Characterizing the Critical Curves of 4x4 Matrices

Reuben P. Retnam '17 and Brian C. Lins

Department of Mathematics and Computer Science, Hampden-Sydney College, Hampden-Sydney, VA 23943

ABSTRACT

We classified and analyzed the critical curves of 4x4 matrices, paying particular attention to unitarily irreducible cases. We investigated formulas for curves with two and three flat portions from the principle that curves with one or two flat portions could have their flat portions rotated to specific axes. We also performed investigations on other related topics, including ways to plot critical curves and ways to analyze their permutations.

Characterizing the Critical Curves of 4x4 Matrices

Dr. Brian Lins and Reuben Retnam

Department of Mathematics, Hampden-Sydney College

Abstract

We classified and analyzed the critical curves of 4x4 matrices, paying particular attention to unitarily irreducible cases. We investigated formulas for curves with two and three flat portions from the principle that curves with one or two flat portions could have their flat portions rotated to specific axes. We also performed investigations on other related topics, including ways to plot critical curves and ways to analyze their permutations.

Introduction

The numerical range is the range of the Rayleigh quotient $R = \frac{(x*)(M)(x)}{(x*)(x)}$, where x is a vector (a $1 * n$ matrix) and $x*$ consists of the conjugate transpose of x. The critical curves of the numerical range are the values where the Jacobian is not rank 2. We already know that for a 2x2 matrix, the numerical range is composed of ellipses and smoothly connected points – and for a 3x3 matrix, there are three cases: a reducible matrix (one that can be factored as $A = P^{-1}SP$), which ends up being the same case as the 2x2 matrix, and two irreducible matrix cases, whose general forms are shown below. Our goal was to be able to characterize the critical curves of 4x4 matrices, focusing on the cases where the matrices were unitarily irreducible.

Methods

First, I wrote a Mathematica program to exploit a possible new method of graphing critical curves. First, the matrix was factored into real (H) and imaginary (K) parts, using the formulas $H = \frac{A^*+A}{2}$ and $K = \frac{A-A^*}{2i}$. We set a differential ($d = 0.01$ for most of the plotting done), and then for $0 < j < 2\pi$ (increasing j by the step size of d each time), we let $n = (\cos j)(H) + (\sin j)(K)$ and $m = \frac{(\cos[j+(diff)])(H)+(\sin[j+(diff)])(K)}{diff}$. To calculate the x-coordinate of the intended point, we calculated $r = Real[(e^{ij})(Eigenvalues[n]) + e^{ij}(i)(Eigenvalues[m])]$ and $c = Imaginary[(e^{ij})(Eigenvalues[n]) + e^{ij}(i)(Eigenvalues[m])]$. Unfortunately, since the program used differentials to compute the graphs, and those derivatives approached infinity at certain points, it was not as effective as hoped.

In order to investigate the order of the eigenvalue crossings further, we created a program to compute the Taylor series for the specific eigenvalue curves of $\cos kH + \sin kK$. We computed the determinant of the matrix form manually, then used Mathematica to solve for the 1st through 4th series coefficients. Sadly, the series were too complicated to have a proper investigative use.

We attempted to investigate the permutations of the critical curves and relate them to properties of the matrix. We were not able to determine a relation between the properties of the matrix and the critical curves, but we determined that adding a single cycle to permutations of the critical curves must result in a single-cycle permutation, and adding two must necessarily result in a two-cycle.

The main result of our work this summer was the following theorem. We showed that all unitarily irreducible 4x4 matrices have three or less flat portions, which goes a long way towards full characterization of the critical curves by flat portions because the unitarily reducible case is fairly simple compared to the unitarily irreducible case (since the numerical range of a normal matrix is simply the convex hull of the eigenvalues.

Theorem: All unitarily irreducible 4x4 matrices have three or less flat portions.

Consider a 4x4 matrix A. Then for the numerical range of the matrix, we can rotate the numerical range, find a unitarily similar matrix with the same numerical range, or deforming the vertical and horizontal flat portions separately. We want $\cos\theta H + \sin\theta K$ to have a crossing (where $H = \frac{A^*+A}{2}$ and $K = \frac{A-A^*}{2i}$ for $W(A)$ for A to have a flat portion. With one flat portion, we can simply apply a rotational transform to A to rotate the flat portion onto the horizontal axis. Upon addition of the second flat portion, we can divide through by $\cos\theta$ to get $H + tK$, where $t \in \mathbb{R}$, since $\tan\theta$ has a range of the entire real line. Thus it is apparent that we can set the angle between the flat portions θ such that the flat portions are perpendicular to each other. Therefore, we can assume that any one flat portion was a horizontal flat portion, and any two flat portions were a horizontal and vertical flat portion. With this information, we want to show that a unitarily irreducible 4x4 matrix can only have three or less flat portions. First, we came up with a form for curves requiring more than two flat portions. First, we construct a matrix with a numerical range that contains two flat portions. We started with a matrix in two parts, real and imaginary – $H = \frac{A^*+A}{2}$ and $K = \frac{A-A^*}{2i}$, respectively, where we required two non-parallel flat portions. We know that a numerical range with these two flat portions will be deformable into a numerical range with perpendicular flat portions, and that since we can rotate the numerical range, we can assume that they are on the real and imaginary axes for simplicity. First, we consider the 'real matrix' H. In order to generate our first flat portion, we want to have two nonzero eigenvalues, a and b. We set the matrix up diagonally, with a and b in the lower right 2x2 block. However, since we want the eigenvalues to be on the diagonal and the matrix is Hermitian, the lower left and upper right 2x2 blocks must be identically zero. In addition, since we want two zero eigenvalues, we want the top left 2x2 block to be identically zero as well. For the 'imaginary matrix' K, we wanted to require the matrix be rank 2 and that it be Hermitian (as previously mentioned, both H and K must be Hermitian due to their definition). Due to the zero blocks in H, we can diagonalize the top left 2x2 block of K. The top right 2x2 block K_1 is unknown, but since we know that K is Hermitian, we can have the lower left 2x2 block equal K_1^*. Since the matrix is required to be rank 2, we know that it has 2 linearly independent rows. It is apparent that due to the zero pattern of the matrix, if we take any rows other than the first two as our linearly independent rows, we will obtain a matrix with a rank higher than 3. Hence, the lower two rows of K must be a linear combination of the first two rows, and thus the bottom right 2x2 block of the matrix must be the top right 2x2 block transformed by the same linear transformation as the transformation from the top left 2x2 block to the bottom left 2x2 block. We now solve for the exact nature of this linear transformation.

We have

$$Q \begin{bmatrix} c & 0 \\ 0 & d \end{bmatrix} = K_1^*$$

$$Q = K_1^* \begin{bmatrix} \frac{1}{c} & 0 \\ 0 & \frac{1}{d} \end{bmatrix}^{-1}$$

$$Q = K_1^* \begin{bmatrix} \frac{1}{c} & 0 \\ 0 & \frac{1}{d} \end{bmatrix}$$

Now, we know that the bottom right 2x2 block of K is equal to QK_1, so therefore it must be equal to $K_1^* \begin{bmatrix} \frac{1}{c} & 0 \\ 0 & \frac{1}{d} \end{bmatrix} K_1$.

So, finally, we end up with $H = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & a & 0 \\ 0 & 0 & 0 & b \end{bmatrix}$ and $K = \begin{bmatrix} K_0 & K_1 \\ K_1^* & K_1^* K_0^{-1} K_1 \end{bmatrix}$, where $K_0 = \begin{bmatrix} c & 0 \\ 0 & d \end{bmatrix}$ and $K_1 = \begin{bmatrix} e & f \\ g & h \end{bmatrix}$.

Now that we have a form for two flat portions, we can require three flat portions from this form by requiring that $\cos\theta H + \sin\theta K$ have a repeat eigenvalue. We perform a simplification by dividing through by $\sin\theta$, producing $\frac{\cos\theta H + \sin\theta K}{\sin\theta} = tH + K$, since $\cot\theta$ has a range of the entire real line. So to produce our third flat portion, we require $tH + K - \lambda I$ to have a repeat eigenvalue of 0, where I is the identity matrix. Since $c = \lambda$ or $d = \lambda$ produces a matrix with a rank no less than 3, we consider the bottom left 2x2 block of our matrix. We can apply the same principle that we used to find the bottom right 2x2 block of K to this new matrix $tH + K - \lambda I$. We have $K_1^* \begin{bmatrix} \frac{1}{c} & 0 \\ 0 & \frac{1}{d} \end{bmatrix} K_1 + \begin{bmatrix} ta - s & 0 \\ 0 & tb - s \end{bmatrix} = K_1^* \begin{bmatrix} \frac{1}{c-s} & 0 \\ 0 & \frac{1}{d-s} \end{bmatrix} K_1$.

So thus $\begin{bmatrix} ta-s & 0 \\ 0 & tb-s \end{bmatrix} = K_1^* \begin{bmatrix} \frac{1}{c-s}-\frac{1}{c} & 0 \\ 0 & \frac{1}{d-s}-\frac{1}{d} \end{bmatrix} K_1$ or rather, a diagonal matrix $D = K_1^* E K_1$, where $E = \begin{bmatrix} \frac{1}{c-s}-\frac{1}{c} & 0 \\ 0 & \frac{1}{d-s}-\frac{1}{d} \end{bmatrix}$. However, by diagonalization, we can make $E = \begin{bmatrix} 1 & 0 \\ 0 & e \end{bmatrix}$, where e is an arbitrary constant. By temporarily considering the values of K_1 as $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and by multiplying out $K_1^* E K_1$, we can show that $\overline{b}a + c\overline{d}e = 0$, and thus $e = -\frac{b\overline{a}}{\overline{c}d} = \frac{\frac{1}{d-s}-\frac{1}{d}}{\frac{1}{c-s}-\frac{1}{c}}$. We want to show this to be one to one. We have $e = f(s) = \frac{c(c-s)}{d(d-s)}$, where $c, d \in \mathbb{R}$. Let $f(b) = f(a)$, where a and b are arbitrary real numbers. Then

$$\frac{c(c-a)}{d(d-a)} = \frac{c(c-b)}{d(d-b)}$$

$$\frac{(c-b)}{(d-b)} = \frac{(c-a)}{(d-a)}$$

$$\frac{c}{(d-b)} - \frac{b}{(d-b)} = \frac{c}{(d-a)} - \frac{b}{(d-a)}$$

$$c(d-a) - b(d-a) = c(d-b) - a(d-b)$$

$$cd - ca - bd + ba = cd - cb - ad + ba$$

$$-ca - bd = -cb - ad$$

$$ad - ac = bd - bc$$

$$a(d-c) = b(d-c)$$

$$a = b$$

. So since $f(b) = f(a) \implies b = a$, f is one to one. Since f is one to one, for every e, there is only one corresponding s, so we know that for every matrix, there is only one more possible flat portion.

I wrote a macro to upload examples of various types automatically to an Imgur account (http://critcurves.imgur.com). Using this, I observed various interesting critical curves. Note, however, that the set of random matrices generated was of somewhat limited utility, since the set of 'interesting matrices' of a certain form is almost guaranteed to be measure 0 on the set of all produced matrices. Using the limited form, I then observed real and complex changes in each of the variables in each zero pattern of the matrix (15 primary forms, with 64 secondary forms). I recorded how each change affected the graph of the critical curve for each matrix. This data can be accessed online along with the graphs.

Conclusions

- We proved that for unitarily irreducible 4x4 matrices, their numerical range can have a maximum of three flat portions.
- We also observed and recorded the graphs that went along with each zero pattern of each matrix.
- We also made attempts at characterizing the critical curves by order of eigenvalue crossings and permutations, but they were unsuccessful.

Forthcoming Research

We would like to extend our research into the 5x5 case this semester.

Acknowledgements

We would like to thank the Hampden-Sydney Mathematics Department and the Hampden-Sydney Honors Council for providing the opportunities necessary to perform this research.

References

Psarrakos, Panayiotis J., and Michael J. Tsatsomeros. ”Numerical Range: (in) a Matrix Nutshell.” Washington State University. N.p., 12 Aug. 2002. Web.

Horn, Roger A., and Charles R. Johnson. Matrix Analysis. 2nd ed. N.p.: Cambridge UP, 2012. Print.