending on one's point of view*, this could not be farther from the truth. There exists an endless breadth of open questions and unsolved problems, both new and old. In this paper, we will establish the understanding necessary for the Crouzeix Conjecture,[3] an unsolved problem first posed in 2004, as well as discuss progress made thus far. Notably, it remains unproven for matrices as small as 3×3. If nothing else, we hope to demonstrate that novel mathematics surrounding relatively simple mathematical objects does exist in a contemporary setting.

Polynomial Functions of Matrices

Polynomial functions are as well-behaved as a dead dog. As such, it is natural and meaningful to extend their definition to allow square matrices as inputs.[4]

Definition. *Let A be an $m \times m$ matrix, I be the $m \times m$ identity matrix, and p be the polynomial function $p(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n$ with real coefficients $a_0, a_1, \ldots, a_n \in \mathbb{R}$, then*

$$p(A) = a_0I_m + a_1A + a_2A^2 + \ldots + a_nA^n$$

Left alone, this expression would be very computationally intensive to solve, however it is possible to use diagonalisation to factor a matrix polynomial in a very similar manner to how one would factor a matrix.

Theorem 1. *Let $n \times n$ matrix A be similar to diagonal matrix D, by invertable matrix X such that $A = XDX^{-1}$ and p be a polynomial function, then*

$$p(A) = Xp(D)X^{-1}$$

The utility of this expression cannot be understated. Taking powers of a diagonal matrix is a relatively effortless calculation.

$$D = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix} \longrightarrow D^n = \begin{bmatrix} a^n & 0 & 0 \\ 0 & b^n & 0 \\ 0 & 0 & c^n \end{bmatrix}$$

When considering polynomials of very large degrees, it becomes clear that the calculation would not be feasible without diagonalisation.

With this in mind, we can consider how the norms of these matrices may yield additional information from the diagonalization. We can define $\mathbb{R}^{n\times n}$ to be a normed vector space such that

$$\max_{\|x\|_2=1} \|Ax\|_2$$

which is equivalent to the definition

$$\|A\| = \sigma_1$$

for $A \in \mathbb{R}^{n\times n}$, such that $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_n > 0$ is the ordered list of nonzero singular values of A which are defined as

$$\sigma_j = \sqrt{\lambda_j}$$

where each λ_j is a unique, non-zero eigenvalue of A^TA. This is the *spectral matrix norm* of A.[5]

We can then compare the norms of these matrices, which are related by the following equation

$$\|p(A)\| = \|X\| \cdot \|p(D)\| \cdot \left\|X^{-1}\right\|$$

which can be rearranged as

$$\|p(A)\| = \|X\| \cdot \left\|X^{-1}\right\| \cdot \|p(D)\|$$

Significant meaning can be drawn from this equation by further simplifying the expression and analysing the norms $\|X\|$ and $\|X^{-1}\|$.

Lemma 1. *Let X be a nonsingular matrix, $\sigma_1(X)$ be the largest non-zero singular value of X, and $\sigma_n(X)$ be the smallest non-zero singular value of X. If we define $\|X\| = \sigma_1(X)$, we can define*

$$\left\|X^{-1}\right\| = \frac{1}{\sigma_n(X)}$$

Therefore the following substitution can be made, noting that it is now an inequality as diagonalizing matrix X can be chosen such that σ_1/σ_n is arbitrarily large.

$$\|p(A)\| \leq \frac{\sigma_1}{\sigma_n} \cdot \|p(D)\|$$

This ratio, σ_1/σ_n, of the largest and smallest singular value of a matrix is the condition number of the matrix.

Definition. *Let A be a nonsingular matrix. Then, the ratio of the largest non-zero singular value of A and the smallest non-zero of A is the condition number of A.*

In the context of this question, we will define the spectral condition number of A, σ_1/σ_n, as the condition number of X, the matrix which diagonalizes A.

With the spectral condition number in mind, let us revisit the algorithm for the diagonalisation of a matrix. The following result is derived in the appendix. It can be shown that for matrix A which is diagonalized by nonsingular matrix X and diagonal matrix D, such that $A = XDX^{-1}$, where $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ the sigular values of A ordered greatest to least, that

$$\|p(D)\| = p(\lambda_1)$$

and thus

$$\|p(A)\| \leq \frac{\sigma_1}{\sigma_n} \cdot p(\lambda_1)$$

Numerical Range

Of all possible subspaces defined around a matrix, only few are useful. The *numerical range* of a matrix is one of these subspaces.

Definition. *Let A be an $n \times n$ matrix and x be a vector in the inner product space $\mathbb{R}^n$. A number $z \in \mathbb{C}$ is in the numerical range of A, $W(A)$, if it satisfies the following*

$$z \in W(A) = \{z = \langle Ax\ ,\ x\rangle : \|x\| = 1\}$$

The numerical range of A can be related to the eigenvalues of A by the following:

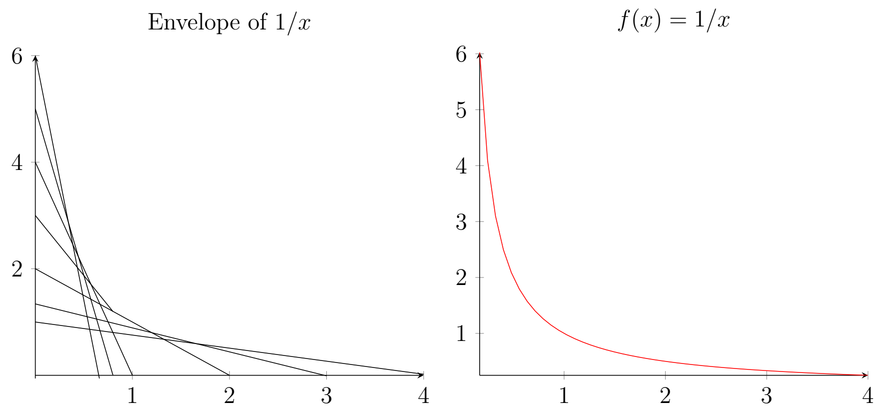
Theorem 2. *Let A be an $n \times n$ matrix, with eigenvalues $(\lambda_1, \lambda_2, \ldots, \lambda_n)$, and let $W(A)$ be the numerical range of A. Then,*

$$(\lambda_1, \lambda_2, \ldots, \lambda_n) \in W(A)$$

While all eigenvalues of A are in the numerical range of A, they are not the only values in A. The proof of the following theorem will not be included in this document.

Theorem 3. *Let A be a matrix with distinct eigenvalues $(\sigma_1, \sigma_2, \ldots, \sigma_n)$, and numerical range $W(A)$. The boundary curve of $W(A)$ is an algebraic curve on $\mathbb{C}$ with n foci equal to the eigenvalues of A.*[9]

This boundary curve can be constructed using an envelope.[8] For any algebraic curve, construct n lines tangent to this curve. Consider the limiting case as $n \to \infty$. The set of intersections of these tangent lines will exactly equal the set of points on the curve. Envelopes are best appreciated visually:



Here it can be seen how the intersections of the tangent lines to $1/x$ are points on the curve itself. An infinite number of these tangent lines would produce this curve exactly.

Although not as simple as taking a derivative, there exists an algorithm for finding the envelope of the boundary of $W(A)$. The first step is to express a matrix A in terms of its real and imaginary component.[9]

Lemma 2. *Let $n \times n$ matrix $A \in \mathbb{C}^{n\times n}$, and let $n \times n$ matrices $H, J \in \mathbb{R}^{n\times n}$. For all A, there exists some symmetric H and skew-symmetric J such that $A = H + iJ$, which can be constructed from the following:*

$$H = \left(\frac{A + A^H}{2}\right)$$

$$J = \left(\frac{A - A^H}{2i}\right)$$

We will now find the line tangent to the point on the boundary of $W(A)$ with the with the largest real component. Recall that $z \in \mathbb{C}$ is in the numerical range of A if there exists some $x \in \mathbb{C}^n$ such that $\{z = \langle Ax\ ,\ x\rangle : \|x\| = 1\}$.

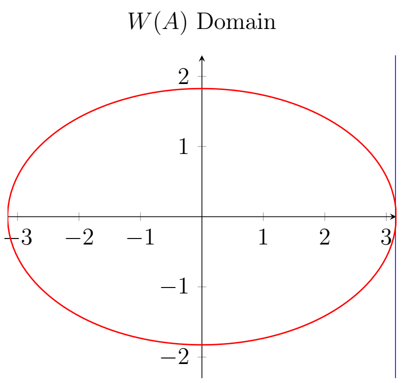
To bound $W(A)$ by its extent on the real axis, we must determine the greatest $\alpha \in \mathbb{R}$ where

$$\alpha = \max_{\|x\|=1}(\Re(\langle Ax\ ,\ x\rangle))$$

As we are interested purely in the real component of this value, we can compute with H instead of A, finding the inner product $\langle Hv\ ,\ v\rangle$, where v is an eigenvector of H, corresponding to the largest eigenvalue, λ_1, where $\|v\| = 1$. It follows that

$$\langle Hv\ ,\ v\rangle = \lambda_1$$

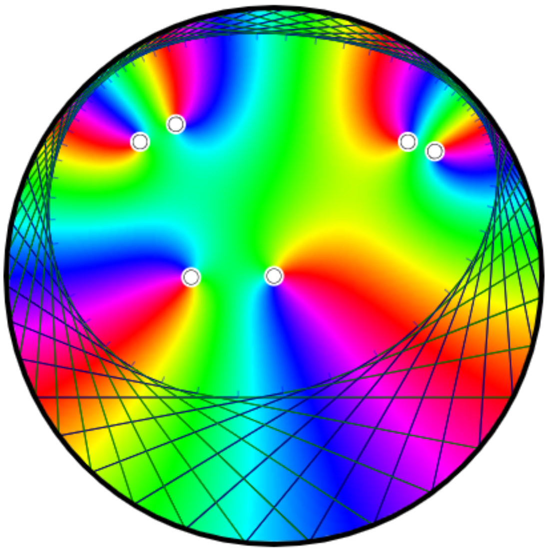
and thus, the real extent of $W(A)$ is the largest eigenvalue of H.



If $W(A)$ is represented by the red ellipse on the complex plane, the blue line represents the set of all $z \in \mathbb{C}$ such that $\lambda_1 = \Re(z)$.

Of course, it requires more than one line to form an envelope for $W(A)$. To find the remaining tangent lines, we simply need to rotate A such that the desired tangent line is orthogonal to the real axis. From there, undoing the rotation will yield the true tangent line. These calculations are demonstrated in the appendix.

When complete, the envelope of $W(A)$ will look as shown below.



Here it can be clearly seen how the envelope lines contain the algebraic curve that is the boundary of $W(A)$.

The Crouzeix Conjecture

We now have the understanding necessary to meaningfully interpret the statement of the Crouzeix Conjecture, which is as follows:[3]

$$\|p(A)\| \leq 2 \cdot \max_{z \in W(A)} |p(z)|$$

Notice how this falls simply from the diagonalization of A, which we expressed as

$$\|p(A)\| \leq \frac{\sigma_1}{\sigma_n} \cdot p(\lambda_1)$$

particularly since the eigenvalues of A necessarily fall within the numerical range. The Crouzeix Conjecture extends this idea to polynomial functions of any element in $W(A)$, claiming that the largest of these will be no less than half the spectral norm of $p(A)$. Thus, further bounding this conjecture is deeply linked to understanding the spectral condition number of A.

In 1997, Jiang et al. were able to demonstrate that the following upper bound

$$\frac{\sigma_1}{\sigma_n} \leq \left(1 + \frac{\alpha}{\delta}\right)^{2n-2}$$

holds for any $n \times n$ diagonalizable matrix A where α is the maximum upper diagonal entry in the Schur decomposition of A and δ is the smallest difference of two eigenvalues of A.[6] For the Crouzeix Conjecture itself, it has only been proven in the case of 2×2 matrices, although the upper bound has been reduced to $1 + \sqrt{2}$.[2]

Generally, it is exciting to see progress made on this problem, and fascinating how linear algebra applies not only in other fields, but as an interesting and developing field of study in and of itself.

References

[1] Carl Cowen and Elad Harel. An effective algorithm for computing the numerical range. 09 1995.

[2] M. Crouzeix and C. Palencia. The numerical range is a (1+sqrt(2))-spectral set. *SIAM Journal on Matrix Analysis and Applications*, 38(2):649–655, 2017.

[3] Michel Crouzeix. Bounds for analytical functions of matrices. *Integral Equations and Operator Theory*, 48(4):461–477, Apr 2004.

[4] I. Gohberg, P. Lancaster, and L. Rodman. *Matrix Polynomials*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104), 1982.

[5] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, 2012.

[6] Erxiong Jiang and Peter C.B. Lam. An upper bound for the spectral condition number of a diagonalizable matrix. *Linear Algebra and its Applications*, 262:165–178, 1997.

[7] S. Leon and L. de Pillis. *Linear Algebra with Applications*. Pearson Education, 2020.

[8] R. C. Yates. A handbook on curves and their properties. *The Mathematical Gazette*, 32(299):90–91, 1948.

[9] Paul F. Zachlin and Michiel E. Hochstenbach. On the numerical range of a matrix. *Linear and Multilinear Algebra*, 56(1-2):185–225, 2008.

Appendix

Applying a Polynomial Function to a Diagonal Matrix

Recall that for matrix A which is diagonalized by nonsingular matrix X and diagonal matrix D, such that $A = XDX^{-1}$, where $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ are unique eigenvalues of A

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

Notably, D is a diagonal matrix with non-zero entries corresponding to the eigenvalues of A. As D is diagonal, $p(D)$ can be easily calculated as

$$p(D) = \begin{bmatrix} p(\lambda_1) & 0 & \cdots & 0 \\ 0 & p(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p(\lambda_n) \end{bmatrix}$$

Lemma 3. *For any diagonal matrix D with non-zero entries $(d_1, d_2, \ldots, d_n)$, $\|D\| = max(d_j)$*

Proof. Let D be a diagonal matrix such that

$$D = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix}$$

As D is diagonal, it therefore must be symmetric, thus $D = D^T$, and therefore

$$D^T D = D^2$$

which can be easily calculated as

$$D^2 = \begin{bmatrix} (d_1)^2 & 0 & \cdots & 0 \\ 0 & (d_2)^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (d_n)^2 \end{bmatrix}$$

The characteristic polynomial $det(D^2 - \lambda I)$ will then be equal to

$$det(D^2 - \lambda I) = ((d_1)^2 - \lambda)((d_2)^2 - \lambda) \cdots ((d_n)^2 - \lambda)$$

Which has solutions $((d_1)^2, (d_2)^2, \ldots, (d_n)^2)$. Therefore, the eigenvalues of $D^T D$ are the squares of the entries of D. Given that $\|D\| = max(\sqrt{\lambda_j})$ where each λ_j is an eigenvalue of $D^T D$, and that the eigenvalues of $D^T D$ are $((d_1)^2, (d_2)^2, \ldots, (d_n)^2)$, it follows that

$$\|D\| = max(d_j)$$

□

Using this lemma, we can substitute $\|p(D)\|$ for $p(\lambda_1)$, where λ_1 is the greatest eigenvalue of A such that

$$\|p(A)\| \leq \frac{\sigma_1}{\sigma_n} \cdot \|p(D)\|$$

becomes

$$\|p(A)\| \leq \frac{\sigma_1}{\sigma_n} \cdot p(\lambda_1)$$

Calculating the Envelope of the Numerical Range

Although similar to the calculation of the vertical tangent line, the algebraic algorithm for applying rotations to matrices and $W(A)$ may not be intuitively obvious. This is the algorithm proposed by Cowen in 1995.[1]

Begin with $n \times n$ matrix A, which may have real or complex entries. Let the real part of A be H defined as shown in Lemma 2.

First, we must rotate matrix A by some angle $-\theta$, which we can express using Euler's formula. All primed matrices will denote a matrix after the rotation.

$$A' = Ae^{-i\theta}$$

We can also rotate both sides of our matrix decomposition $A = H + iJ$ such that

$$Ae^{-i\theta} = (H + iJ)\, e^{-i\theta}$$

We can now substitute using the full identity

$$e^{i\theta} = \cos\theta + i\sin\theta$$

which yields

$$H\left(\cos\theta - i\sin\theta\right) + iJ\left(\cos\theta - i\sin\theta\right)$$

Since matrix A' must be factorable into some matrices H', J' such that $A' = H' + iJ'$,[9] it follows that

$$H' = H\left(\cos\theta - i\sin\theta\right)$$

We will now follow the same steps as previously, however now considering H', rather than H. Notibly, this is possible since $W(A)$ is invarient under rotation.[9] We will take λ_1' as the largest eigenvalue of H', and find the vertical line on the complex plane corresponding to all $z \in \mathbb{C}$ such that $z = \lambda_1' + bi$ for any $b \in \mathbb{R}$. This line is tangent to $W(A')$, and thus the rotation must be reversed to find a line tangent to $W(A)$.

The standard rotation matrix in $\mathbb{R}^2$ is defined as

$$R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

To reverse the rotation, we simply multiply

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \lambda_1' \\ b \end{bmatrix}$$

We then find that the line tangent to $W(A)$ are $z \in \mathbb{C}$ such that $z = a' + b'i$ where

$$a' = \lambda_1' \cos\theta - b\sin\theta$$

$$b' = \lambda_1' \sin\theta - b\cos\theta$$

To find the complete envelope for $W(A)$, repeat this process for all θ on $[0, 2\pi]$

Proofs

Proof of Theorem 1

Proof. Let $p(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n$ and $A = XDX^{-1}$

We can apply the polynomial function to both sides of the equation

$$\begin{aligned} p(A) &= p(XDX^{-1}) \\ p(A) &= a_0I_n + a_1(XDX^{-1}) + a_2(XDX^{-1})^2 + \ldots + a_n(XDX^{-1})^n \\ p(A) &= a_0(XX^{-1}) + a_1(XDX^{-1}) + a_2(XDX^{-1})^2 + \ldots + a_n(XDX^{-1})^n \end{aligned}$$

Recall[7] that $(XDX^{-1})^n = XD^nX^{-1}$, thus

$$p(A) = a_0XX^{-1} + a_1XDX^{-1} + a_2XD^2X^{-1} \ldots + a_nXD^nX^{-1}$$

Which, after factoring yields

$$p(A) = X(a_0I_n + a_1D + a_2D^2 + \ldots + a_nD^n)X^{-1}$$

Otherwise stated as

$$p(A) = Xp(D)X^{-1}$$

□

Proof of Theorem 2

Proof. Let A be an $n \times n$ matrix, with unique eigenvalues $(\lambda_1, \lambda_2, \ldots, \lambda_n)$, and corresponding normed eigenvectors $(v_1, v_2, \ldots, v_n)$ such that $\|v_j\| = 1$. By definition,

$$Av_j = \lambda_j v_j$$

And therefore,

$$\langle Av_j \ , \ v_j \rangle = \langle \lambda_j v_j \ , \ v_j \rangle$$

Factoring λ_j out of this inner product yields

$$\lambda_j \langle v_j \ , \ v_j \rangle$$

Since $\|v_j\| = 1$, the square of the norm, $\|v_j\| = \langle v_j \ , \ v_j \rangle = 1$. Thus, for $\|v_j\| = 1$

$$\langle Av_j \ , \ v_j \rangle = \lambda_j$$

Which is the definition of an element of $W(A)$, therefore,

$$\lambda_j \in W(A)$$

□

Proof of Lemma 1

Proof. Let $B = X^T X$, with eigenvalues $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ corresponding to eigenvectors $(v_1, v_2, \ldots, v_n)$. Then,

$$\begin{aligned} Bv_j &= \lambda_j v_j \\ B^{-1}Bv_j &= \lambda_j B^{-1} v_j \\ v_j &= \lambda_j B^{-1} v_j \\ \frac{1}{\lambda_j} v_j &= B^{-1} v_j \end{aligned}$$

Therefore the inverse of the eigenvalues of B are the eigenvalues of B^{-1}. We also know these are the non-zero eigenvalues of XX^T, since both right and left products of a matrix and its transpose, XX^T and $X^T X$, will always have the same non-zero eigenvalues.

The largest singular value of X is $\sigma_1(X) = \sqrt{\lambda_1}$, and since the largest eigenvalue of $\left(X^T X\right)^{-1}$ is $\frac{1}{\lambda_n}$, the largest singular value of X^{-1} must be $\frac{1}{\sigma_n}$. Therefore,

$$\left\|X^{-1}\right\| = \frac{1}{\sigma_n(X)}$$

□

Proof of Lemma 2

Proof. If we define

$$A = H + iJ$$

Then we can substitute

$$A = \left(\frac{A + A^H}{2}\right) + i\left(\frac{A - A^H}{2i}\right)$$

Which simplifies to

$$A = \left(\frac{A + A}{2}\right)$$

And

$$A = A$$

Therefore this is a valid decomposition of A into its real and imaginary component. □