"Spectral Inclusion Sets for Matrices"

by

Mark D. Tronzo, P.E.

Submitted in Partial Fulfillment of the Requirements

For the Degree of

Master of Science

In the

Mathematics

Program

Youngstown State University

May, 2005

"Spectral Inclusion Sets for Matrices"

Mark D. Tronzo

I hereby release this thesis to the public. I understand that this thesis will be made available from the OhioLINK ETD Center and the Maag Library Circulation Desk for public access. I also authorize the University or other individuals to make copies of this thesis as needed for scholarly research.

Signature:

Mark D. Tronzo 5-17-05

Mark D. Tronzo, Student Date

Approvals:

John J. Buoni 5-19-05

John J. Buoni, Thesis Advisor **Date**

Angela Spalsbury 5-19-05

Angela Spalsbury, Thesis Advisor Date

David Pollack 5/20/05

David Pollack, Committee Member Date

8/9/05

Peter J. Kasvinsky, Dean of Graduate Studies Date

Introduction

This thesis investigates various methods of producing spectral inclusion sets for matrices. The purpose of this investigation is threefold. First of all, various methods will be compared in order to give insight into the fastest and sharpest method of producing spectral inclusion sets for various types of matrices (chapter eight). Secondly, new types of inclusion sets will be introduced by intersecting inclusion sets that are produced using very simple calculations (chapter nine). Finally, new results will be presented and a new method will be introduced for producing spectral inclusion sets of certain types of Toeplitz matrices (chapter ten).

Methods of producing spectral inclusion sets of matrices and operators have been developed, primarily, because of two shortcomings in the exact calculations of the eigenvalues. First of all, actual calculation of the spectrum for large matrices can take a great deal of time even on fast computers. Secondly, such calculations may produce erroneous results due to round-off errors. Such erroneous results are particularly prevalent when attempting to calculate the spectrum of ill-conditioned matrices. All of the methods examined in this thesis avoid one or both of these problems.

Two groups of methods will be considered in this thesis. One group is classified as 'simple' (chapters one through three). The 'simple' methods are those methods which are limited to adding, subtracting, multiplying, dividing and raising matrix elements to powers. Such methods are both fast and avoid round-off errors. These 'simple' methods include what are called in this paper 'pre-Gerschgorin', Gerschgorin's method and Parker's second theorem.

A second group of methods considered in this paper are classified as 'involved' (chapters four, five, and seven). Such methods utilize extensive searching, large numbers of similarity transformations, and/or a large number of incremental calculations. These methods avoid round-off error and produce very small inclusion sets but may require considerable calculation time. Among the 'involved' methods are Cassini, Brualdi, minimal Gerschgorin, the numerical range, and the pseudospectra.

No one will be surprised that the major drawback with all of these methods is that they produce sets that are only guaranteed to include the spectrum. While each method produces a set that includes the eigenvalues, the set produced is usually somewhat larger than the actual spectrum of the matrix or operator. Furthermore, each method, except perhaps the pseudospectra, have varying, and sometimes unpredictable, degrees of 'sharpness' depending upon the application.

In chapters seven and eight it will be shown that the pseudospectra is the most powerful of the 'involved' methods. In most cases, the pseudospectra will produce a significantly smaller spectral inclusion set than any other method. Even in those few instances in which another method produces a smaller inclusion set, that set will only be slightly smaller than the pseudospetra's set. Therefore, it can be said that *the pseudospectra is the only method that consis-*

tently produces small spectral inclusion sets

In chapter nine it will also be shown that it is possible to produce relatively sharp spectral inclusion sets by *intersecting* sets produced by the 'simple' methods. This means that small spectral inclusion sets may be produced by using a minimal amount of calculation time. Therefore, this thesis will establish new methods of producing sharp spectral inclusion sets very quickly through the intersection of sets.

In the chapter ten of this thesis, two new theorems will be presented and a new method will be introduced for producing spectral inclusion sets of certain types of Toeplitz matrices. The new theorems will be based on Gerschgorin's theorem and the minimal Gerschgorin theorem. It will be demonstrated that the minimal Gerschgorin set can be used, in a new way, to *very quickly* produce relatively small inclusion set for Toeplitz matrices.

Note: Throughout the paper, $\sigma(A)$, is the spectrum of A and, unless otherwise indicated, is computed in Matlab with single precision.

Acknowledgements

I want to thank Dr. John Buoni and Dr. Angela Spalsbury, my advisors, for all of their help with this thesis. Dr, Spalsbury first suggested that I consider research in the Psuedo-Spectra early in 2003. This led to Dr. Buoni's suggestion that I also consider other spectral inclusion sets such as the Gerschgorin Disks and the Numerical Range. The suggestions and guidance of Dr. Buoni and Dr. Spalsbury during the research and writing of this thesis proved to be invaluable to its final production. I want to particularly thank them for their painstaking reading and rereading of my thesis.

I also want to thank Dr. David Pollack for agreeing to be on my thesis committee on very short notice. Even though he was given the opportunity to read only the later versions of the thesis, he made very important suggestions and corrections that helped to bring the thesis to a higher level.

I want to use this opportunity to offer a *very* belated thank you to Dr. Hyun W. Kim, Chair of the Mechanical Engineering Department here at Youngstown State University. Dr. Kim was my mentor during my first Master's Degree Program (in Engineering) at Youngstown State during 1983-1985. Dr. Kim's guidance helped greatly during my engineering studies. Dr. Kim's instruction prepared me very well for what turned out to be a very rewarding engineering career.

I want to thank my parents, Thomas C. Tronzo and Olga P. Tronzo, who supported me and encouraged me throughout all of my life. Without their help and support I would have made no progress in any area of my life.

Above all I offer this thesis to the glory of the Lord Jesus Christ, the creator of heaven and earth in the hope that we will live with Him forever and ever in heaven. "For by Him (Jesus Christ) all things were created, both in the heavens and on earth, visible and invisible, whether thrones or dominions or rulers or authorities – all things have been created by Him and for Him. And He is before all things, and in Him all things hold together." (Colossians 1:16-17).

Mark D. Tronzo, P.E.

New Brighton, Pa
August 2005

Contents

1 Pre-Gerschgorin and Gerschgorin Methods

This is the first of two chapters that deal with 'simpler', norm-type methods of creating spectral inclusion sets of matrices. 'Simpler' includes methods that involve nothing more than arithmetic operations in real numbers. 'Simpler' does not include methods that implement extensive searches, similarity transformations, iterations, etc. Therefore, spectral inclusion sets may be created very quickly using these simpler methods.

The 'simpler' methods considered in this chapter date through the 1930's and 1940's. During the study of these methods for this thesis the idea of trying to intersect the various inclusions sets was developed. Therefore, these first two chapters introduce these simpler methods and lay the foundation for the 'intersection method' that is presented in chapter nine.

Section 1.1 Pre-Gerschgorin (Bounds on spectral radius and numerical Range)

This section will consider methods of bounding the spectrum that were discovered before 1947. 'Pre-Gerschgorin', therefore, may seem like a misnomer since Gerschgorin produced his theorem in 1931! Yet the title seems appropriate since the Gerschgorin theorem was either forgotten or simply ignored until Olga Taussky and Alfred Brauer studied and propagated it in the 1940's. So, 'Pre-Gerschgorin' includes those methods that were developed before Gerschgorin's work was widely known.

The 'pre-Gerschgorin' methods accomplish two things: they approximate the spectral radius and produce bounds on the numerical range.(For $A \in C^{nxn}$, the numerical range is defined as $W(A) = \{\langle A\mathbf{x}, \mathbf{x}\rangle : \mathbf{x} \in C^n \text{ and } \|x\| = 1\}$. For a more complete discussion of the numerical range see chapter five). The approximation of the spectral radius takes the form of an upper bound of the absolute value of the eigenvalues. At least some of developers of these early techniques specifically had the numerical range on their minds as they developed their methods. It was understood that numerical range contained the spectrum of the matrix but calculating the entire numerical range for a matrix was very difficult. Creating sets that included the numerical range was the next best thing. These bounds on the spectral radius and the numerical range lead to a very natural way of producing spectral inclusion sets. Such sets will be examined in this section.

Since these methods 'bound' the numerical range a natural question arises: since computers and algorithms now exist for calculating the entire numerical range why not just use the numerical range? After all, the numerical range will produce a subset of the 'pre-Gerschgorin' methods and, therefore, produce sharper results. Actually, it is true that the numerical range produces sharper results than these older methods but for large matrices, the numerical range may take several minutes to calculate even on very fast computers. On the other hand, the spectral inclusion sets produced by the pre-Gerschgorin methods can be calculated very quickly on the computer . Furthermore, the pre-Gerschgorin methods have the advantage that they are very easy to use in combination with other methods. This feature will be very important for developing the 'Intersection Methods' of chapter nine.

As each of the pre-Gerschgorin methods was published in the 1930s and 1940s, the author would say something like 'this method appears to give sharper results than previous methods'. That is, each author suspected that his or her method *always* produced a subset of the previous methods but could not actually prove it. Well, as it turns out, it could not be proven because it was not true! It will be shown that there are about five methods that are not, *in general*, subsets of the other methods but for any given matrix one method will produce a subset of the others. This fact could not have been of much practical value in the 1930's and 1940's because it would be very time consuming to do the hand calculations based on all of these different methods in order to find the sharpest one for each application. Today, this can be done very easily with a computer. Therefore, computers allow a fresh look at these old methods.

General Form of the Pre-Gerschgorin inclusion sets

All of the 'pre-Gerschgorin' methods bound the eigenvalues in generally the same way. Each method produces a bound on the spectral radius which when plotted takes the form of a circle centered at the origin. In addition, each method produces separate bounds on the real and imaginary parts of the eigenvalues which, when plotted, take the form of a rectangular box, also centered at the origin.

So, each of the 'pre-Gerschgorin' methods produces a spectral inclusion set that has one of the following forms:

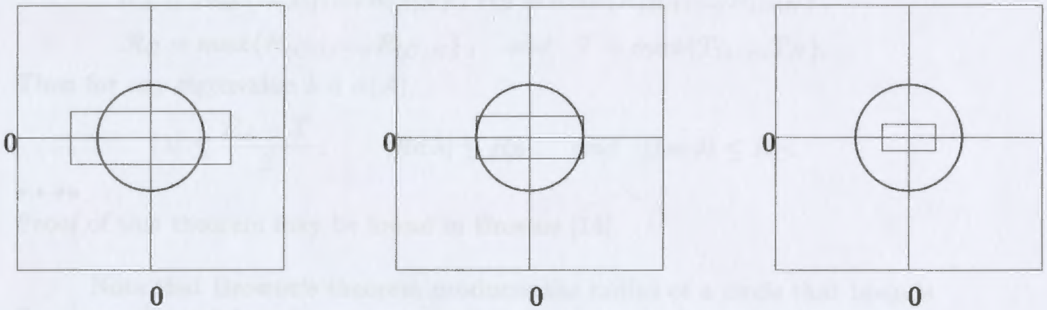


Figure 1.1

The circle and the rectangular box each contain the spectrum. So, the spectrum is contained in the intersection of the rectangular box and the circle. Note that in the graph in the far right of figure 1.1 that the intersection of the circle and box will leave only the rectangular box.

As it turns out, each method in this chapter will produce exactly the same bound on the real and imaginary parts of the eigenvalues. That is, all the methods will produce the same rectangular box. Therefore, each succeeding method is an attempt to bound the spectral radius. Graphically, each new method is an attempt to reduce the radius of the circle centered at the origin.

Section 1.1.1 Browne's Theorem

A method for bounding eigenvalues was proposed by Browne [14] in 1930. Browne estimated the upper bounds of eigenvalues using row sums of the original matrix and row sums of two other matrices related to the original.

Beginning with a square, complex matrix A, Browne defined two Hermitian matrices, B and C, as follows:

$$B = \frac{A + A^*}{2}, \quad and \quad C = \frac{A - A^*}{2i}.$$

Using the row sums from matrices A,B, and C and column sums from A, Browne was able to find a bound for the eigenvalues of A as well as separate bounds for the real and imaginary parts of the eigenvalues.
Browne's Theorem may then be stated as follows:

Theorem 1.1 (Browne's) Let $A \in C^{nxn}$. Let

$$B = \frac{A + A^*}{2} \quad and \quad C = \frac{A - A^*}{2i}.$$

Let $R_{(A)i}$,$R_{(B)i}$,and $R_{(C)i}$,be the sums of the absolute values of the elements in the i^{th} row of the matrices A,B, and C, respectively. Let T_i be the sum of the absolute values of the elements in the i^{th} column of A. Define:

$$R_A = max\{R_{(A)1}, ..., R_{(A)N}\}\ ,\ R_B = max\{R_{(B)1}, ..., R_{(B)N}\}\ ,$$

$$R_C = max\{R_{(C)1}, ..., R_{(C)N}\}\ , \quad and \quad T = max\{T_1, ..., T_N\}.$$

Then for any eigenvalue $\lambda \in \sigma(A)$,

$$|\lambda| \leq \frac{R_A + T}{2}, \qquad |Re\,\lambda| \leq R_B\ , \quad and \quad |Im\,\lambda| \leq R_C.$$

● ● ●●

Proof of this theorem may be found in Browne [14].

Note that Browne's theorem produces the radius of a circle that bounds the eigenvalues of A and a rectangular box that bounds the real and imaginary parts of the eigenvalues of A. This means that the circle contains the spectrum and the rectangular box contains the spectrum. Therefore, the spectrum is contained in the intersection of the circle and the rectangular box.

Example 1.2 Let

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & -1 & 2 \\ 1 & 2 & 1 & -1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & .5 & -1 \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{-1.79, 2.17, .81 + 341i, .81 - 341i\}.$$

(Note that the eigenvalues in this thesis were computed with Matlab in single precision). (The solution details of this particular example may be found in the Appendix).

Then

$$\mathbf{B} = \frac{A + A^*}{2} = \begin{pmatrix} 0 & .5 & -.5 & 1.5 \\ .5 & 2 & .5 & 0 \\ -.5 & .5 & 1 & .75 \\ 1.5 & 0 & .75 & -1 \end{pmatrix},$$

$$\mathbf{C} = \frac{A - A^*}{2i} = \begin{pmatrix} 0 & .5i & .5i & -.5i \\ -.5i & 0 & -.5i & i \\ -.5i & .5i & 0 & -.25i \\ .5i & -i & .25i & 0 \end{pmatrix},$$

and for any $\lambda \in \sigma(A)$,

$$|\lambda| \leq \frac{R_A + T}{2} = \frac{5+5}{2} = 5\,, \qquad |Re\,\lambda| \leq R_B = 3.25\,, \quad and \quad |Im\,\lambda| \leq R_C = 2.$$

This is graphed in figure 1.2 below. Note that the bounds on 'a' and 'b' (represented by the rectangle in figure 1.2) are contained completely within the bounds for λ (represented by the circle in figure 1.2). Therefore, in this example, the bound on λ (represented by the circle) is superfluous and may be ignored.

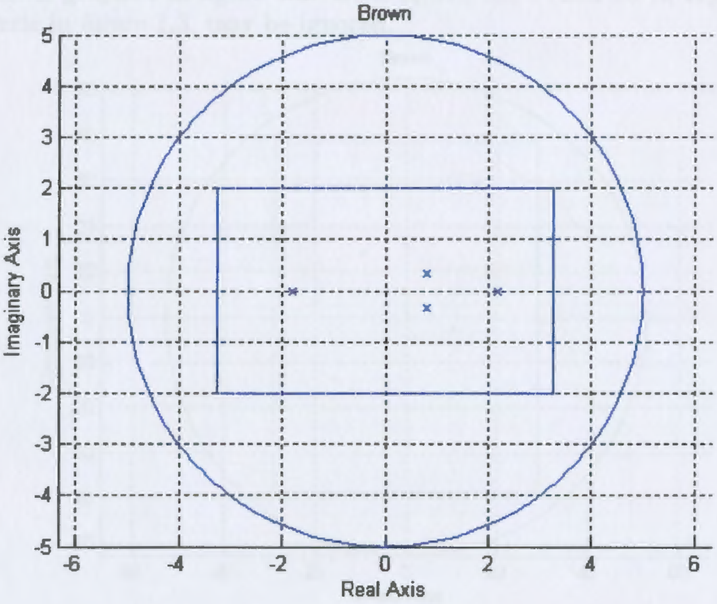


Figure 1.2

In these figures, the eigenvalues are represented by the X's

Example 1.3 Let

$$\mathbf{A} = \begin{pmatrix} -5+i & 2 & -6 & 15i \\ 7-6i & 15i & -7+3i & 8+5i \\ 19 & 8-4i & 13+9i & 10 \\ 16+9i & 15+2i & -9+3i & -7+2i \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{-15.42-13.78i, 15.37+29.14i, 10.33-2.72i, -9.28+14.36i\}.$$

Then

$$\mathbf{B} = \frac{A+A^*}{2} = \begin{pmatrix} -5 & 4.5+3i & 6.5 & 8+3i \\ 4.5-3i & 0 & .5+3.5i & 11.5+1.5i \\ 6.5 & .5-3.5i & 13 & .5-1.5i \\ 8-3i & 11.5-1.5i & .5+1.5i & -7 \end{pmatrix},$$

$$\mathbf{C} = \frac{A-A^*}{2i} = \begin{pmatrix} 1 & -3+2.5i & 12.5i & 12+8i \\ -3-2.5i & 15 & -.5+7.5i & 3.5+3.5i \\ -12.5i & -.5-7.5i & 9 & 1.5-9.5i \\ 12-8i & 3.5-3.5i & 1.5+9.5i & 2 \end{pmatrix},$$

and for any $\lambda \in \sigma(A)$,

$$|\lambda| \leq \frac{R_A+T}{2} = 52.72, \qquad |Re\lambda| \leq R_B = 28.72, \quad \textit{and} \quad |Im\lambda| \leq R_C = 38.63.$$

This is graphed in figure 1.3. Once again, the bound on λ, represented by the circle in figure 1.3, may be ignored.

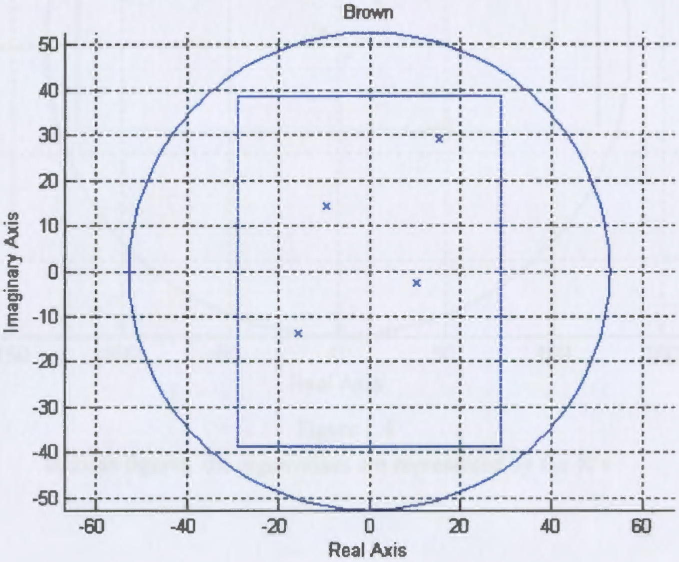


Figure 1.3

In these figures, the eigenvalues are represented by the X's

The matrix in the previous example was altered slightly to produce the matrix in the next example.

Example 1.4 Let

$$\mathbf{A} = \begin{pmatrix} 100 & 2 & -6 & 15i \\ 7-6i & 15i & -7+3i & 8+5i \\ 19 & 8-4i & 13+9i & 10 \\ 16+9i & 15-2i & -9+3i & 0 \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{97.33 + 2.07i, 17.51 + 20.11i, 5.14 - 4.56i, -7.0 + 6.38i\}.$$

Then for any $\lambda \in \sigma(A)$,

$$|\lambda| \leq \frac{R_A + T}{2} = 134.79, \qquad |Re\lambda| \leq R_B = 120.45, \quad \textit{and} \quad |Im\lambda| \leq R_C = 38.63.$$

This is graphed in figure 1.4. Once again, the rectangular box fits completely inside the circle. Therefore, the circle can be ignored.

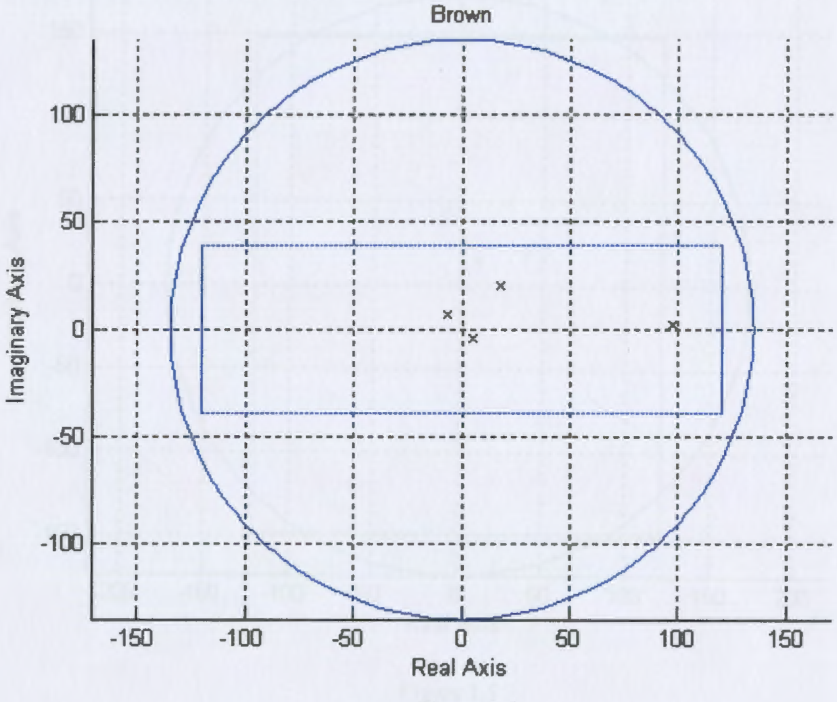


Figure 1.4

In these figures, the eigenvalues are represented by the X's

Example 1.5 Let

$$\mathbf{A} = \begin{pmatrix} 100 & 2 & -6 & 15i \\ 7-6i & 15i & -7+3i & 8+5i \\ 19 & 8-4i & 13+9i & 10 \\ 16+9i & 15-2i & -9+3i & -60-120i \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{98.87+1.12i, -60.76-121.3i, 3.32+9.03i, 11.56+15.16i\}.$$

Then for any $\lambda \in \sigma(A)$,

$$|\lambda| \le \frac{R_A+T}{2} = 172.87, \qquad |Re\lambda| \le R_B = 120.45, \quad and \quad |Im\lambda| \le R_C = 147.87.$$

This is graphed in figure 1.5. Notice that, this time, the circle intersects the rectangular box and, therefore, very slightly reduces the size of the inclusion set.

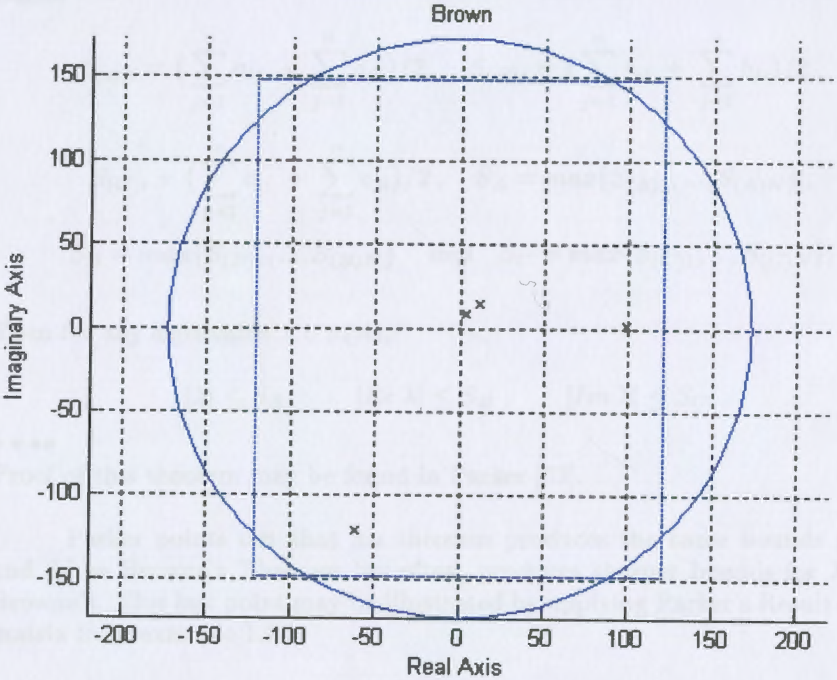


Figure 1.5

In these figures, the eigenvalues are represented by the X's

Section 1.1.2 Parker's First Theorem

In the last four examples, the circular bound on the eigenvalues either completely enclosed the rectangular box or intersected it only slightly. Therefore, the circular bound on λ provided little or no help in reducing the size of the inclusion set. In the 1930s and 1940s, methods were proposed to reduce the radius of that circular bound.

One of these methods was proposed by Parker [62] in 1937. Parker built on Browne's work by using the same matrices A,B, and C as Browne but summed both the rows and the columns of every matrix. The result was a rectangular box(the bound on the real and imaginary parts of the eigenvalues) of the same size as Browne but Parker's circle (the bound on λ) was equal to or smaller than Browne's.

The details of Parker's work are stated in the following Theorem.

Theorem 1.6 (Parker's First Theorem) Let $A \in C^{nxn}$.

Define

$$S_{(A)i} = (\sum_{j=1}^{n} a_{ij} + \sum_{j=1}^{n} a_{ji})/2\,, \quad S_{(B)i} = (\sum_{j=1}^{n} b_{ij} + \sum_{j=1}^{n} b_{ji})/2\,,$$

$$S_{(C)i} = (\sum_{j=1}^{n} c_{ij} + \sum_{j=1}^{n} c_{ji})/2\,, \quad S_A = max\{S_{(A)1}, ..., S_{(A)N}\}\,,$$

$$S_B = max\{S_{(B)1}, ..., S_{(B)N}\} \quad and \quad S_C = max\{S_{(C)1}, ..., S_{(C)N}\}.$$

Then for any eigenvalue $\lambda \in \sigma(A)$,

$$|\lambda| \leq S_A\,, \qquad |Re\,\lambda| \leq S_B\,, \qquad |Im\,\lambda| \leq S_C.$$

• • ••

Proof of this theorem may be found in Parker [62].

Parker points out that his theorem produces the same bounds for 'a' and 'b' as Browne's Theorem but,often, produces sharper bounds for λ than Browne's. This last point may be illustrated by applying Parker's Result to the matrix from example 1.2:

Example 1.7 Let

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & -1 & 2 \\ 1 & 2 & 1 & -1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & .5 & -1 \end{pmatrix}$$

(The solution details of this example may be found in the Appendix).

The spectrum, $\sigma(A)$, was previously found to be

$$\{-1.79, 2.17, .81 + 341i, .81 - 341i\}.$$

$$|\lambda| \le S_A = 4.25\,, \qquad |Re\ \lambda| \le S_B = 3.25\,, \quad and \quad |Im\ \lambda| \le S_C = 2.$$

Note that the bound on $|\lambda|$ is 4.25 for Parker, compared with 5 for Browne. However, in this example, the bounds on 'a' and 'b' (represented by the rectangle in figure 1.7) are contained completely within the bounds for λ (represented by the rectangle in figure 1.7). Therefore, in this example, the bound on λ is superfluous so that, Browne and Parker, for practical purposes, produce the same results.

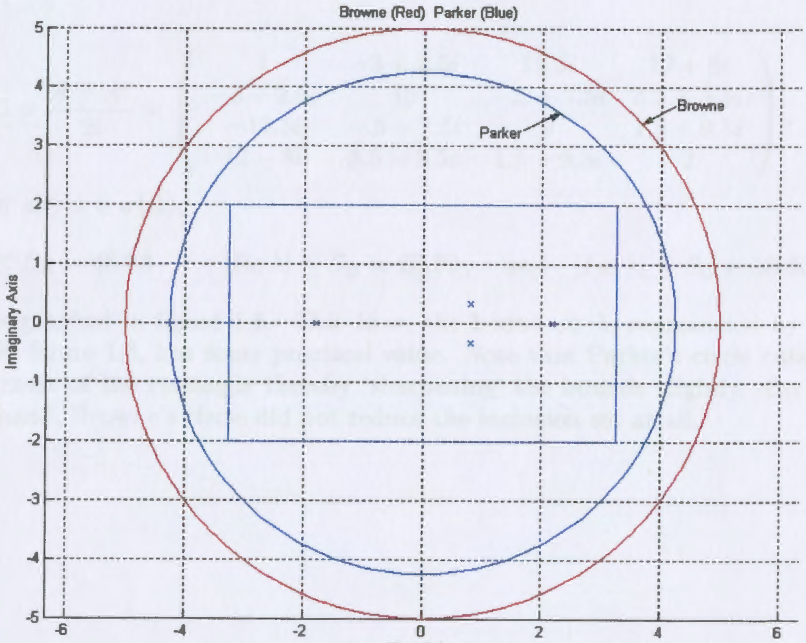


Figure 1.7
In these figures, the eigenvalues are represented by the X's

Recalculating Example 1.3 will produce practical, usable bounds that are different for Browne and Parker.

Example 1.3 is reworked as,
Example 1.8 Let

$$\mathbf{A} = \begin{pmatrix} -5+i & 2 & -6 & 15i \\ 7-6i & 15i & -7+3i & 8+5i \\ 19 & 8-4i & 13+9i & 10 \\ 16+9i & 15+2i & -9+3i & -7+2i \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{-15.42-13.78i, 15.37+29.14i, 10.33-2.72i, -9.28+14.36i\}.$$

Then

$$\mathbf{B} = \frac{A+A^*}{2} = \begin{pmatrix} -5 & 4.5+3i & 6.5 & 8+3i \\ 4.5-3i & 0 & .5+3.5i & 11.5+1.5i \\ 6.5 & .5-3.5i & 13 & .5-1.5i \\ 8-3i & 11.5-1.5i & .5+1.5i & -7 \end{pmatrix},$$

$$\mathbf{C} = \frac{A-A^*}{2i} = \begin{pmatrix} 1 & -3+2.5i & 12.5i & 12+8i \\ -3-2.5i & 15 & -.5+7.5i & 3.5+3.5i \\ -12.5i & -.5-7.5i & 9 & 1.5-9.5i \\ 12-8i & 3.5-3.5i & 1.5+9.5i & 2 \end{pmatrix},$$

and for any $\lambda \in \sigma(A)$,

$$|\lambda| \le S_A = 46.33\,, \qquad |Re\,\lambda| \le S_B = 28.72\,, \quad \textit{and} \quad |Im\,\lambda| \le S_C = 38.63.$$

This is graphed in figure 1.8. *This time*, the bound on λ, represented by the circle in figure 1.8, has some practical value. Note that Parker's circle cuts off the corners of the rectangle thereby 'sharpening' the bounds slightly. On the other hand, Browne's circle did not reduce the inclusion set at all.

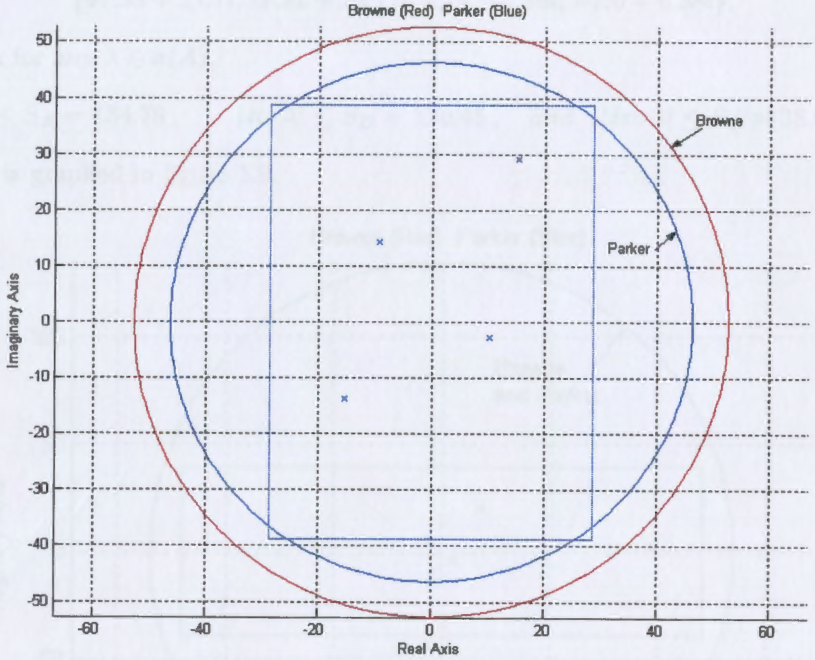


Figure 1.8

In these figures, the eigenvalues are represented by the X's

Revisiting Example 1.4,

Example 1.9 Let

$$\mathbf{A} = \begin{pmatrix} 100 & 2 & -6 & 15i \\ 7-6i & 15i & -7+3i & 8+5i \\ 19 & 8-4i & 13+9i & 10 \\ 16+9i & 15-2i & -9+3i & 0 \end{pmatrix}.$$

The spectrum, $\sigma(A)$, again is

$$\{97.33 + 2.07i, 17.51 + 20.11i, 5.14 - 4.56i, -7.0 + 6.38i\}.$$

Then for any $\lambda \in \sigma(A)$,

$$|\lambda| \leq S_A = 134.79\,, \qquad |Re\,\lambda| \leq S_B = 120.45\,, \quad and \quad |Im\,\lambda| \leq S_C = 38.63.$$

This is graphed in figure 1.9.

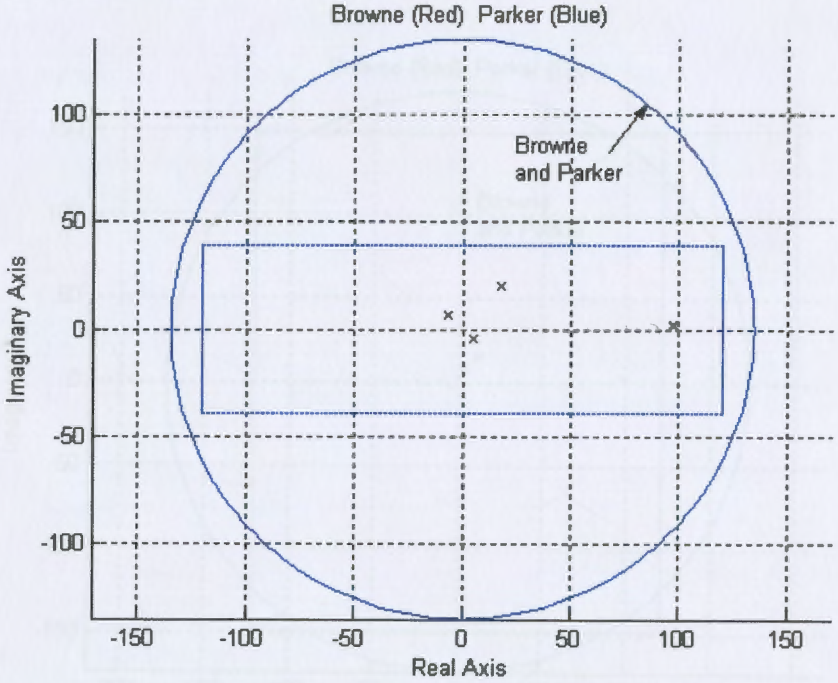


Figure 1.9

In these figures, the eigenvalues are represented by the X's

Revisiting Example 1.5,
Example 1.10 Let

$$\mathbf{A} = \begin{pmatrix} 100 & 2 & -6 & 15i \\ 7-6i & 15i & -7+3i & 8+5i \\ 19 & 8-4i & 13+9i & 10 \\ 16+9i & 15-2i & -9+3i & -60-120i \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{98.87 + 1.12i, -60.76 - 121.3i, 3.32 + 9.03i, 11.56 + 15.16i\}.$$

Then for any $\lambda \in \sigma(A)$,

$$|\lambda| \leq S_A = 172.87\,, \qquad |Re\ \lambda| \leq S_B = 120.45\,, \quad and \quad |Im\ \lambda| \leq S_C = 147.87.$$

This is graphed in figure 1.10. In this case, Parker and Browne produce the same inclusion set.

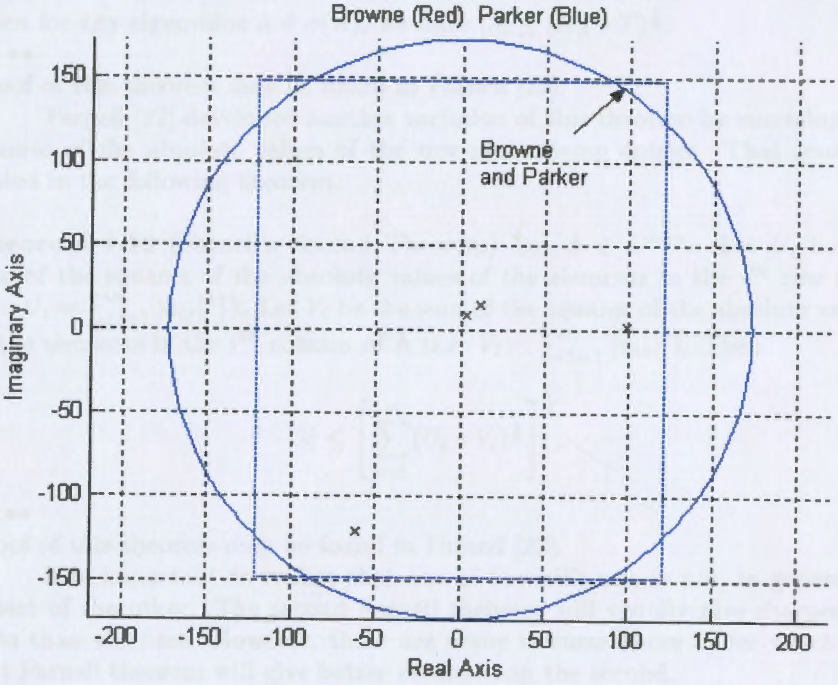


Figure 1.10

In these figures, the eigenvalues are represented by the X's

The preceding examples show that Parker's Theorem produces a small improvement over Browne in a limited number of cases. More substantial improvements were to come.

Section 1.1.3 Farnell and Brauer

In the 1940's, improvements were made to the bounds established by Browne and Parker. The first improvement was made by A.B Farnell [22] in 1944. Farnell established a bound on λ that was analogous to Browne's bound except that Farnell used the geometric mean instead of the arithmetic mean. Farnell's findings lead to the following theorem.

Theorem 1.11 (Farnell's First Theorem) Let $A \in C^{nxn}$. Let $R_{(A)i}$ be the sum of the absolute values of the elements in the i^{th} row of A. Let T_i be the sum of the absolute values of the elements in the i^{th} column of A. Define:

$$R_A = max\{R_{(A)1}, ..., R_{(A)N}\} \quad and \quad T = max\{T_1, ..., T_N\}.$$

Then for any eigenvalue $\lambda \in \sigma(A)$, we have $|\lambda| \leq (R_A * T)^{\frac{1}{2}}$.

• • ••

Proof of this theorem may be found in Farnell [22].

Farnell [22] developed another variation of this theorem by summing the squares of the absolute values of the row and column entries. That result is stated in the following theorem.

Theorem 1.12 (Farnell's Second Theorem) Let $A \in C^{nxn}$. Let U_i be the sum of the squares of the absolute values of the elements in the i^{th} row of A (i.e. $U_i = \sum_{j=1}^{n} |a_{ij}|^2\}$). Let V_i be the sum of the squares of the absolute values of the elements in the i^{th} column of A (i.e. $V_i = \sum_{k=1}^{n} |a_{ki}|^2$). Then

$$|\lambda| \leq \left[\sum_{i=1}^{n} (U_i * V_i)^{\frac{1}{2}}\right]^{\frac{1}{2}}.$$

• • ••

Proof of this theorem may be found in Farnell [22].

It is important to realize that one of Farnell's sets is not, in general, a subset of the other. The second Farnell theorem will usually give sharper results than the first. However, there are many circumstances under which the first Farnell theorem will give better results than the second.

In 1946, Alfred Brauer [5] developed a bound that will always be better than Farnell's first theorem.

Theorem 1.13 (Brauer's First Theorem) Let $A \in C^{nxn}$. Let $R_{(A)i}$ be the sum of the absolute values of the elements in the i^{th} row of A. Let T_i be the sum of the absolute values of the elements in the i^{th} column of A. Define:

$$R_A = max\{R_{(A)1}, ..., R_{(A)N}\} \quad and \quad T = max\{T_1, ..., T_N\} .$$

Then for any eigenvalue $\lambda \in \sigma(A)$, we have $|\lambda| \leq min\{R_A, T\}$.

● ● ●●

Proof of this theorem may be found in Brauer [5].

Section 1.1.4 Brauer's Power Method

In 1946, Brauer [5] considered taking the sums of the rows of *powers* of the original matrix A. It turns out that this is a rather 'sharp' method. However, since the method involves taking powers of the original matrix, round-off error will often rear its ugly head. In any case, the theorem is stated as follows:

Theorem 1.14 (Brauer's Power Method) Let $A \in C^{nxn}$. Let $R_{(A^{2^r})i}$ be the sum of the absolute values of the elements in the i^{th} row of A^{2^r} where r is natural number. This means that the actual powers on the matrix will be 2,4,8,16,32,64,...). Define:

$$R_{A^{2^r}} = max\{R_{(A^{2^r})1}, ..., R_{(A^{2^r})N}\}.$$

Then for any eigenvalue $\lambda \in \sigma(A)$, we have

$$|\lambda| \leq (R_{(A^{2^r})})^{\frac{1}{r}}.$$

● ● ●●

Proof of this theorem may be found in Brauer [5].
Suspecting the possibility of roundoff error, many runs were done for this thesis using Brauer's Power method. The following examples are a representative sampling of what was found:

Example 1.15 Let

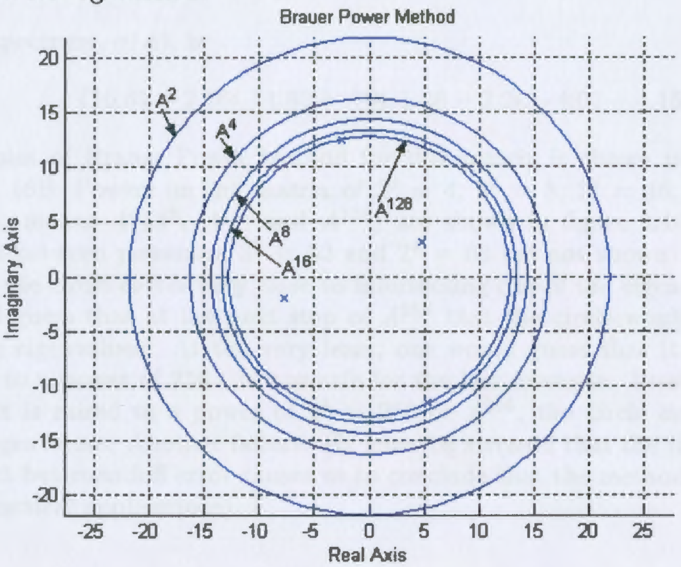
$$\mathbf{A} = \begin{pmatrix} 3i & 3-2i & -5 & -7+4i \\ 3-2i & -7i & 2-11i & -8 \\ -5 & 2-11i & 11i & 5+5i \\ -7+4i & -8 & 5+5i & -5i \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{-2.45+12.45i, 4.82+3.10i, -7.69-2.04i, 5.36-11.51i\}.$$

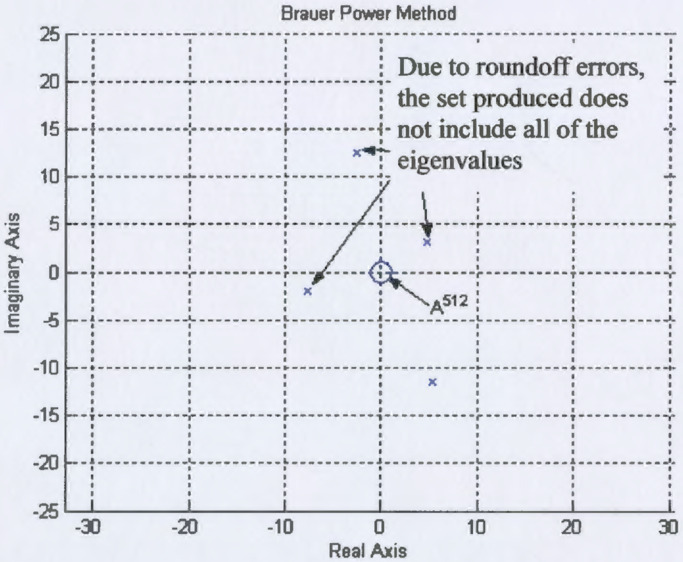
The plot of Brauer Power Method for this matrix is shown in figures 1.15A and 1.15B. Powers on the matrix of $2^1 = 2$, $2^2 = 4$, $2^3 = 8$, $2^4 = 16$, and $2^7 = 128$ (which means A^2,A^4,A^8, A^{16}, and A^{128}) are shown in figure 1.15A (for clarity, the in-between powers of $2^5 = 32$ and $2^6 = 64$ are not shown).

These were calculated in Matlab using single precision. Note that at A^{128}, the circle actually intersects one of the eigenvalues. When the matrix is raised to a power of $2^8 = 256$ or A^{256} (not shown in the figure), the circle is the same as A^{128}. This is good. However, at the next step, $2^9 = 512$, **round off error causes problems** as shown in figure 1.15B. At this power, the circle *does not* enclose the eigenvalues.



For clarity, powers of 32 and 64 are not shown. The power of 256 is the same as 128.

Figure 1.15A



When the matrix A is raised to a power of 512, roundoff error causes problems.

Figure 1.15B

In these figures, the eigenvalues are represented by the X's

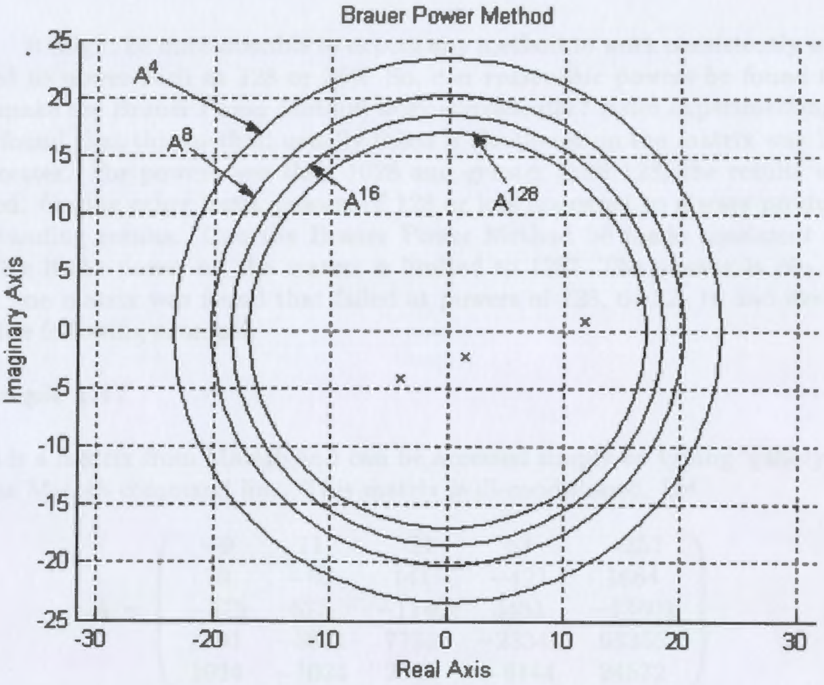
Example 1.16 Let

$$\mathbf{A} = \begin{pmatrix} 6 & 2 & 1 & -6 \\ 4 & -3i & 1 & 7i \\ 2+4i & 5i & 5 & 9 \\ 6 & 8 & 10 & 15 \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

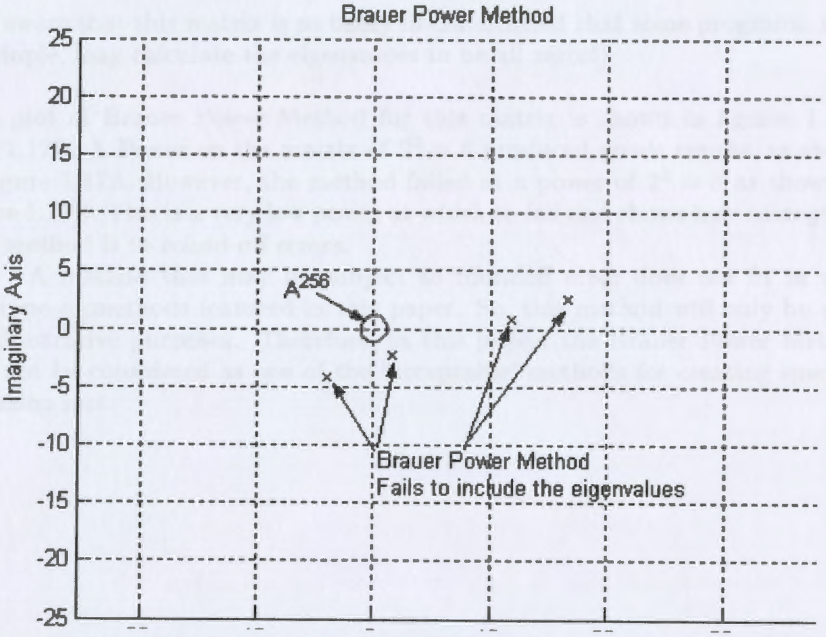
$$\{16.61 + 2.26i, 11.82 + .79i, 1.58 - 2.2i, -4.01 - 4.15i\}.$$

The plot of Brauer Power Method for this matrix is shown in figures 1.16A and 1.16B. Powers on the matrix of $2^2 = 4$, $2^3 = 8$, $2^4 = 16$, and $2^7 = 128$ (which means A^4,A^8, A^{16}, and A^{128}) are shown in figure 1.16A (for clarity, the in-between powers of $2^5 = 32$ and $2^6 = 64$ are not shown). Note that at A^{128}, the circle *comes very close* to intersecting one of the eigenvalues. So, one would guess that at the next step of A^{256} that the circle would intersect one of the eigenvalues. At the very least, one would guess that it would be safe to go to a power of 256 - it was safe for the last example. However, when the matrix is raised to a power of $2^8 = 256$ or A^{256}, the circle *does not* enclose the eigenvalues. Another failure. (It must be stressed that the theorem itself is correct but roundoff error causes us to conclude that the method is not reliable for practical applications).



For Clarity, powers of 32 and 64 are not shown.

Figure 1.16A



The Brauer Power Method fails when the matrix A is raised to a power of 256.

Figure 1.16B

In these figures, the eigenvalues are represented by the X's

It might be unreasonable to expect any method to work consistently when raised to power such as 128 or 256. So, can *reasonable* powers be found that will make the Brauer Power Method work consistently? After experimenting, it was found that this method usually failed if the power on the matrix was 1028 or greater. For powers less than 1028 and greater than 128, the results were mixed. On the other hand, powers of 128 or less appeared to always produced outstanding results. Can the Brauer Power Method be made consistent and reliable if the power on the matrix is limited to 128? The answer is **No.** At least one matrix was found that failed at powers of 128, 64 32, 16 and even 8. See the following example.

Example 1.17

This is a matrix from Matlab and can be accessed simply by typing 'gallery(5)' at the Matlab command line. This matrix is ill-conditioned. Let

$$\mathbf{A} = \begin{pmatrix} -9 & 11 & -21 & 63 & -252 \\ 70 & -69 & 141 & -421 & 1684 \\ -575 & 575 & -1149 & 3451 & -13801 \\ 3891 & -3891 & 7782 & -23345 & 93365 \\ 1024 & -1024 & 2048 & -6144 & 24572 \end{pmatrix}.$$

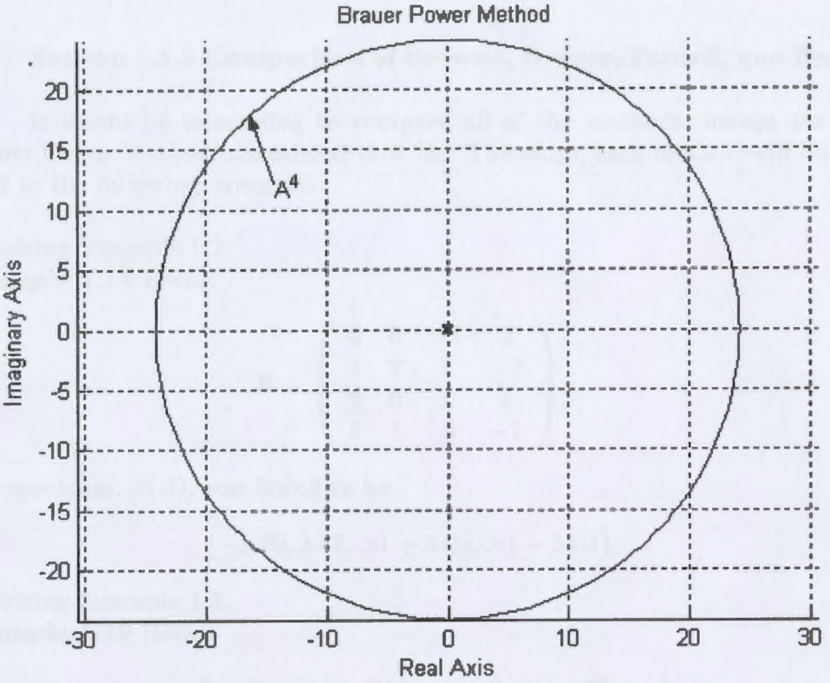
The spectrum, $\sigma(A)$, is

$$\{-.040844, -.011876{+}.038593i, -.011876{-}.038593i, .032298{+}.022998i, .032298{-}.022998i\}.$$

(Be aware that this matrix is so badly ill-conditioned that some programs, such as Maple, may calculate the eigenvalues to be all zeros!)

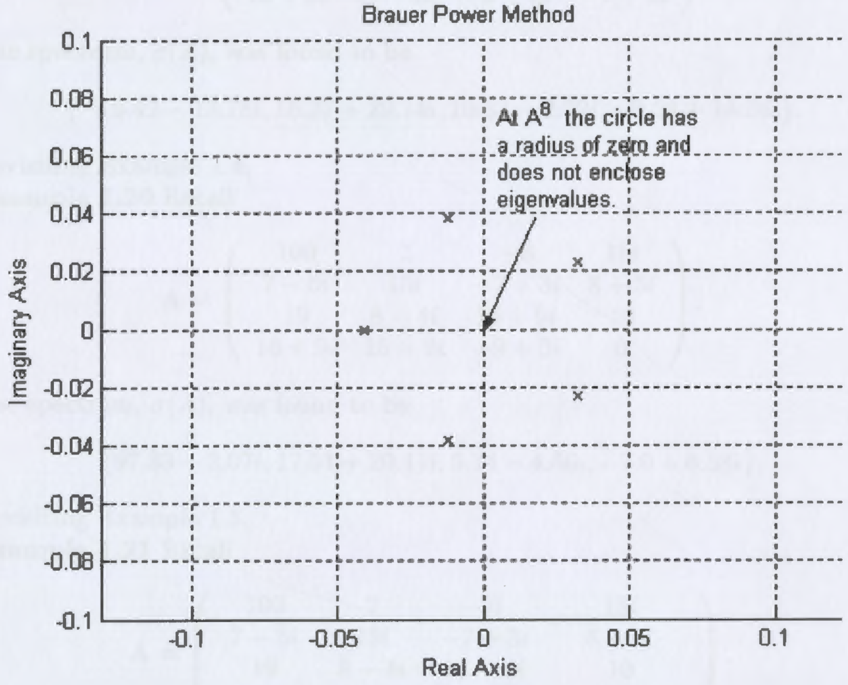
The plot of Brauer Power Method for this matrix is shown in figures 1.17A and 1.17B. A Power on the matrix of $2^2 = 4$ produced goods results, as shown in figure 1.17A. However, the method failed at a power of $2^3 = 8$ as shown in figure 1.17B. This is a very low power at which to fail and shows how susceptible this method is to round-off errors.

A method that may be subject to roundoff error does not fit in with the type of methods featured in this paper. So, this method will only be used for illustrative purposes. Therefore, in this paper, the Brauer Power Method will not be considered as one of the 'acceptable' methods for creating spectral inclusion sets.



The Brauer Power Method works for powers on the matrix A of 4

Figure 1.17A



At a power on the ma-

Figure 1.17B

In these figures, the eigenvalues are represented by the X's

Section 1.1.5 Comparison of Browne, Brauer, Farnell, and Parker

It would be interesting to compare all of the methods, except for the Brauer Power Method, considered thus far. Therefore, each method will be applied to the following examples.

Revisiting Example 1.2,
Example 1.18 Recall

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & -1 & 2 \\ 1 & 2 & 1 & -1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & .5 & -1 \end{pmatrix}.$$

The spectrum, $\sigma(A)$, was found to be

$$\{-1.79, 2.17, .81 + 341i, .81 - 341i\}.$$

Revisiting Example 1.3,
Example 1.19 Recall

$$\mathbf{A} = \begin{pmatrix} -5+i & 2 & -6 & 15i \\ 7-6i & 15i & -7+3i & 8+5i \\ 19 & 8-4i & 13+9i & 10 \\ 16+9i & 15+2i & -9+3i & -7+2i \end{pmatrix}.$$

The spectrum, $\sigma(A)$, was found to be

$$\{-15.42 - 13.78i, 15.37 + 29.14i, 10.33 - 2.72i, -9.28 + 14.36i\}.$$

Revisiting Example 1.4,
Example 1.20 Recall

$$\mathbf{A} = \begin{pmatrix} 100 & 2 & -6 & 15i \\ 7-6i & 15i & -7+3i & 8+5i \\ 19 & 8-4i & 13+9i & 10 \\ 16+9i & 15-2i & -9+3i & 0 \end{pmatrix}.$$

The spectrum, $\sigma(A)$, was found to be

$$\{97.33 + 2.07i, 17.51 + 20.11i, 5.14 - 4.56i, -7.0 + 6.38i\}.$$

Revisiting Example 1.5,
Example 1.21 Recall

$$\mathbf{A} = \begin{pmatrix} 100 & 2 & -6 & 15i \\ 7-6i & 15i & -7+3i & 8+5i \\ 19 & 8-4i & 13+9i & 10 \\ 16+9i & 15-2i & -9+3i & -60-120i \end{pmatrix}.$$

The spectrum, $\sigma(A)$, was found to be

$$\{98.87 + 1.12i, -60.76 - 121.3i, 3.32 + 9.03i, 11.56 + 15.16i\}.$$

The results of these last for examples are plotted in figures 1.18, 1.19, 1.20, and 1.21. The radii of the circular bounds on λ are listed below.

Example 1.18		**Example 1.19**	
Farnell's 2nd	4.049	Farnell's 2nd	46.286
Parker	4.25	Parker	46.335
Browne	5.000	Brauer	51.676
Brauer	5.000	Farnell's 1st	52.706
Farnell's 1st	5.000	Browne	52.716

Example 1.20		**Example 1.21**	
Farnell's 2nd	110.108	Brauer	168.598
Brauer	123.000	Farnell's 1st	172.817
Farnell's 1st	134.272	Browne	172.870
Browne	134.789	Parker	172.870
Parker	134.789	Farnell's 2nd	172.870

The results are rather fascinating. Notice that in examples 1.18,1.19, and 1.20, Farnell's Second Method produced the best results but in example 1.21, Farnell's Second Method was among the poorest. On the other hand, Brauer was among the worst in example 1.18, third best in 1.19, second best in 1.20, and best in 1.21. The other methods produced mixed results even though it can be said that Browne, the oldest method considered, usually produced poorer results than the others.

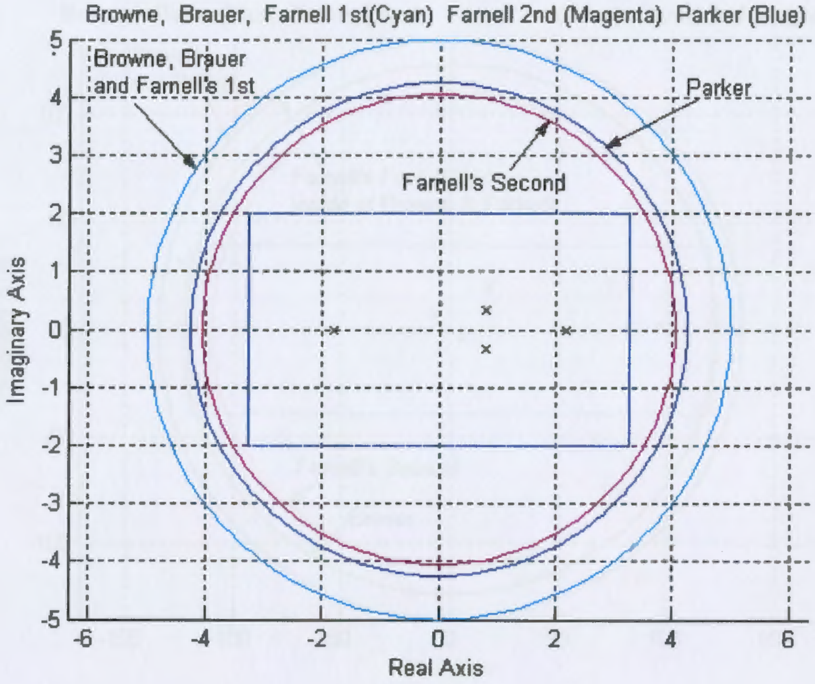


Figure 1.18

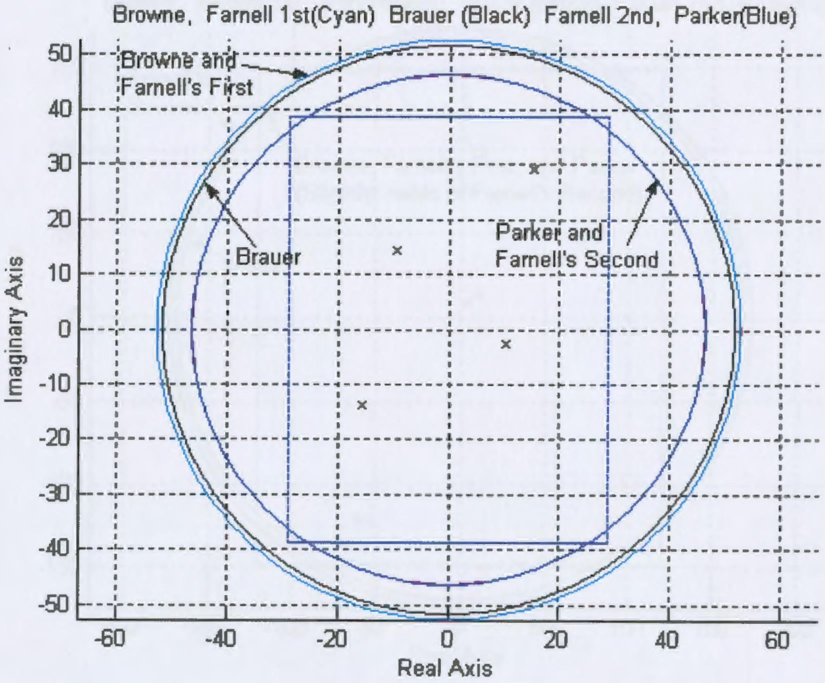


Figure 1.19

In these figures, the eigenvalues are represented by the X's

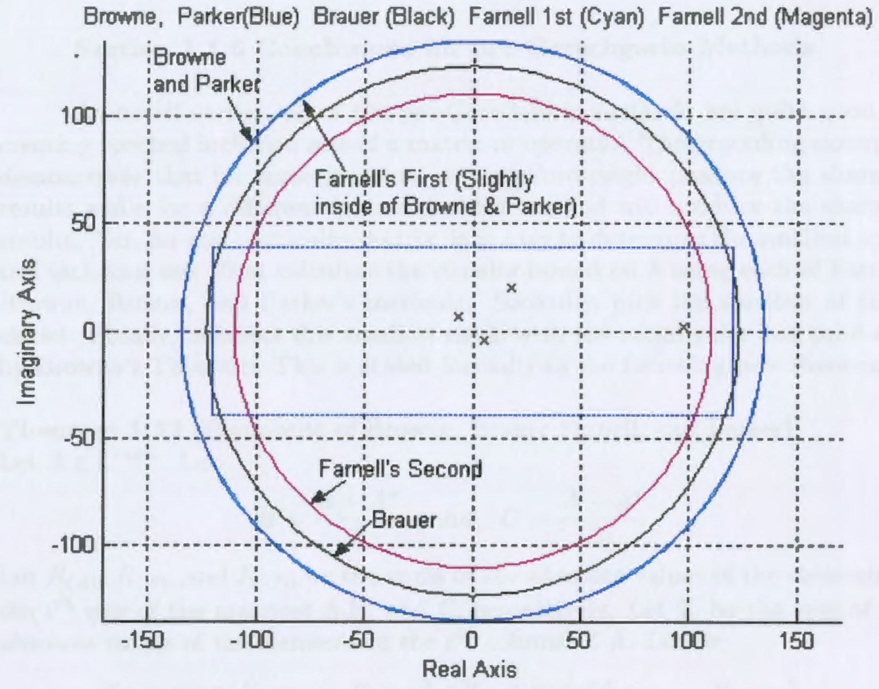


Figure 1.20

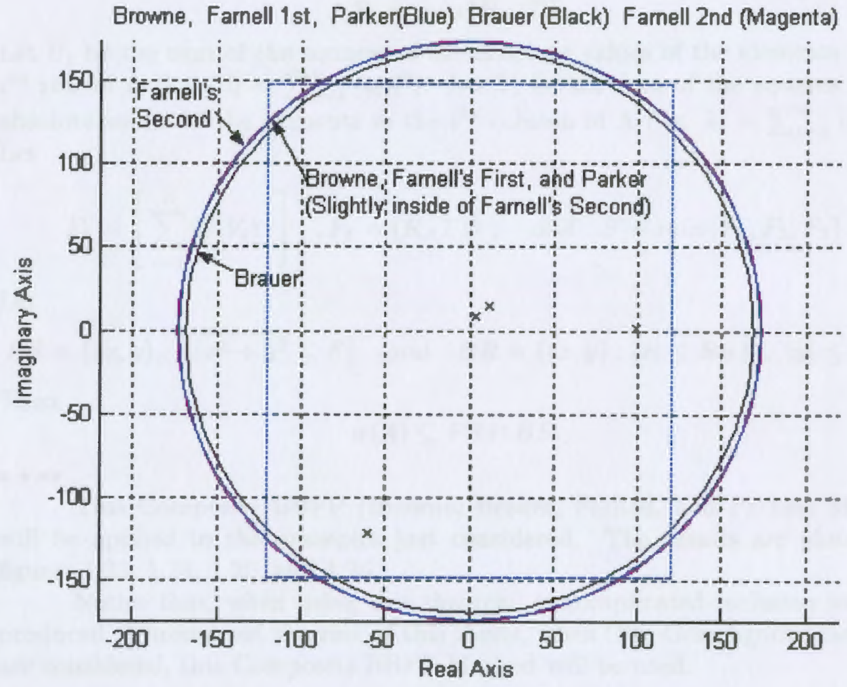


Figure 1.21

In these figures, the eigenvalues are represented by the X's

Section 1.1.6 Conclusions for pre-Gerschgorin Methods

As noted earlier, all of the pre-Gerschgorin methods are quite good for creating spectral inclusion sets of a matrix or operator. The preceding examples demonstrate that for some matrices, one method might produce the sharpest results while for a different matrix another method will produce the sharpest results. Yet, for any particular matrix, it is easy to determine the smallest spectral inclusion set: First calculate the circular bound on λ using each of Farnell, Browne, Brauer, and Parker's methods. Secondly, pick the smallest of these circles. Finally, intersect this smallest circle with the rectangular box produced by Browne's Theorem. This is stated formally in the following new theorem.

Theorem 1.22 (Composite of Browne, Brauer Farnell, and Parker)
Let $A \in C^{nxn}$. Let

$$B = \frac{A + A^*}{2} \quad and \quad C = \frac{A - A^*}{2i}.$$

Let $R_{(A)i}$,$R_{(B)i}$,and $R_{(C)i}$,be the sums of the absolute values of the elements in the i^{th} row of the matrices A,B, and C, respectively. Let T_i be the sum of the absolute values of the elements in the i^{th} column of A. Define:

$$R_A = max\{R_{(A)1}, ..., R_{(A)N}\}, \; R_B = max\{R_{(B)1}, ..., R_{(B)N}\},$$
$$R_C = max\{R_{(C)1}, ..., R_{(C)N}\}, \quad and \quad T = max\{T_1, ..., T_N\}.$$

Let

$$F_1 = min\{R_A, T\}.$$

Let U_i be the sum of the squares of the absolute values of the elements in the i^{th} row of A (i.e. $U_i = \sum_{j=1}^{n} |a_{ij}|^2$). Let V_i be the sum of the squares of the absolute values of the elements in the i^{th} column of A (i.e. $V_i = \sum_{k=1}^{n} |a_{ki}|^2$). Let

$$F_2 = \left[\sum_{i=1}^{n} (U_i V_i)^{\frac{1}{2}} \right]^{\frac{1}{2}}, F_3 = (R_A T)^{\frac{1}{2}}, \quad and \quad F = min\{F_1, F_2, F_3\}.$$

Let

$$FS = \{(x,y) : \sqrt{x^2 + y^2} \leq F\} \quad and \quad BR = \{(x,y) : |x| \leq R_B \mid :, \; |y| \leq R_C\}.$$

Then

$$\sigma(A) \subseteq FS \cap BR.$$

• • ••

This Composite BBFP (Browne, Brauer, Farnell, and Parker) Method will be applied to the examples just considered. The results are plotted in figures 1.23, 1.24, 1.25, and 1.26.

Notice that, when using this theorem, uncomplicated inclusion sets are produced. Throughout the rest of this thesis, when 'Pre-Gerschgorin' methods are considered, this Composite BBFP Method will be used.

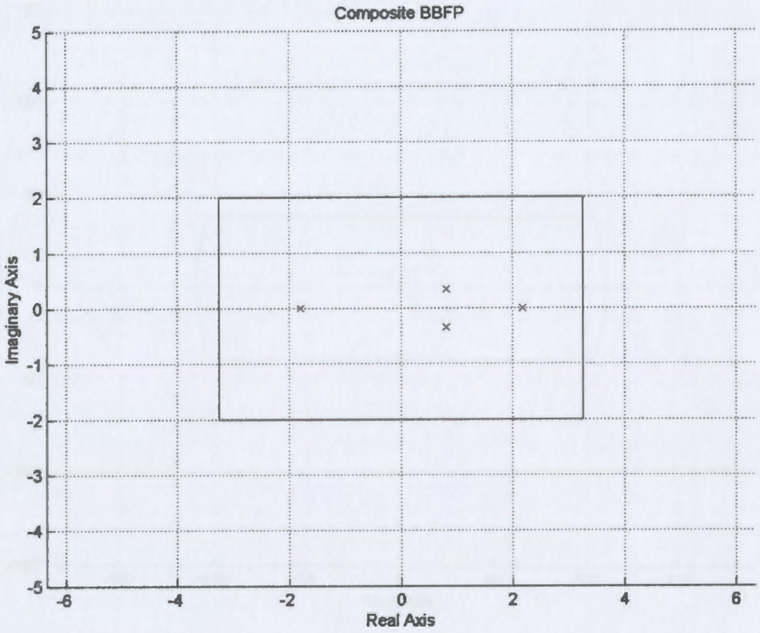


Figure 1.23

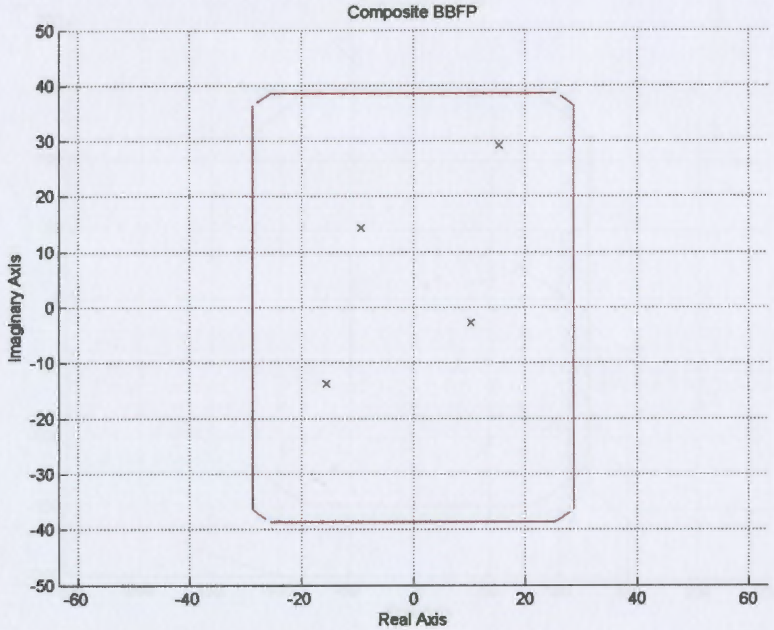


Figure 1.24

In these figures, the eigenvalues are represented by the X's

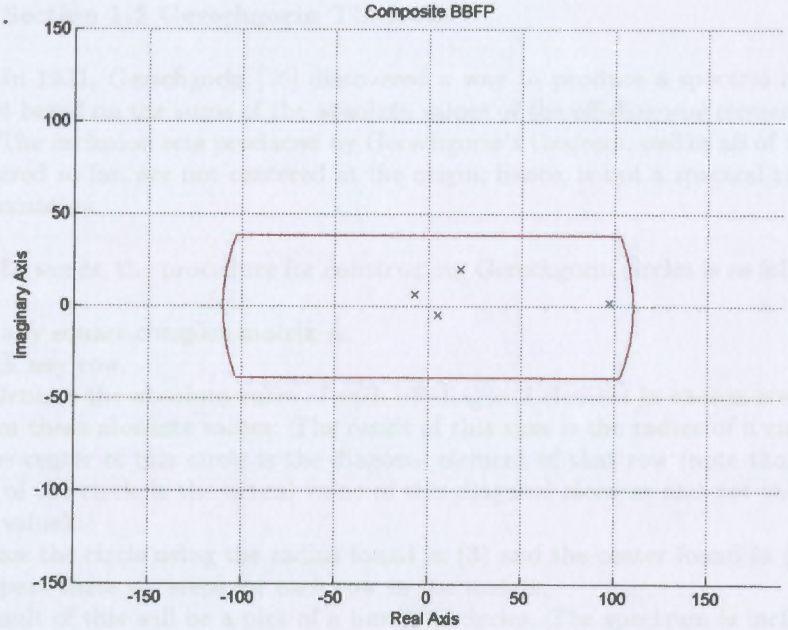


Figure 1.25

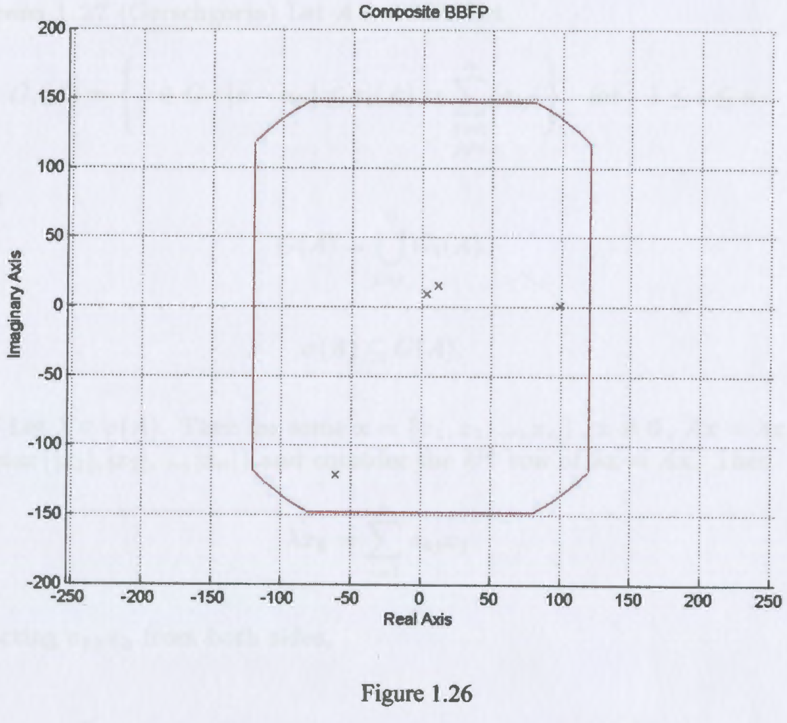


Figure 1.26

In these figures, the eigenvalues are represented by the X's

Section 1.2 Gerschgorin Theorem

In 1931, Gerschgorin [26] discovered a way to produce a spectral inclusion set based on the sums of the absolute values of the off-diagonal elements of rows. The inclusion sets produced by Gerschgorin's theorem, unlike all of those considered so far, are not centered at the origin; hence, is not a spectral radius approximation.

In words, the procedure for constructing Gerschgorin circles is as follows:

Given any square,complex matrix A
(1) Pick any row.
(2) Calculate the absolute value of each off-diagonal element in chosen row.
(3) Sum these absolute values. The result of this sum is the radius of a circle.
(4) The center of this circle is the diagonal element of that row (note that the center of the circle is the actual value of this diagonal element and not the absolute value).
(5) Draw the circle using the radius found in (3) and the center found in (4)
(6) Repeat these six steps for each row in the matrix.
The result of this will be a plot of a bunch of circles. The spectrum is included in the *union of these circles.*

A formal statement of Gerschgorin's Theorem is as follows:

Theorem 1.27 (Gerschgorin) Let $A \in C^{nxn}$. Let

$$G_i(A) = \left\{ z \in C : |z - a_{ii}| \leq r_i(A) = \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| \right\} \quad \text{for} \quad 1 \leq i \leq n$$

and let

$$G(A) = \bigcup_{i=1}^{n} G_i(A).$$

Then

$$\sigma(A) \subseteq G(A).$$

Proof Let $\lambda \in \sigma(A)$. Then for some $\mathbf{x} = \{x_1, x_2, ..., x_n\}$, $\mathbf{x} \neq \mathbf{0}$, $A\mathbf{x} = \lambda\mathbf{x}$. Set $x_k = max\{|x_1|, |x_2|, ..., |x_n|\}$ and consider the k^{th} row of $\lambda\mathbf{x} = A\mathbf{x}$. Then

$$\lambda x_k = \sum_{j=1}^{n} a_{kj} x_j.$$

Subtracting $a_{kk}x_k$ from both sides,

$$\lambda x_k - a_{kk}x_k = \sum_{\substack{j=1\\ j\neq k}}^{n} a_{kj}x_j,$$

$$(\lambda - a_{kk})x_k = \sum_{\substack{j=1\\ j\neq k}}^{n} a_{kj}x_j.$$

Taking the absolute value of both sides and using the triangle inequality, we have:

$$|(\lambda - a_{kk})x_k| = \left|\sum_{\substack{j=1\\ j\neq k}}^{n} a_{kj}x_j\right| \leq \sum_{\substack{j=1\\ j\neq k}}^{n} |a_{kj}x_j|,$$

and thus

$$|(\lambda - a_{kk})||x_k| \leq \sum_{\substack{j=1\\ j\neq k}}^{n} |a_{kj}||x_j|.$$

Since $|x_k| \geq |x_j|$,

$$|(\lambda - a_{kk})||x_k| \leq \sum_{\substack{j=1\\ j\neq k}}^{n} |a_{kj}||x_j| \leq \sum_{\substack{j=1\\ j\neq k}}^{n} |a_{kj}||x_k|$$

Dividing both sides by $|x_k|$, we see that

$$|\lambda - a_{kk}| \leq \sum_{\substack{j=1\\ j\neq k}}^{n} |a_{kj}|.$$

● ● ●●

Revisiting Example 1.2,
Example 1.28 Find the Gerschgorin radii and disks for the following matrix:

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & -1 & 2 \\ 1 & 2 & 1 & -1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & .5 & -1 \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{-1.79, 2.17, .81 + 341i, .81 - 341i\}.$$

(The solution details of this example may be found in the Appendix).

The Gerschgorin disks for this matrix are plotted in figure 1.28.

● ● ●●

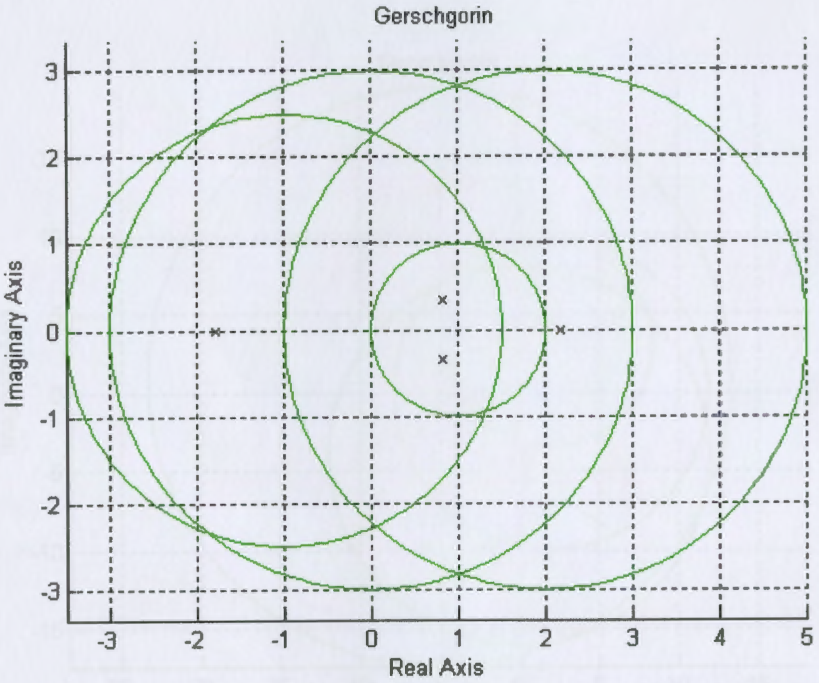


Figure 1.28

In these figures, the eigenvalues are represented by the X's

The next example uses a complex matrix

Example 1.29 Find the Gerschgorin radii and disks for the following matrix:

$$\mathbf{A} = \begin{pmatrix} 2+3i & i & 4 & i+1 \\ 4-4i & 2 & 1+i & 2+2i \\ 3i & 4 & -4i & 5i \\ -7 & 2-5i & 6 & -5+i \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{7.79 - .12i, .97 + 6.19i, -5.47 - 5.89i, -4.29 - .18i\}.$$

(The solution details of this example may be found in the Appendix).

The Gerschgorin disks for this matrix are plotted in figure 1.29.

● ● ●●

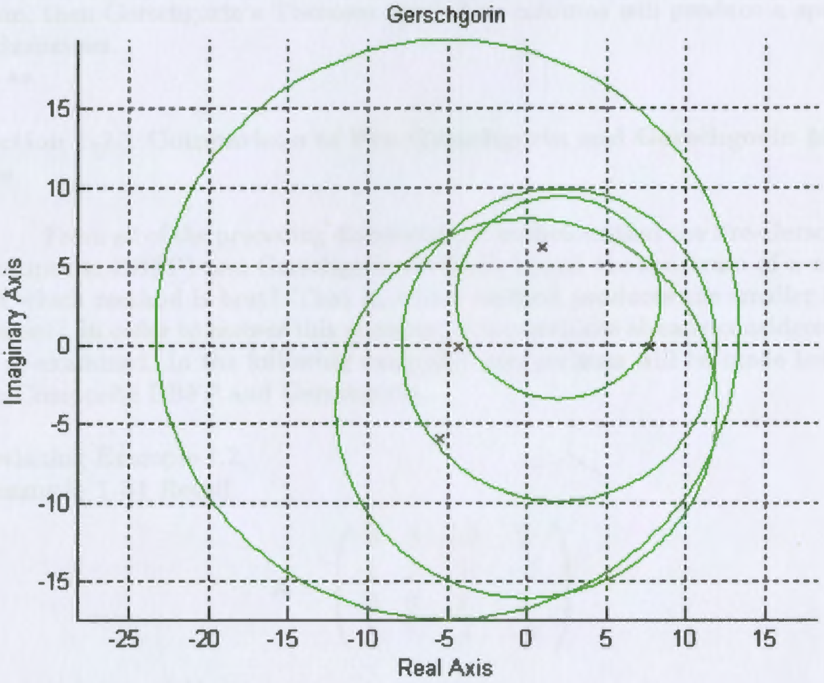


Figure 1.29

In these figures, the eigenvalues are represented by the X's

A spectral inclusion set can be produced by using Gerschgorin's theorem based on column sums rather than row sums. The theorem related to column sums is stated as follows:

Theorem 1.30 (Gerschgorin Column Theorem) Let $A \in C^{nxn}$. Let

$$G_j^{\mathbf{T}}(A) = \{z \in C : |z - a_{jj}| \leq r_j(A) = \sum_{\substack{i=1 \\ i \neq j}}^{n} |a_{ij}|\} \qquad 1 \leq j \leq n$$

and let

$$G^{\mathbf{T}}(A) = \bigcup_{j=1}^{n} G_j^{\mathbf{T}}(A).$$

Then

$$\sigma(A) \subseteq G^{\mathbf{T}}(A).$$

Proof Gerschgorin's Column Theorem uses column elements instead of row elements. Therefore, it is equivalent to applying Gerschgorin's original theorem to the transpose of a matrix. Since the transpose of a matrix preserves the spectrum, then Gerschgorin's Theorem applied to columns will produce a spectral inclusion set.

● ● ●●

Section 1.2.1 Comparison of Pre-Gerschgorin and Gerschgorin Methods

From all of the preceding discussions, it is obvious that the Pre-Gerschgorin (Composite BBFP) and Gerschgorin methods bound the spectrum of a matrix but which method is best? That is, which method produces the smaller inclusion set? In order to answer this question, some matrices already considered will be re-examined. In the following examples, comparisons will be made between the Composite BBFP and Gerschgorin.

Revisiting Example 1.2,
Example 1.31 Recall

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & -1 & 2 \\ 1 & 2 & 1 & -1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & .5 & -1 \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{-1.79, 2.17, .81 + 341i, .81 - 341i\}.$$

The Gerschgorin, Gerschgorin Column, and the Composite BBFP sets are shown in figure 1.31A and 1.31B. Notice that the area enclosed by the Gerschgorin circles is someone larger than the area enclosed by the Composite BBFP rectangle.

In fact, the Composite BBFP is almost a subset of Gerschgorin. In this case, Gerschgorin and Gerschgorin Column looks rather crude compared to the Pre-Gerschgorin methods.

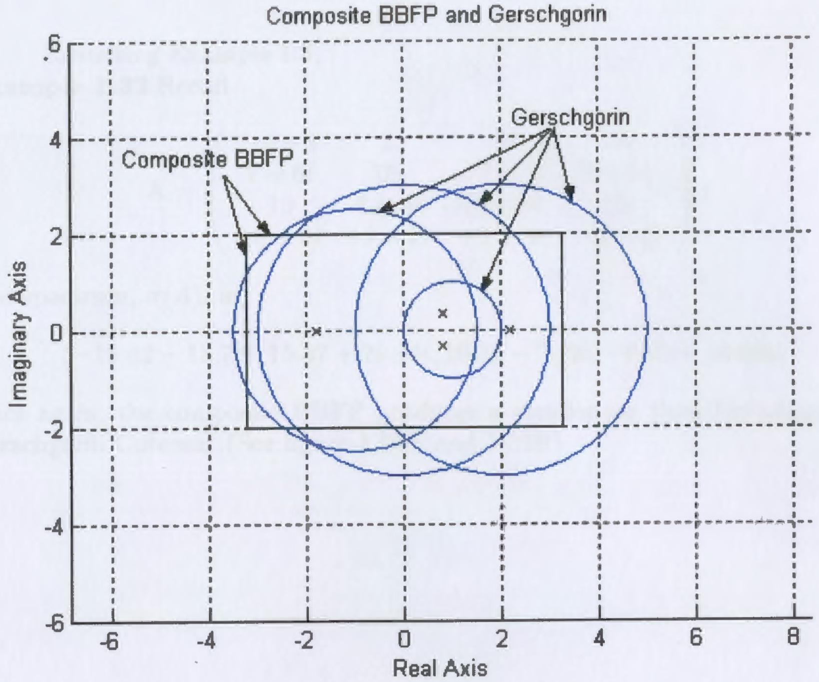


Figure 1.31A

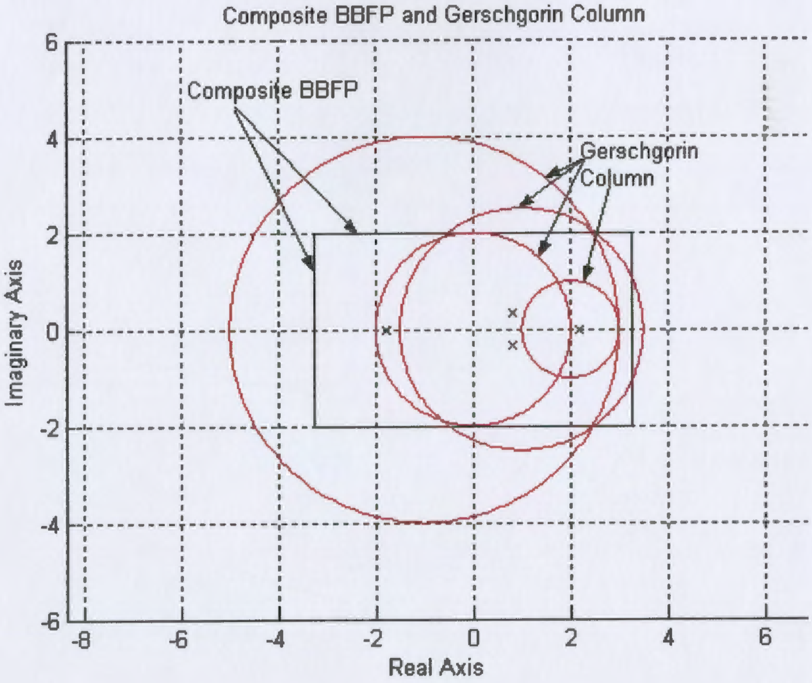


Figure 1.31B

In these figures, the eigenvalues are represented by the X's

Revisiting Example 1.3,

Example 1.32 Recall

$$\mathbf{A} = \begin{pmatrix} -5+i & 2 & -6 & 15i \\ 7-6i & 15i & -7+3i & 8+5i \\ 19 & 8-4i & 13+9i & 10 \\ 16+9i & 15+2i & -9+3i & -7+2i \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{-15.42 - 13.78i, 15.37 + 29.14i, 10.33 - 2.72i, -9.28 + 14.36i\}.$$

Once again, the composite BBFP produces a smaller set than Gerschgorin or Gerschgorin Column. (See figure 1.32A and 1.32B)

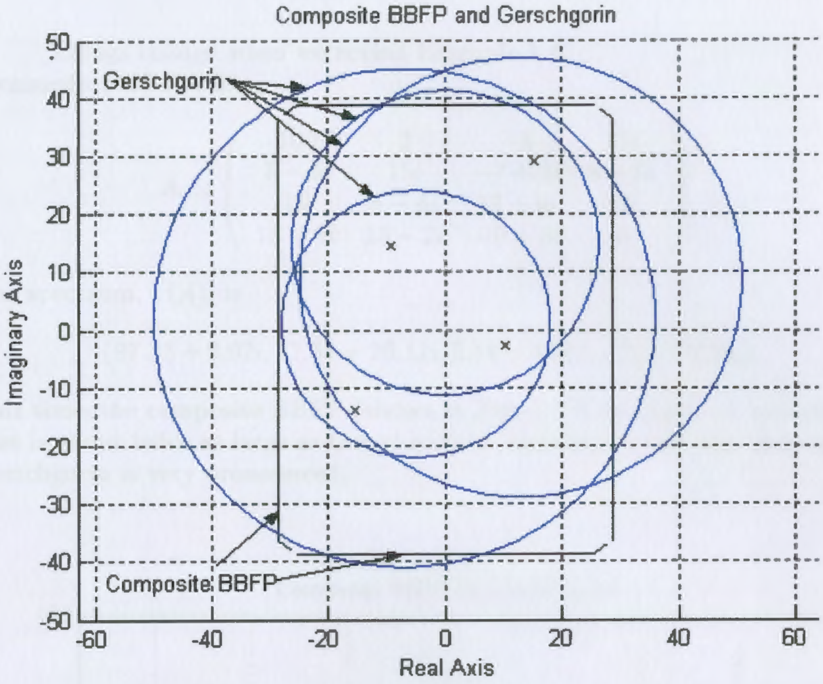


Figure 1.32A

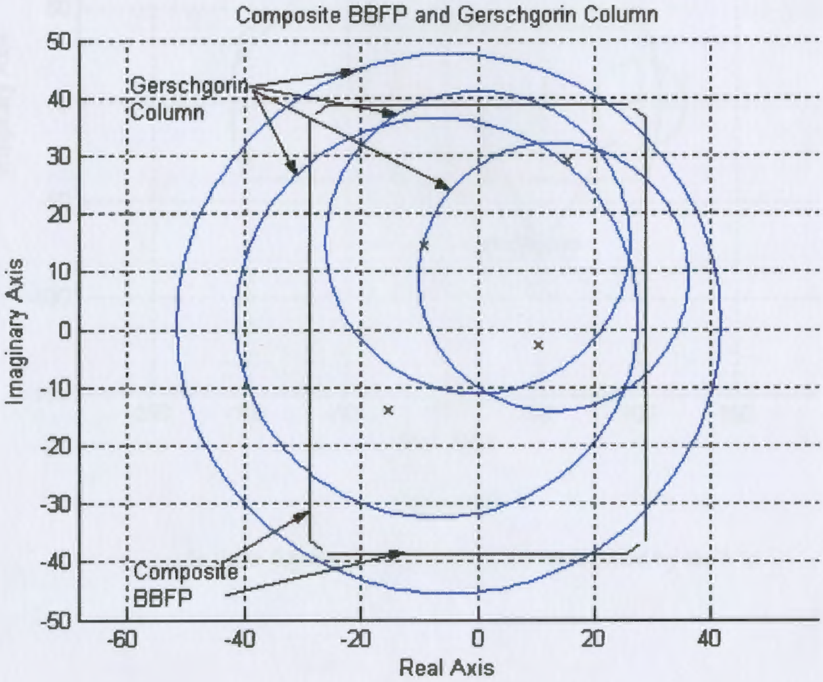


Figure 1.32B

In these figures, the eigenvalues are represented by the X's

Things change when we revisit Example 1.4,
Example 1.33 Recall

$$\mathbf{A} = \begin{pmatrix} 100 & 2 & -6 & 15i \\ 7-6i & 15i & -7+3i & 8+5i \\ 19 & 8-4i & 13+9i & 10 \\ 16+9i & 15-2i & -9+3i & 0 \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{97.33 + 2.07i, 17.51 + 20.11i, 5.14 - 4.56i, -7.0 + 6.38i\}.$$

This time, the composite BBFP (shown in Fig. 1.33) produces an inclusion set that is about twice as large as Gerschgorin's. In this example, the 'economy' of Gerschgorin is very pronounced.

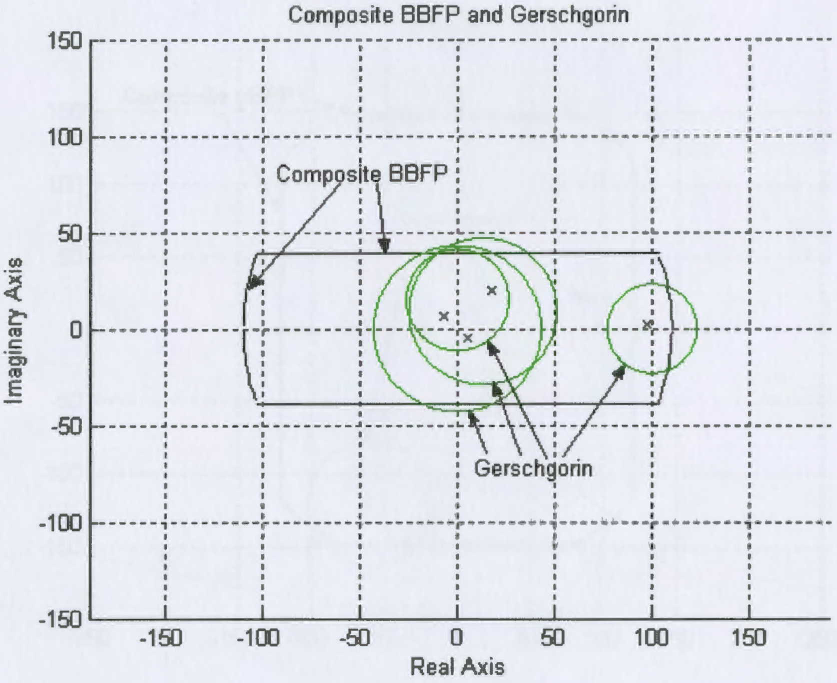


Figure 1.33

In these figures, the eigenvalues are represented by the X's

The Gerschgorin disks are even more impressive when we revisit Example 1.5,

Example 1.34 Recall

$$\mathbf{A} = \begin{pmatrix} 100 & 2 & -6 & 15i \\ 7-6i & 15i & -7+3i & 8+5i \\ 19 & 8-4i & 13+9i & 10 \\ 16+9i & 15-2i & -9+3i & -60-120i \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{98.87 + 1.12i, -60.76 - 121.3i, 3.32 + 9.03i, 11.56 + 15.16i\}.$$

The results, shown in figure 1.34, reveal that the composite BBFP set is more than three times larger than Gerschgorin's. In this example, the Pre-Gerschgorin methods look crude.

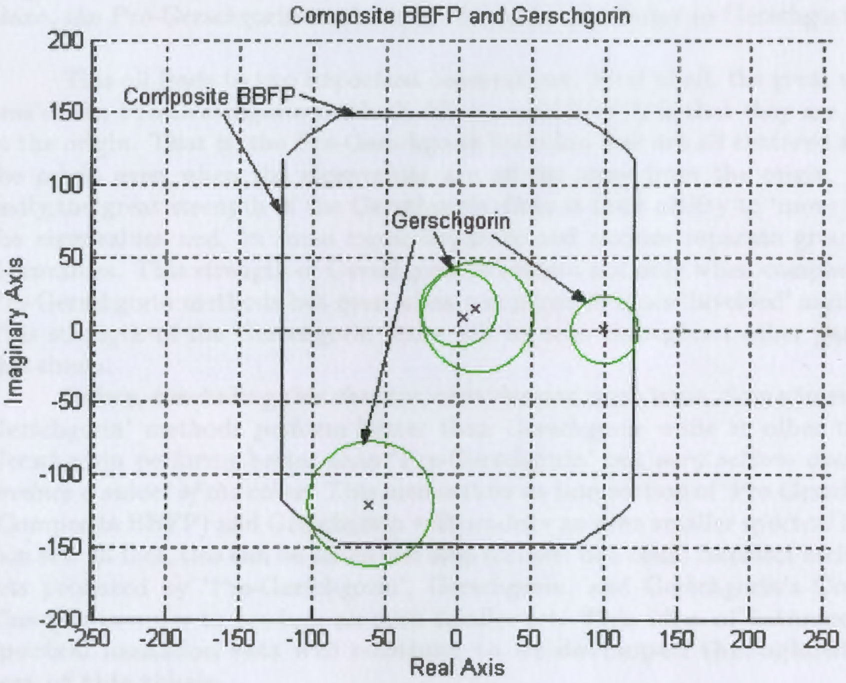


Figure 1.34

In these figures, the eigenvalues are represented by the X's

What is going on here? In these examples, both sides appear - cases in which the Composite BBFP is superior to Gerschgorin and cases in which Gerschgorin is much better than the Composite BBFP. Is there a way to predict the best method for a particular matrix? If there is a way to predict it, how is it done? The short answers are 'yes, there is a way' and it is done by first calculating the Gerschgorin disks and considering the following:

(1) If the Gerschgorin disks form one connected set that includes the origin,such as examples 1.31 and 1.32, then the Pre-Gerschgorin (Composite BBFP) methods will perform very well relative to Gerschgorin.

(2) If the Gerschgorin disks consist of separate groups of disks and the distance between the groups is relatively large, such as examples 1.33 and 1.34, the Pre-Gerschgorin methods are likely to be inferior to Gerschgorin.

(3) If the Gerschgorin disks form one connected set that is far from the origin or the Gerschgorin disks are confined to one quadrant in the complex plane, the Pre-Gerschgorin methods are likely to be inferior to Gerschgorin.

This all leads to two important observations. First of all, the great weakness of the Pre-Gerschgorin methods (Composite BBFP) is that they are 'tied' to the origin. That is, the Pre-Gerschgorin inclusion sets are all centered about the origin even when the eigenvalues are all far away from the origin. Secondly,the great strength of the Gerschgorin disks is their ability to 'move with' the eigenvalues and, in some cases, separate and enclose separate groups of eigenvalues. This strength of Gerschgorin is evident not only when compared to Pre-Gerschgorin methods but even when compared to more 'involved' methods. This strength of the Gerschgorin disks will be seen throughout other parts of this thesis.

Before, concluding this chapter, consider one more issue. Sometimes Pre-Gerschgorin' methods perform better than Gerschgorin while at other times, Gerschgorin performs better than 'Pre-Gerschgorin' but *very seldom does one produce a subset of the other.* This means than an intersection of 'Pre-Gerschgorin' (Composite BBFP) and Gerschgorin will produce an even smaller spectral inclusion set. In fact, this can be taken one step farther: one could intersect inclusion sets produced by 'Pre-Gerschgorin', Gerschgorin, and Gerschgorin's Column Theorem in order to produce an even smaller set. **This idea of intersecting spectral inclusion sets will continue to be developed throughout the rest of this thesis.**

2 Improvements to the Pre-Gerschgorin Methods

As noted in chapter one, the Pre-Gerschgorin methods produced spectral inclusion sets centered at the origin. In many cases, this did not cause the inclusion set to be too large but in some cases, the spectral inclusion sets generated by Pre-Gerschgorin methods were excessively large simply because the sets were 'tied' so closely to the origin. Consider the following example.

Example 2.1

$$\mathbf{A} = \begin{pmatrix} 8+8i & 2 & 1 & 2 \\ 4 & 7+7i & 1 & 3 \\ 3 & 3 & 10+4i & 2 \\ 1 & -2 & 2 & 2+9i \end{pmatrix}$$

The spectrum, $\sigma(A)$, is

$$\{12.54+5.84i, 7.39+5.83i, 4.74+6.90i, 2.34+9.44i\}$$

When the best of the Pre-Gerschgorin methods, the Composite BBFP formulated in chapter one, is applied to the matrix A, the inclusion set shown in figure 2.1 is produced.

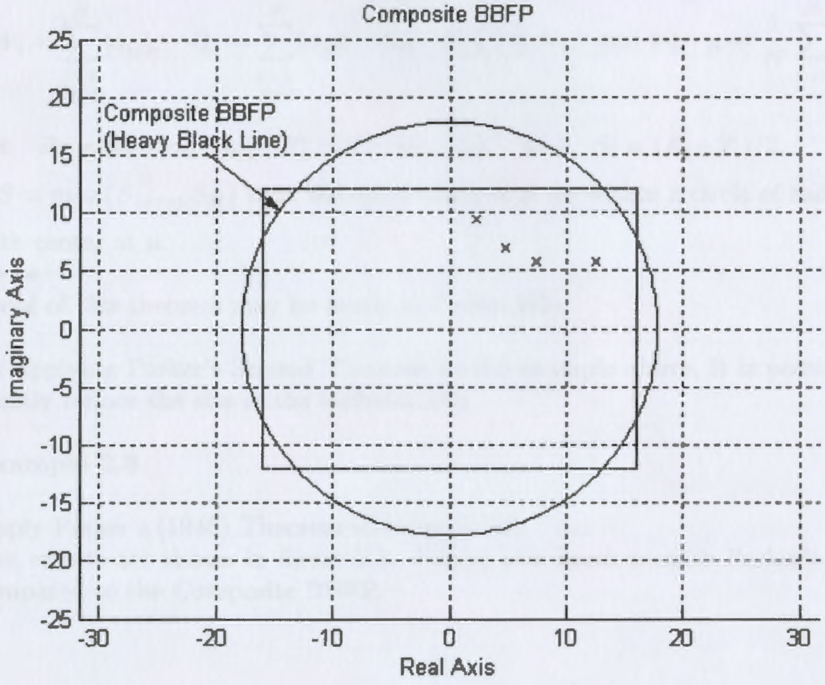


Figure 2.1

Notice that, even though the eigenvalues are all in the first quadrant, the Composite BBFP sweeps a large radius through all four quadrants in order to create the inclusion set. Again, this is due to the fact that the Pre-Gerschgorin inclusion sets are centered around the origin and are bounds for the spectral radius. On the other hand, it was noted in chapter one that one of the great advantages of Gerschgorin's Theorem is that its inclusion set is not centered at the origin but is centered around the diagonal elements of the matrix. What if one could somehow combine the best ideas from the Pre-Gerschgorin methods with Gerschgorin's idea of centering the inclusion set around diagonal elements of the matrix. Would this produce a smaller inclusion set?

Parker's Second Theorem (1948)

W.V. Parker is, apparently, the first person to adapt the Pre-Gerschgorin methods so that they might be centered at places other than the origin. By 1948, thanks to some papers written by Brauer, the Gerschgorin Theorem began to be widely known. In 1948, Parker [63] took Gerschgorin's idea of using the diagonal elements as centers of circles and applied this idea to the Pre-Gerschgorin methods. Unlike Gerschgorin, however, Parker took the *average* value of all of the diagonal elements and used this average value as the center of one big circle. At the same time, Parker retained some of the Pre-Gerschgorin ideas. Parker's Second Theorem is stated as follows:
Theorem 2.2 (Parker's Second Theorem (1948)) Let $A \in C^{NxN}$. Let

$$W_i = \sum_{\substack{j=1\\ j\neq i}}^{N} |a_{ij}|, \quad Q_i = \sum_{\substack{j=1\\ j\neq i}}^{N} |a_{ji}|, \quad \text{for} \quad 1 \leq i \leq N, \quad \text{and let} \quad \mu = \frac{1}{N}\sum_{i=1}^{N} a_{ii}.$$

Set $R_i = W_i + |a_{ii} - \mu|, \quad T_i = Q_i + |a_{ii} - \mu|, \quad$ and $\quad S_i = (R_i + T_i)/2.$

If $S = max\{S_1, ..., S_N\}$ then the eigenvalues of A lie within a circle of radius S, with center at μ.

• • ••

Proof of this theorem may be found in Parker [63].

By applying Parker's Second Theorem to the example above, it is possible to greatly reduce the size of the inclusion set:

Example 2.3

Apply Parker's (1948) Theorem to example 2.1.
The results are shown in figure 2.3. Notice how much smaller Parker's set is compared to the Composite BBFP.

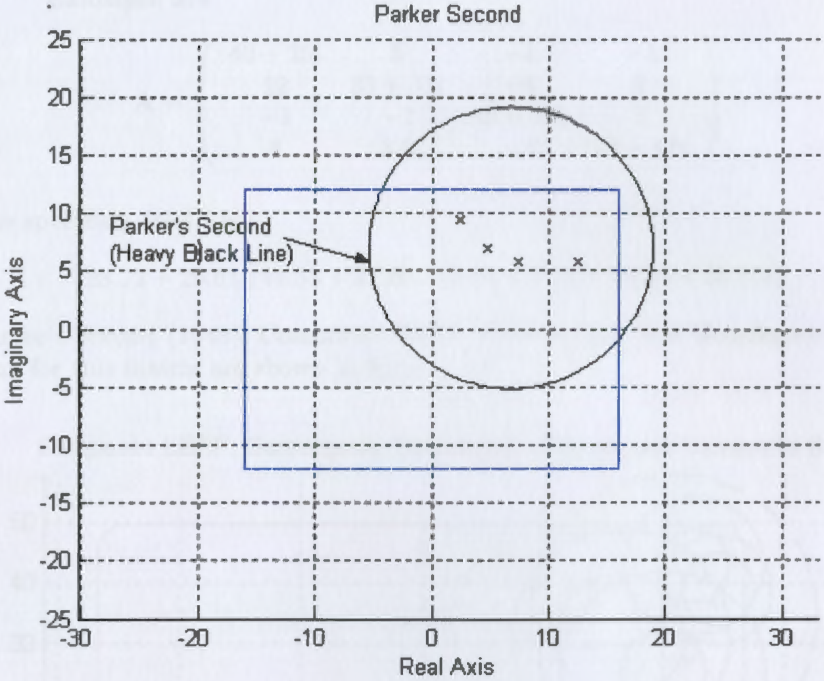


Figure 2.3

Two important things must be noted. First of all, the inclusion set produced by Parker's Second Theorem is usually smaller than the set produced by the Composite BBFP. Secondly, Parker's does not necessarily produce a subset of the Composite BBFP set, as the last example shows.

So, on the one hand, Parker's does not necessarily produce a subset of the Composite BBFP set. On the other hand, after applying Parker's Second Theorem to a number of different matrices, it began to appear that the intersection of the Composite BBFP and Gerschgorin may be a subset of Parker's Second Theorem. If that is the case, there will be no need to even consider Parker's Second. It does seem reasonable that the one is a subset of the other since Parker is using both Gerschgorin and Pre-Gerschgorin ideas. However, the following example shows that this is not the case.

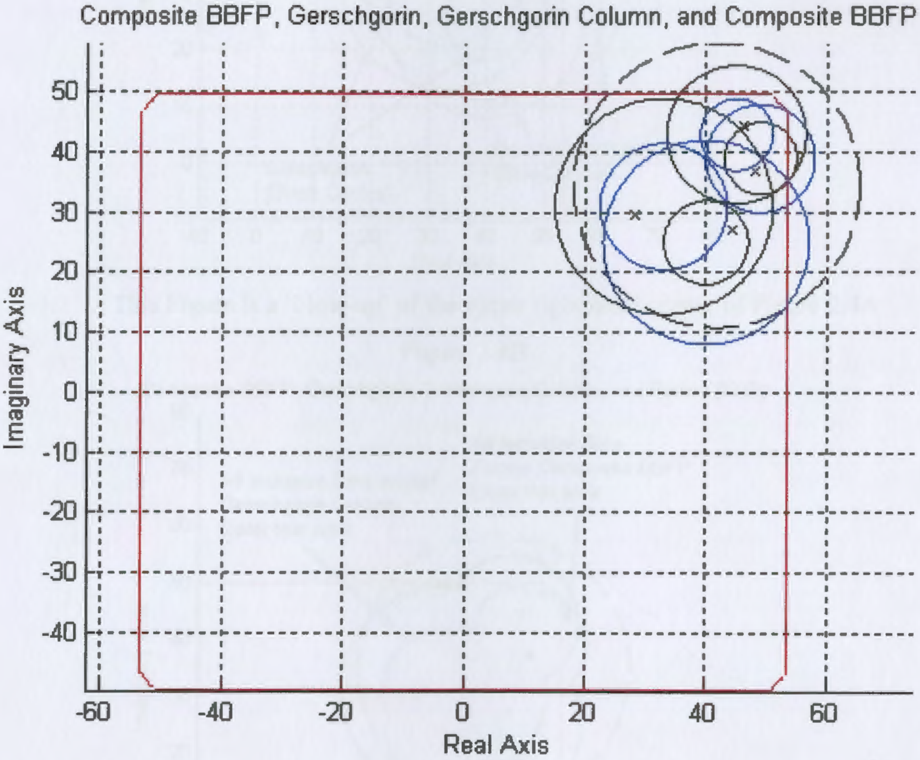
Example 2.4

$$\mathbf{A} = \begin{pmatrix} 40+25i & 5 & -1 & -1 \\ 12 & 33+31i & -4 & 2 \\ -1 & -2 & 49+39i & 3 \\ 4 & 3.5 & -4 & 45+43i \end{pmatrix}$$

The spectrum, $\sigma(A)$, is

$$\{28.22+29.61i, 44.35+27.09i, 46.33+44.45i, 48.10+36.85i\}$$

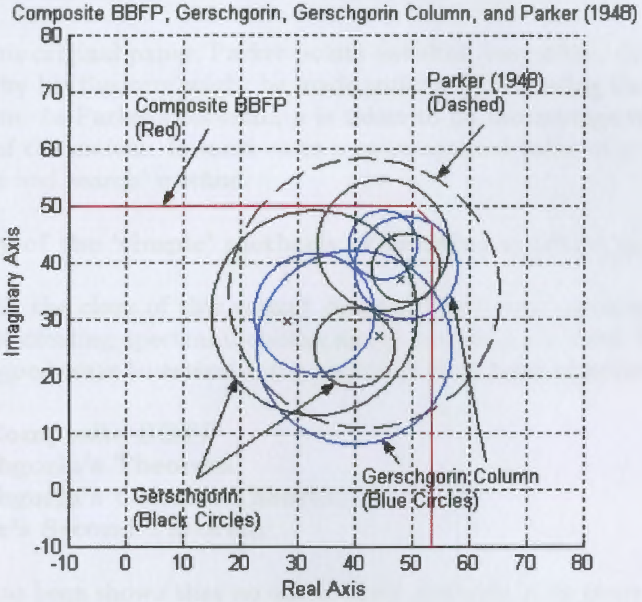
Parker's Second (1948), Composite BBFP, Gerschgorin, and Gerschgorin Column for this matrix are shown in figure 2.4A.



In this Figure, the Composite BBFP Set is in red; the Gerschgorin Set is in Black; the Gerschgorin Column Set is in blue; and Parker's (1948) Set is enclosed by the dashed circle. In this example, each of these sets will contribute to reducing the size of the Spectral Inclusion Set. This fact will become clear by studying Figures 2.4B and 2.4C.

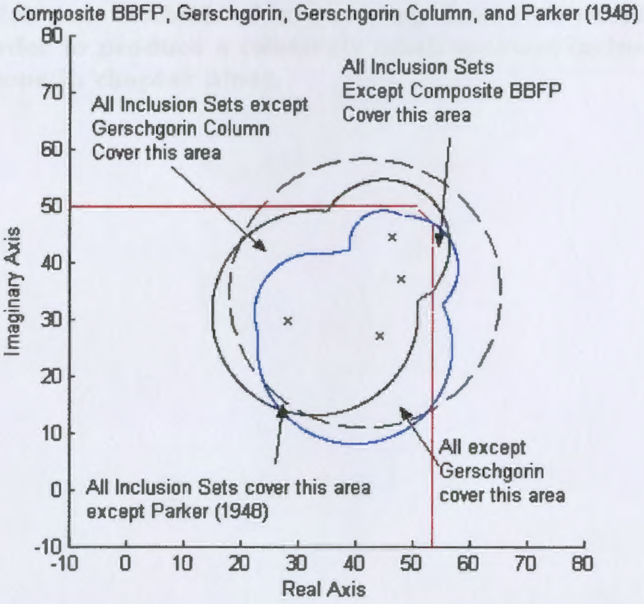
Figure 2.4A

The 'blow-up' of a portion of figure 2.4A is shown in 2.4B and 2.4C. The arrows in 2.4C are highlighting regions that are covered by three of the inclusion sets but not the fourth.



This Figure is a 'blow-up' of the upper right hand corner of Figure 2.4A

Figure 2.4B



This figure is the same as Figure 2.4 B, except the superfluous lines in the Gerschgorin and Gerschgorin Column Sets have been removed.

Figure 2.4C

Note that each of the four sets is not covering parts of the complex plane that are covered by the other three. This means that each of the four sets is in some way helping to reduce the spectral inclusion set. Therefore, no one set is a superset of the intersection of the other three.

In his original paper, Parker points out that, very often, the inclusion set produced by his theorem might be made smaller by adjusting the value of μ in his theorem. In Parker's theorem, μ is taken to be the average of the diagonal elements of the matrix. In most cases a more optimal value of μ may be found by the 'hit and search' method.

Summary of the 'simple' methods of creating spectral inclusion sets

With the close of this second chapter, the examination of the 'simple' methods of creating spectral inclusion sets is complete. In these first two chapters, four good ways to estimate the spectrum have been established:

1. **The Composite BBFP**
2. **Gerschgorin's Theorem**
3. **Gerschgorin's Column Theorem**
4. **Parker's Second Theorem**

It has been shown that no one of these methods is, in general, a subset of any of the others. More importantly, in general, *intersecting the sets generated by all four of these methods will produce a smaller inclusion set than by Intersecting any three sets.* **In fact, one of the purposes of this thesis will be to establish new methods of intersecting these four simply-generated sets in order to produce a relatively small spectral inclusion set (this will be done in chapter nine).**

3 Methods related to Gerschgorin (Varga-Medley)

Once the Gerschgorin disks have been found for a particular matrix or operator, certain analysis can be done on the disks in order to better define the spectrum. This chapter deals with some of that analysis.

Before the analysis on the Gerschgorin disks can be considered, a couple of facts must be established. In the 1940s, Olga Taussky [74], [75] formulated some very important theorems related to isolated Gerschgorin disks:

Theorem 3.1 (Olga Taussky) If a group of k Gerschgorin disks are isolated from the other disks, this group of isolated disks contain exactly k eigenvalues.

Proof (This proof is based on Meyer [58]) Let $A \in C^{nxn}$ such that the Gerschgorin disks of A include a group of k isolated Gerschgorin disks. Let $D =$ diag $\{a_{11}, ..., a_{nn}\}$. Let $B = A - D$ so that $A = B + D$. Let $C(t) = Bt + D$ where $t \in [0, 1]$, so $C(0) = D$ and $C(1) = A$. Then the Gerschgorin disks of C(t) consist of the $z's$ that satisfy:

$$|z - a_{ii}| \leq tr_i = t \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|.$$

Notice that when t=0, the k isolated Gerschgorin disks consists of only the points $a_{11}, ..., a_{kk}$ which are also the eigenvalues of C(0). As t is increased from 0 to 1, the Gerschgorin disks grow and the eigenvalues change. Now a property of eigenvalues states that the eigenvalues vary *continuously with the entries of the matrix.* So, as t increases from 0 to 1, the eigenvalues of the group of k isolated disks will trace out k continuous curves. For example, the i^{th} curve will start at a_{ii} when $t = 0$ and end at λ_i when $t = 1$. Since each of the k curves begins within this group of k isolated disks centered at a_{ii}, no curve can leave this group of disks without causing a discontinuity. Since each curve is continuous, the k eigenvalues must remain within this group of isolated disks. ● ● ●●

This theorem is illustrated in the next example.

Example 3.2 Let

$$\begin{pmatrix} 38+32i & 2 & 4-5i & 1 & .5 & 1-.5i \\ 2-3i & -10-8i & 3 & 5i & 2 & -1 \\ i & 2 & -15-12i & 6i & 2 & -3 \\ -5i & 2 & 7i & -12-4i & .5 & 6 \\ 1-i & -5-2i & 3 & 5i & 25+30i & -1 \\ i & 2 & 2i & .5 & 2 & -20-5i \end{pmatrix}.$$

The Gerschgorin disks for this example are shown in figure 3.2. Notice that the group of two isolated Gerschgorin disks contain exactly two eigenvalues as predicted by Theorem 3.1.

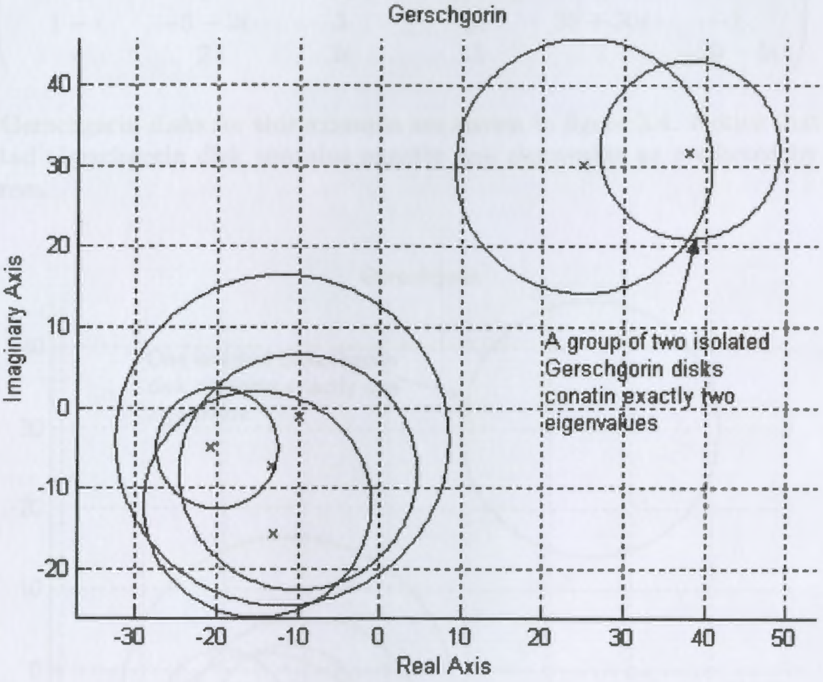


Figure 3.2

Corollary 3.3 (Olga Taussky) If one Gerschgorin disk is isolated from the other disks, this isolated disk contains exactly one eigenvalue. Corollary 3.3 is a special case of Theorem 3.1.

• • ••

This Corollary is illustrated in the next example.

Example 3.4 Let

$$\begin{pmatrix} -12-6i & 2 & 2-4i & 1 & .5 & 1-.5i \\ 2-3i & -10-8i & 3 & 5i & 2 & -1 \\ i & 2 & -15-12i & 6i & 2 & -3 \\ -5i & 2 & 7i & -12-4i & .5 & 6 \\ 1-i & -5-2i & 3 & 5i & 25+30i & -1 \\ i & 2 & 2i & .5 & 2 & -20-5i \end{pmatrix}.$$

The Gerschgorin disks for this example are shown in figure 3.4. Notice that the isolated Gerschgorin disk contains exactly one eigenvalue as predicted by the theorem.

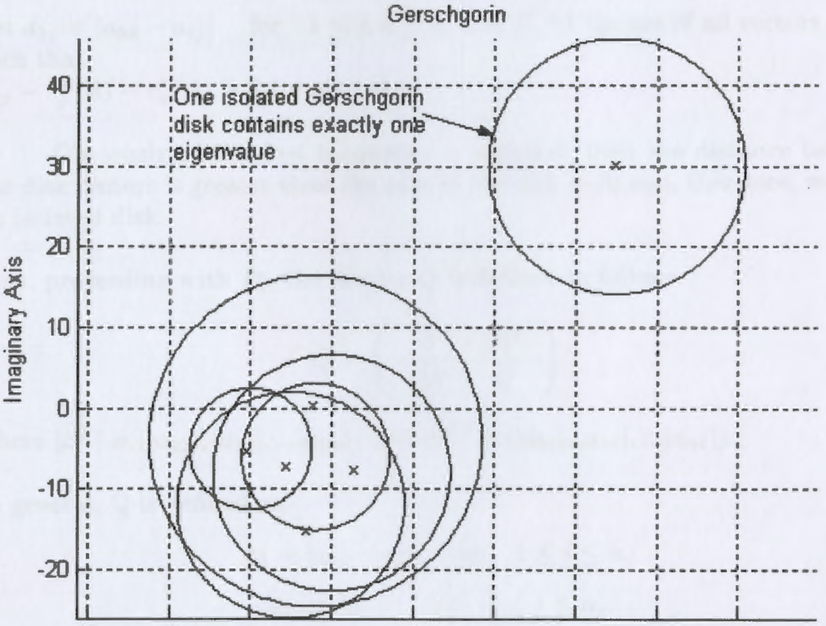


Figure 3.4

The actual eigenvalues are represented by the X's

Once these facts about isolated Gerschgorin disks became known, it was possible to develop methods to calculate the exact value of the eigenvalue(s) located in these disks.

Varga-Medley Method

In the early 1960's Richard Varga and Helen Medley devised methods for taking advantage of isolated Gerschgorin disks. That is, they developed methods for computing the *exact* value of the eigenvalues located inside of the isolated Gerschgorin disks.

The equations that appear in the Varga and Medley papers [55],[86] are presented below. A Matlab program appears in the appendix. The reader should look at the Matlab code in combination with the equations below in order to understand the practical use of the Varga-Medley method.

Let $A \in C^{nxn}$. Define

$$r_i^{\mathbf{x}}(A) = \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{|a_{ij}|x_j}{x_i} \quad \text{for} \quad 1 \leq i \leq n \quad \text{and} \quad x_i > 0.$$

Let $d_{kj} = |a_{kk} - a_{jj}|$ for $1 \leq j, k \leq n$. Let P_k be the set of all vectors $\mathbf{x} > \mathbf{0}$ such that
$d_{kj} - r_j^{\mathbf{x}}(A) - r_k^{\mathbf{x}}(A) \geq 0$ for all $j \neq k$.

Obviously, if this last inequality is satisfied, then the distance between the disk centers is greater than the sum of the disk radii and, therefore, we have an isolated disk.

Now, proceeding with P_1, the matrix Q is defined as follows:

$$Q = \left(\begin{array}{c|c} 0 & |\widehat{\beta}|^T \\ \hline -|\widehat{\gamma}| & \widehat{Q} \end{array} \right)$$

where $|\widehat{\beta}^T| = (|a_{12}|, |a_{13}|, ...|a_{1n}|)$ and $|\widehat{\gamma}^T| = (|a_{21}|, |a_{31}|, ...|a_{n1}|)$.

In general, Q is defined as:

$$q_{ii} = |a_{11} - a_{ii}| \quad \text{for} \quad 1 \leq i \leq n,$$

$$q_{1j} = |a_{1j}| \quad \text{for} \quad 2 \leq j \leq n,$$

$$q_{ij} = -|a_{ij}| \quad \text{for} \quad i \neq j \quad \text{and} \quad i \neq 1.$$

Medley and Varga partition the matrix A as follows

$$A = \left(\begin{array}{c|c} a_{11} & \widehat{\beta}^T \\ \hline \widehat{\gamma} & A_{22} \end{array} \right)$$

The matrix B is defined as:

$$B = \left(\begin{array}{c|c} 0 & \widehat{\beta}^T \\ \hline \widehat{\gamma} & \widetilde{B} \end{array} \right)$$

where $\widetilde{B} = A_{22} - a_{11} I_{n-1}$ and I_{n-1} is the (n-1)x(n-1) identity matrix.

Note that if the isolated disk under consideration is *not* associated with the first row of the matrix, then a_{11} will not be used. In that case, the isolated disk is associated with the k^{th} row and $\widetilde{B}$ becomes:

$$\widetilde{B} = A_{22} - a_{kk} I_{n-1}.$$

Once all of the above calculation are done, the eigenvalue of this isolated disk can be found by an iterative method using following steps:

Algorithm

Step 1. Initialize z=.05.

Step 2. Calculate a new value of z using:

$$z = -\widehat{\beta}^T (\widehat{\beta} - z I_{n-1})^{-1} \widehat{\gamma}.$$

Step 3. Substitute this new value of z back into Step 2.

Notice that $\widetilde{\beta}$ and $\widetilde{\gamma}$ are constructed just once.

Continue Steps 2 and 3 until z converges. The eigenvalue for this isolated disk is given by:

$$\lambda = a_{kk} + z.$$

Example 3.5 Let

$$\mathbf{A} = \begin{pmatrix} -7 & -1.5 & 5 & 7i & 3 \\ 2 & 3i & 3 & 4 & -7 \\ 3 & 2 & 30-8i & 1.5 & 2.5 \\ 6 & 3 & 4 & 2 & 3 \\ -7 & -4 & 3 & 2 & 5 \end{pmatrix}.$$

The Gerschgorin disks for this matrix are shown in figure 3.5.

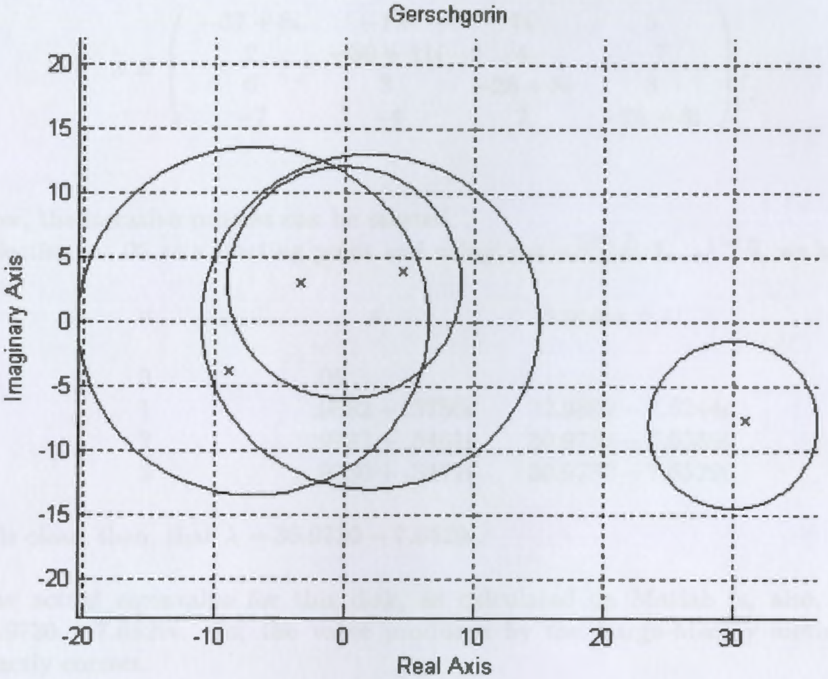


Figure 3.5

Notice that one of the Gerschgorin disks for this matrix is isolated. That isolated disk is associated with the third row. Thus, the third row and the third column will be used in the construction of γ,β, and A_{22}. So, γ will be made up of a_{13},a_{23},a_{43},a_{53} and β will be made up of a_{31},a_{32},a_{34},a_{35}. The matrix A_{22} is made from A by including all of A except the third row and the third column. This gives,

$$\gamma = \begin{pmatrix} 5 \\ 3 \\ 4 \\ 3 \end{pmatrix}, \quad \beta = \begin{pmatrix} 3 & 2 & 1.5 & 2.5 \end{pmatrix}, \quad \mathbf{A_{22}} = \begin{pmatrix} -7 & -1.5 & 7i & 3 \\ 2 & 3i & 4 & -7 \\ 6 & 3 & 2 & 3 \\ -7 & -4 & 2 & 5 \end{pmatrix}.$$

Since the isolated disk is associated with the 3rd row, $\widetilde{\beta} = A_{22} - a_{33}I_{n-1}$. Then

$$\widetilde{\beta} = \begin{pmatrix} -7 & -1.5 & 7i & 3 \\ 2 & 3i & 4 & -7 \\ 6 & 3 & 2 & 3 \\ -7 & -4 & 2 & 5 \end{pmatrix} - \begin{pmatrix} 30-8i & 0 & 0 & 0 \\ 0 & 30-8i & 0 & 0 \\ 0 & 0 & 30-8i & 0 \\ 0 & 0 & 0 & 30-8i \end{pmatrix}.$$

The result is

$$\widetilde{\beta} = \begin{pmatrix} -37+8i & -1.5 & 7i & 3 \\ 2 & -30+11i & 4 & -7 \\ 6 & 3 & -28+8i & 3 \\ -7 & -4 & 2 & -25+8i \end{pmatrix}.$$

Now, the iterative process can be started.
Selecting z=.05 as a starting point and using $z = -\widehat{\beta}^T(\widetilde{B}zI_{n-1})^{-1}\widehat{\gamma}$, we have

n	z	$\lambda = a_{33} + z$
0	$.05$	
1	$.9882 + .3756i$	$30.9882 - 7.6244i$
2	$.9732 + .3461i$	$30.9732 - 7.6539i$
3	$.9730 + .3471i$	$30.9730 - 7.6529i$.

It is clear, then, that $\lambda = 30.9730 - 7.6529i$.

The actual eigenvalue for this disk, as calculated on Matlab is, also, $\lambda = 30.9730 - 7.6529i$. So, the value produced by the Varga-Medley method is exactly correct.

Consider the next example.

Example 3.6 Let

$$\mathbf{A} = \begin{pmatrix} -5+10i & 4 & 2 & 3 & 1 \\ 2 & 8-7i & 2 & 1.5 & 1 \\ 2.5 & 2 & 18+19i & 3 & 1 \\ -11 & 2 & 3 & -22-18i & 3 \\ 2 & 2 & 3 & 1 & -25+10i \end{pmatrix}.$$

The Gerschgorin disks for this matrix are located in figure 3.6. Notice that *all five* of the Gerschgorin disks for this matrix are isolated. The Medley/Varga method yields the following five eigenvalues:

$\lambda_1 = -5.7214 + 10.5932i$
$\lambda_2 = 8.2069 - 6.5804i$
$\lambda_3 = 18.4312 + 18.6291i$
$\lambda_4 = -21.7234 - 18.5776i$
$\lambda_5 = -25.1933 + 9.9357i.$

Once again, these numbers match the actual eigenvalues *exactly.*

The Medley/Varga method is very fast and very efficient and shows the power and richness of the Gerschgorin disks.

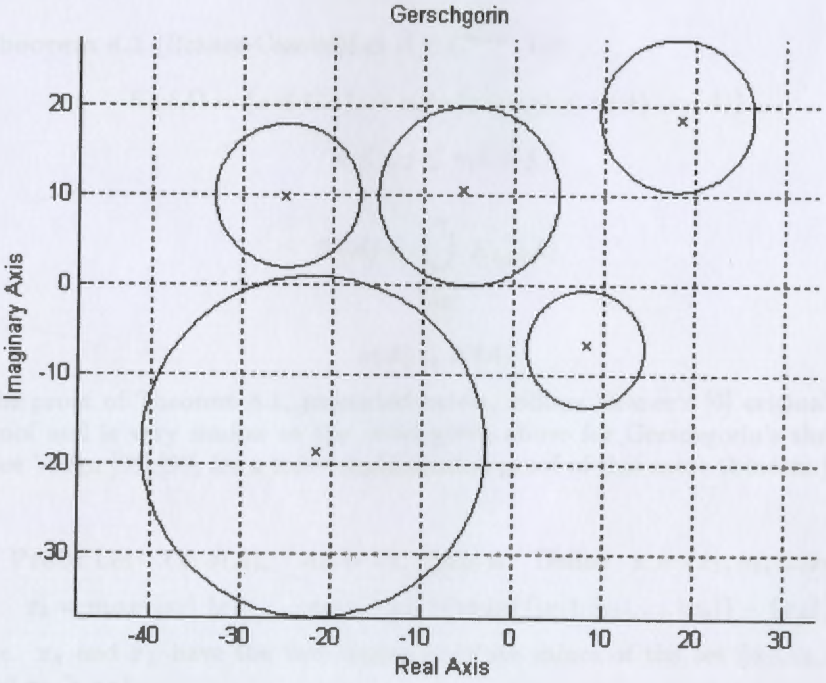


Figure 3.6

The actual eigenvalues are represented by the X's

4 Gerschgorin-Type Methods

This chapter begins the study of the 'involved' methods of producing spectral inclusion sets. In this context, 'involved' does not necessarily imply complicated calculations but 'involved' does mean a larger *quantity* of computations, transformations, and/or iterations in order to produce a result. Such methods, therefore, will require more computer time than the methods previously considered.

This particular chapter covers the Gerschgorin-Type methods. That is, methods that are based on Gerschgorin's original theorem but go much further. As with the regular Gerschgorin Theorem, Richard Varga, has done extensive work with these Gerschgorin-Type methods. In particular he has, over the past fifty years, greatly advanced our understanding of the minimal Gerschgorin sets both on the theoretical and the practical level.

Section 4.1 Brauer's Ovals of Cassini

In the 1940's Alfred Brauer [6] improved on the original Gerschgorin theorem. Brauer began with Gerschgorin's idea of using the diagonal elements of the matrix as the centers of disks and the sums absolute values of the off-diagonal elements as the radii of the disks but extended the idea by considering *two rows at a time.* Brauer realized that he could reduce the size of Gerschgorin's inclusion set by considering two rows at a time instead of one. His theorem is stated as follows:

Theorem 4.1 (Brauer-Cassini)Let $A \in C^{nxn}$. Let

$$K_{ij}(A) = \{z \in C : |z - a_{ii}| \cdot |z - a_{jj}| \leq r_i(A) \cdot r_j(A)\} \qquad (1)$$

$$1 \leq i, j \leq n; i \neq j$$

$$K(A) = \bigcup_{\substack{i,j=1 \\ j \neq i}}^{n} K_{i,j}(A)$$

$$\sigma(A) \subseteq K(A)$$

The proof of Theorem 4.1, presented below, follows Brauer's [6] original 1947 proof and is very similar to the proof given above for Gerschgorin's theorem. (See Varga [88],[90] for a more sophisticated proof of this same theorem.)

Proof Let $\lambda \in \sigma(A)$, $A\mathbf{x} = \lambda\mathbf{x}$, $\mathbf{x} \neq \mathbf{0}$. Define $\mathbf{x} = \{x_1, x_2, ..., x_n\}$,

$$x_k = max\{|x_1|, |x_2|, ..., |x_n|\}, \quad x_L = max\{\{|x_1|, |x_2|, ..., |x_n|\} - \{x_k\}\}.$$

(i.e. x_k and x_L have the two largest absolute values of the set $\{x_1, x_2, ...x_n\}$ and $x_k \geq x_L$)

Consider the k^{th} row of $\lambda\mathbf{x} = A\mathbf{x}$,

$$\lambda x_k = \sum_{j=1}^{n} a_{kj}x_j.$$

Subtracting $a_{kk}x_k$ from both sides,

$$\lambda x_k - a_{kk}x_k = \sum_{\substack{j=1\\ j\neq k}}^{n} a_{kj}x_j, \quad \text{and} \quad (\lambda - a_{kk})x_k = \sum_{\substack{j=1\\ j\neq k}}^{n} a_{kj}x_j.$$

Consider the L^{th} row of $\lambda\mathbf{x} = A\mathbf{x}$:

$$\lambda x_L = \sum_{j=1}^{n} a_{Lj}x_j.$$

Subtracting $a_{LL}x_L$ from both sides, (recall x_L is the second largest absolute value from the set $\{x_1, x_2, ...x_n\}$),

$$\lambda x_L - a_{LL}x_L = \sum_{\substack{j=1\\ j\neq L}}^{n} a_{Lj}x_j, \tag{2}$$

and thus

$$(\lambda - a_{LL})x_L = \sum_{\substack{j=1\\ j\neq L}}^{n} a_{Lj}x_j. \tag{3}$$

Multiplying (2) and (3), we have

$$(\lambda - a_{kk})(\lambda - a_{LL})x_k x_L = \Big(\sum_{\substack{j=1\\ j\neq k}}^{n} a_{kj}x_j\Big)\Big(\sum_{\substack{j=1\\ j\neq L}}^{n} a_{Lj}x_j\Big),$$

and thus

$$|\lambda - a_{kk}||\lambda - a_{LL}||x_k x_L| \leq |x_k|\Big(\sum_{\substack{j=1\\ j\neq k}}^{n} |a_{kj}|\Big)|x_L|\Big(\sum_{\substack{j=1\\ j\neq L}}^{n} |a_{Lj}|\Big).$$

● ● ●●

Note that the set produced by Theorem 4.1 is sometimes called 'Brauer's Ovals of Cassini', 'Brauer-Cassini', 'the Ovals of Cassini', and 'Cassini'. All of these names will be used in this thesis.

A valuable feature of Brauer's Ovals of Cassini is that the resulting set is always a subset of the Gerschgorin disks. This means that when using Brauer's

Ovals of Cassini, it is not necessary to consider the Gerschgorin disks. This result is stated in the next theorem.

Theorem 4.2 (Brauer's Ovals of Cassini is a subset of Gerschgorin) Let $A \in C^{nxn}$. Then

$$K(A) \subseteq G(A)$$

Proof This proof follows R.S. Varga [88]
Fix i and j with $1 \leq i, j \leq$ and $i \neq j$.Let $z_o \in K_{ij}(A)$ then

$$|z_o - a_{ii}| \cdot |z_o - a_{jj}| \leq r_i(A) \cdot r_j(A).$$

Rearranging, we have:

$$(\frac{|z_o - a_{ii}|}{r_i(A)})(\frac{|z_o - a_{jj}|}{r_j(A)}) \leq 1.$$

Note that for this inequality to hold, either

$$(\frac{|z_o - a_{ii}|}{r_i(A)}) \leq 1 \quad or \quad (\frac{|z_o - a_{jj}|}{r_j(A)}) \leq 1.$$

So, without loss of generality, assume that

$$(\frac{|z_o - a_{ii}|}{r_i(A)}) \leq 1.$$

Rearranging, we have

$$|z_o - a_{ii}| \leq r_i(A).$$

Therefore,z_o is in one of the Gerschgorin disks:

$$G_i(A) = \{z \in C : |z - a_{ii}| \leq r_i(A) = \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|\} \quad \text{for} \quad 1 \leq i \leq n.$$

Therefore,

$$K(A) \subseteq G(A)$$

● ● ●●

Section 4.1.1 Applying the Brauer-Cassini Theorem

Obviously, generating a Brauer-Cassini set is slightly more difficult than generating a Gerschgorin set and is best done by computer. Since it is necessary to find values of z that satisfy,

$$|z - a_{ii}| \cdot |z - a_{jj}| \leq r_i(A) \cdot r_j(A) \tag{4}$$

this inequality will be treated as any equation that uses complex numbers. Therefore, let
$z = x + yi$ and $a_1 = \mathrm{Re}(a_{ii})$; $b_1 = \mathrm{Im}(a_{ii})$; $a_2 = \mathrm{Re}(a_{jj})$; $b_2 = \mathrm{Im}(a_{jj})$

The left side of (4) becomes

$$|[x + yi - (a_1 + b_1 i)][x + yi - (a_2 + b_2 i)]| \quad =$$

$$\begin{aligned}
&= |[(x - a_1) + (y - b_1)i][(x - a_2) + (y - b_2)i]| \\
&= |(x - a_1)(x - a_2) + [(x - a_1)(y - b_2) + (x - a_2)(y - b_1)]i - (y - b_1)(y - b_2)| \\
&= |(x - a_1)(x - a_2) - (y - b_1)(y - b_2) + [(x - a_1)(y - b_2) + (x - a_2)(y - b_1)]i| \\
&= |x^2 - (a_1 + a_2)x + a_1a_2 - (y^2 - (b_1 + b_2)y + b_1b_2 \\
&\qquad + [xy - b_2x - a_1y + a_1b_2 + xy - b_1x - a_2y + a_2b_1]i|
\end{aligned}$$

Applying the absolute value,

$$\sqrt{\begin{aligned}&[x^2 - (a_1 + a_2)x + a_1a_2 - (y^2 - (b_1 + b_2)y + b_1b_2]^2 \\ &\qquad + [xy - b_2x - a_1y + a_1b_2 + xy - b_1x - a_2y + a_2b_1]^2\end{aligned}}$$

This square root is the left hand side of (4). Squaring both sides of (4) produces,

$$\begin{aligned}&[x^2 - (a_1 + a_2)x + a_1a_2 - (y^2 - (b_1 + b_2)y + b_1b_2]^2 + [xy - b_2x - a_1y + a_1b_2 + \\ &xy - b_1x - a_2y + a_2b_1]^2 \\ &\qquad \leq (r_i(A) \cdot r_j(A))^2 \qquad (5)\end{aligned}$$

So, the Brauer-Cassini set consists of values of x and y that satisfy inequality (5).

The above is a natural way of approaching the Brauer-Cassini theorem. However, there is a more efficient way to produce the Brauer-Cassini set when using a computer. Some years after Brauer developed his theorem, he suggested a much simpler method [9] to calculate the Brauer-Cassini set. Combining Brauer's suggestions along with considerations for computer applications, it is possible to develop an efficient algorithm to produce the Brauer-Cassini sets. That algorithm is presented below.

Algorithm 4.1

Let $A \in C^{nxn}$.

Step 1
For $k = 1, 2, ...n$ calculate the P_k's:

$$\sum_{\substack{j=1 \\ j \neq k}}^{n} a_{kj} = P_k$$

Step 2
Select a point 'z' from some spectral inclusion set of the matrix A.

Step 3
Begin with k=1 and see if 'z' satisfies the following inequality:

$$|z - a_{kk}| \leq P_k$$

If 'z' does not satisfy the inequality, then try k=2, k=3, etc. Then consider the following:

3a If 'z' does not satisfy the inequality for one or more of the k's , then 'z' is not in the Brauer-Cassini set for the matrix. In that case, change the value of 'z' by some increment and go back to step 2.

3b If 'z' satisfies the inequality for one or more of the k's , then go to step 4.

Step 4
For each of the k's that z satisfies, check to see if 'z' satisfies the following inequality.

$$|z - a_{kk}||z - a_{LL}| \leq P_k P_L$$

If 'z' satisfies this inequality for the current 'k' and for any $L = 1, 2, ..n$ with $k \neq L$ then 'z' is in the Brauer-Cassini set.

Step 5
Change 'z' by some increment and return to step 2.

This process is to be continued until all of the points in the inclusion set have been covered.

● ● ●●

When using the Algorithm 4.1, it is necessary to have some inclusion set available from which the z's can be selected. Any inclusion set such as the Norm of the matrix or the Gerschgorin set will serve that purpose.

Different algorithms for producing the Brauer-Cassini sets were tested on the computer, this Algorithm 4.1 proved to be the simplest and most efficient. (Find the Matlab code in the Appendix.)

Example 4.3 Let

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & -1 & 2 \\ 1 & 2 & 1 & -1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & .5 & -1 \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{-1.79, 2.17, .81 + 341i, .81 - 341i\}$$

The Brauer-Cassini set for this matrix is plotted in figure 4.3. The Brauer-Cassini set consists of the points represented by the red stars. This graph also shows the Gerschgorin circles for the same matrix. The basic Gerschgorin set is the set enclosed by the union of the green circles. As guaranteed by Theorem 4.2, the Brauer-Cassini set is a subset of Gerschgorin.

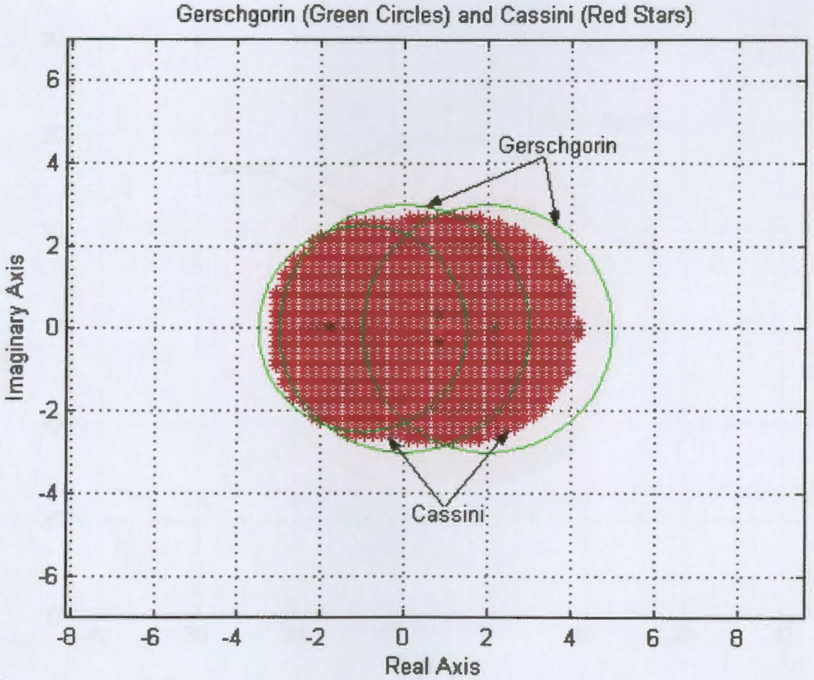


Figure 4.3

In these figures, the eigenvalues are represented by the X's

Revisiting Example 1.29,
Example 4.4 Recall

$$\mathbf{A} = \begin{pmatrix} 2+3i & i & 4 & i+1 \\ 4-4i & 2 & 1+i & 2+2i \\ 3i & 4 & -4i & 5i \\ -7 & 2-5i & 6 & -5+i \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{7.79 - .12i, .97 + 6.19i, -5.47 - 5.89i, -4.29 - .18i\}$$

The Brauer-Cassini set for this matrix is shown in figure 4.4. The Brauer-Cassini set consists of the points represented by the red stars. The basic Gerschgorin set is the set enclosed by the union of the green circles.

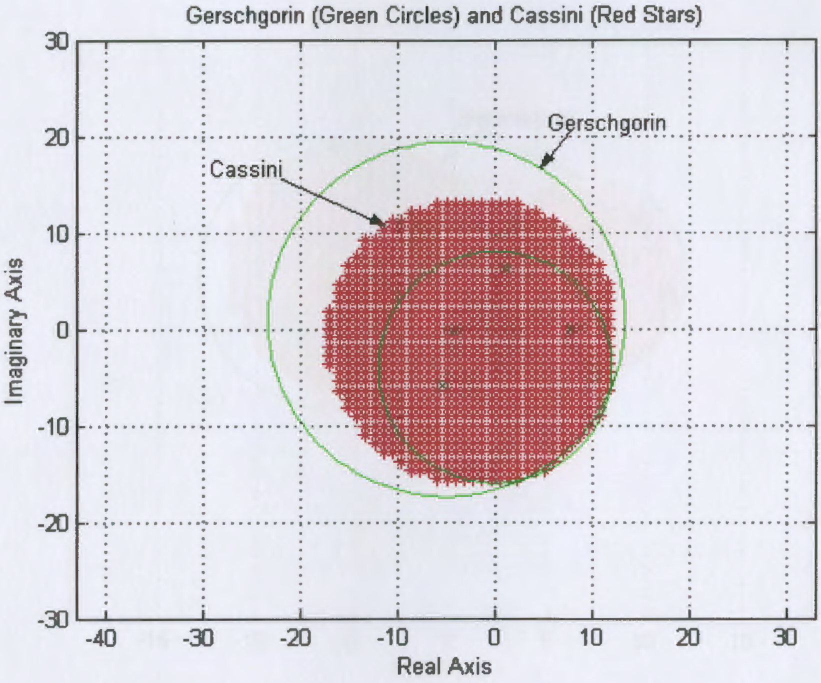


Figure 4.4

In these figures, the eigenvalues are represented by the X's

Example 4.5 Let

$$\mathbf{A} = \begin{pmatrix} -4+4i & -2 & 3 & -1 \\ 1 & 6+2i & -2 & 2 \\ 2 & 2 & -6+3i & 3 \\ .5 & .5 & -4 & 7+3i \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{-6.46+3.43i, -2.17+3.61i, 5.67+1.78i, 5.96+3.18i\}$$

The Brauer-Cassini set for this matrix is shown in figure 4.5. The Brauer-Cassini set consists of the points represented by the red stars. The basic Gerschgorin set is the set enclosed by the union of the green circles.

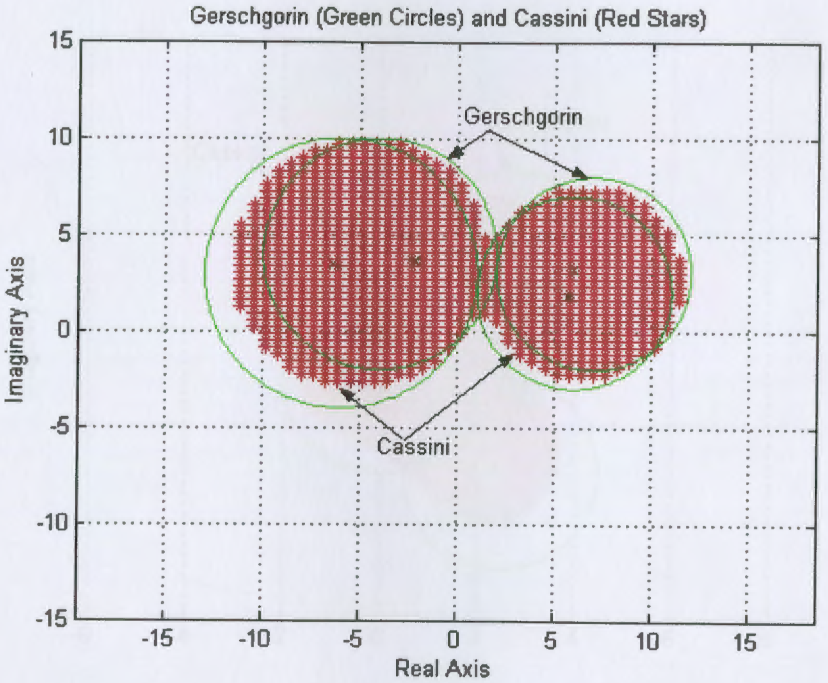


Figure 4.5

In these figures, the eigenvalues are represented by the X's

Example 4.6 Let

$$\mathbf{A} = \begin{pmatrix} 2.75+i & 1 & .75 & -.5 \\ -.75 & -i & .5 & .75 \\ .75 & .5 & .1+i & 1 \\ 1 & .5i & -.5i & 2.5-3i \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{2.41 - 3.03i, 2.75 + 1.12i, .36 - 1.08i, -.17 + .99i\}$$

The Brauer-Cassini set for this matrix is shown in figure 4.6. The Brauer-Cassini set consists of the points represented by the red stars. The basic Gerschgorin set is enclosed by the union of the green circles.

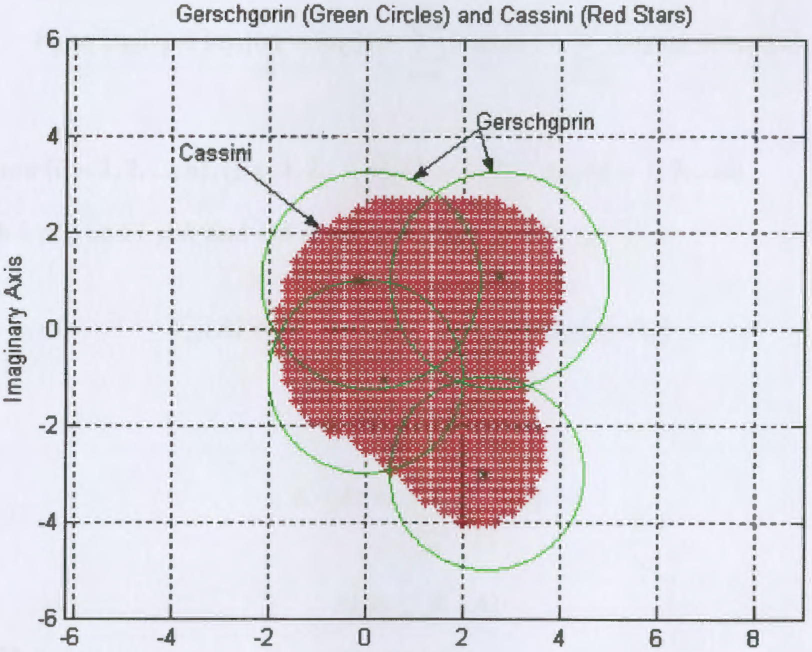


Figure 4.6

In in each of the last four examples, the Brauer-Cassini set was a bit smaller than the Gerschgorin set. The performance of Gerschgorin and Brauer-Cassini in these examples is very typical of the relative 'sharpness' of these two methods.

Section 4.2 Cassini sets for Real matrices

In 1952, two separate articles appeared on the subject of Brauer-Cassini sets for real matrices. One article was by Gene W. Medlin [57] of the University of North Carolina and the other article was by Alfred Brauer [8] also of UNC. Both articles presented essentially the same equations for determining a Brauer-Cassini set for real matrices. The equations they propose will, in general, produce slightly smaller spectral inclusion sets than the general Cassini theorem when dealing with real matrices.

Theorem 4.7 (Brauer-Medlin-Cassini) Let $A \in R^{nxn}$. Let

$$r_s(A) = \sum_{\substack{t=1 \\ t \neq s}}^{n} |a_{st}| \quad \text{for} \quad 1 \leq s \leq n.$$

Let

$$P_{ij} = |a_{ij}|r_j + |a_{ji}|(r_i - |a_{ij}|) + \sum_{k=1}^{n} |a_{ik}a_{jk}| + \sum_{k<q} |a_{ik}a_{jq} + a_{iq}a_{jk}|.$$

Where $(i = 1, 2, ..., n), (j = 1, 2, ..., n), (k = 1, 2, ..., n), (q = 1, 2, ...n)$

with $i \neq j$ and $i \neq k$ and $i \neq q$ and $j \neq q$ and $j \neq k$.

Let

$$K^{'}_{ij}(A) = \{z \in C : |z - a_{ii}| \cdot |z - a_{jj}| \leq P_{ij}\}$$

$$1 \leq i, j \leq n; i \neq j$$

$$K^{'}(A) = \bigcup_{\substack{i,j=1 \\ j \neq i}}^{n} K^{'}_{i,j}(A)$$

$$\sigma(A) \subseteq K^{'}(A)$$

● ● ● ●

Proof of this theorem may be found in Medlin [57] or Brauer [8].

Note that the difference between the general Cassini theorem and this theorem is that $r_i(A) \cdot r_j(A)$ is replaced by P_{ij}. A careful study will reveal that $P_{ij} \leq r_i(A) \cdot r_j(A)$. This of course means that the inclusion set produced by this theorem is as small or smaller than the inclusion set produced by the original Brauer-Cassini theorem.

Example 4.8 Consider the following real matrix,

$$\mathbf{A} = \begin{pmatrix} -123 & 12 & -16 & 22 & 6 & 8 \\ -14 & 77 & 7 & -8 & 9 & 3 \\ 4 & 9 & 6 & -25 & -18 & 29 \\ 9 & -8 & -15 & -5 & 12 & -10 \\ -24 & 16 & -17 & 9 & -8 & 7 \\ -12 & 8 & -15 & -16 & 12 & 22 \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{-122.19, 80.32, 23.7 + 12.1i, 23.7 - 12.1i, -21.84, -14.70\}.$$

The Brauer-Cassini set and the Brauer-Medlin-Cassini set for real matrices are shown in figure 4.8. Notice that the Cassini set for real matrices is slightly smaller than the general Cassini set. This is very typical of the relative performance of these two theorems.

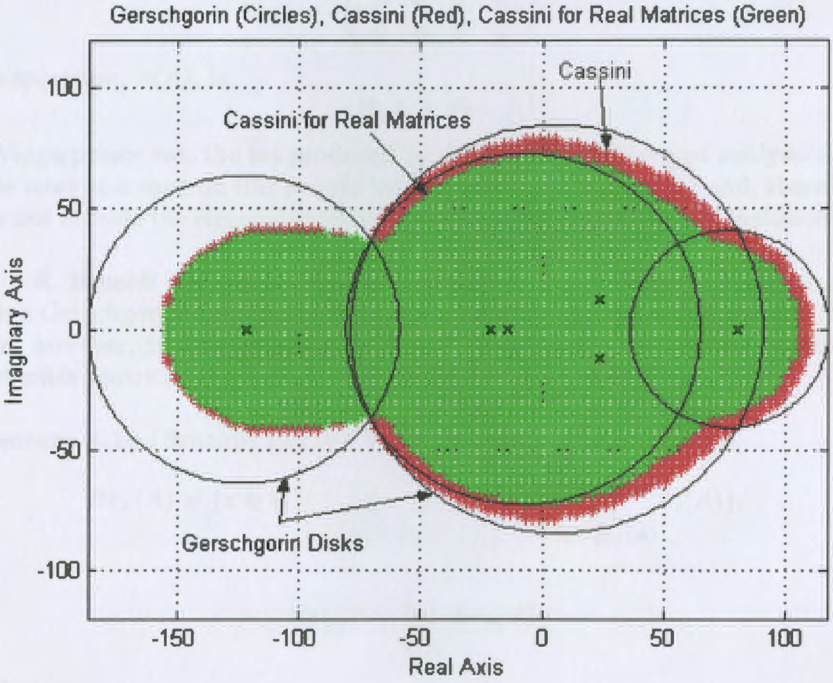


Figure 4.8

In this figure, the eigenvalues are represented by the X's

Section 4.3 Brualdi sets

As we have seen thus far, Gerschgorin produced a spectral inclusion set by considering one row of a matrix at a time. Brauer-Cassini produced a subset of Gerschgorin by using two rows at a time. So, what will happen if multiple rows are used in the calculation? Will the set based on, say, three rows at a time be smaller than the Cassini set? The answer is 'yes'. The resulting set will be smaller. However, **sets produced by using multiple rows will not always include the spectrum!**. In other words, it is not possible to produce a Gerschgorin-type spectral inclusion set by using more than two rows at a time - at least not without restrictions and modifications.

Varga [88] supplies an example, attributed to Morris Newman, that illustrates the problem described above.

Example 4.9 Let

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{0, 1, 1, 2\}.$$

As Varga points out, the set produced by doing Gerschgorin-type analysis using three rows at a time on this matrix includes only one point, z=1 and, therefore, does not include the eigenvalues 0 or 2. So, the set is not a spectral inclusion set.

R. Brualdi [16] was able, by the use of elementary graph theory, to produce a Gerschgorin-type theorem for multiple rows of a matrix. Brualdi's Theorem, however, is rather hard to use and its application is limited to weakly irreducible matrices. The Brualdi Theorem is stated as follows:

Theorem 4.10 (Brualdi) Let $A \in C^{nxn}$ be weakly irreducible. Let

$$Br_k(A) = \{z \in C : \prod_{i \in Cyc_k(A)} |z - a_{i,i}| \leq \prod_{i \in Cyc_k(A)} r_i(A)\},$$

and

$$Br(A) = \bigcup_{k \in M} Br_k(A).$$

Then

$$\sigma(A) \subseteq Br(A).$$

Where M is the number of cycles in A and $Cyc_k(A)$ contains the set of vertices in the kth cycle.

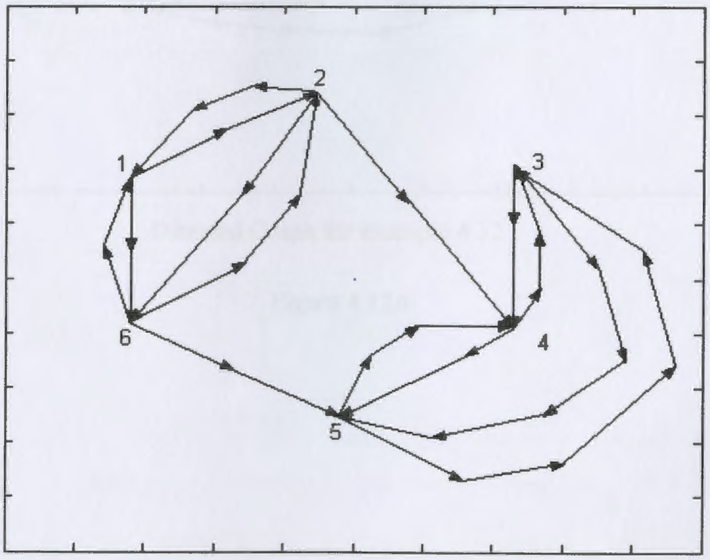
● ● ●●

Some definitions and examples will be required in order to make Theorem 4.10 clear. First, it is necessary to recall some definitions from graph theory. A *directed graph* of a matrix is *strongly connected* if for any vertices v_i and v_j there exists a directed path from v_i to v_j. A matrix is *irreducible* if and only if its associated graph is strongly connected. On the other hand, a matrix is *weakly irreducible* if each vertex of its associated graph belongs to some cycle in the graph. (Note that an irreducible matrix is weakly irreducible).

Example 4.11 Consider,

$$\mathbf{A} = \begin{pmatrix} 5 & 3 & 0 & 0 & 0 & -7 \\ 5 & 9 & 0 & 6 & 0 & -3 \\ 0 & 0 & 4 & 9 & 8 & 0 \\ 0 & 0 & -6 & 8 & 7 & 0 \\ 0 & 0 & 8 & 7 & 9 & 0 \\ -41 & 7 & 0 & 0 & 0 & 14 \end{pmatrix}.$$

The directed graph of matrix A is shown in the figure 4.11. (The details of how this graph was constructed may be found in the Appendix). Notice that this graph *is not strongly connected.* For example, there does not exist a directed path from 5 to 2. Since the graph is not strongly connected, it is not irreducible. On the other hand, each vertex belongs to some cycle in the graph. That is, 1 belongs to C_{126}; 2 belongs to C_{126}; 3 belongs to C_{345};4 belongs to C_{345};5 belongs to C_{35}; and 6 belongs to C_{126}. Therefore, the matrix is weakly irreducible and the Brualdi Theorem may be applied to this matrix.



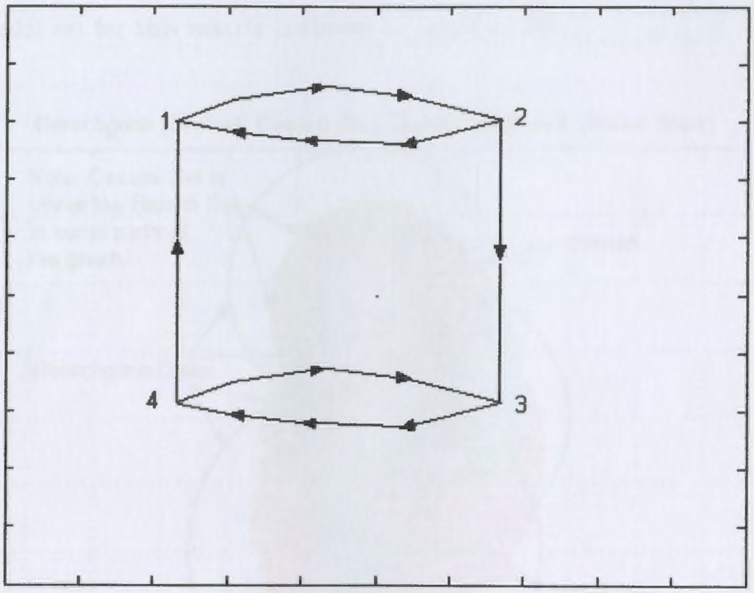
This is the directed graph for the matrix in example 4.11

Figure 4.11

Example 4.12 Let

$$\mathbf{A} = \begin{pmatrix} .5 & 1 & 0 & 0 \\ .75 & i & .25 & 0 \\ 0 & 0 & -.5 & .25 \\ 1 & 0 & .3 & -i \end{pmatrix}.$$

Notice that each vertex is in some cycle (see figure 4.12A). Therefore, the matrix is weakly irreducible and Brualdi's Theorem may be applied to this matrix.



Directed Graph for example 4.12

Figure 4.12A

The cycles in this graph are C_{12}, C_{34} and C_{1234}. So, the Bruadi set will consist of the union of these three sets:

For C_{12},

$$\{z \in C : |z-a_{11}|\cdot|z-a_{22}| \le r_1(A)\cdot r_2(A)\} = \{z \in C : |z-.5|\cdot|z-i| \le (1)\cdot(.75+.25)\}.$$

For C_{34},

$$\{z \in C : |z-a_{33}|\cdot|z-a_{44}| \le r_3(A)\cdot r_4(A)\} = \{z \in C : |z-(-.5)|\cdot|z-(-i)| \le (.25)\cdot(1+.3)\}.$$

For C_{1234},

$$\{z \in C : |z-a_{11}|\cdot|z-a_{22}|\cdot|z-a_{33}|\cdot|z-a_{44}| \le r_1(A)\cdot r_2(A)\cdot r_3(A)\cdot r_4(A)\}$$
$$= \{z \in C : |z-.5|\cdot|z-i|\cdot|z-(-.5)|\cdot|z-(-i)| \le (1)\cdot(.75+.25)\cdot(.25)\cdot(1+.3)\}.$$

The Brualdi set for this matrix is shown in figure 4.12B.

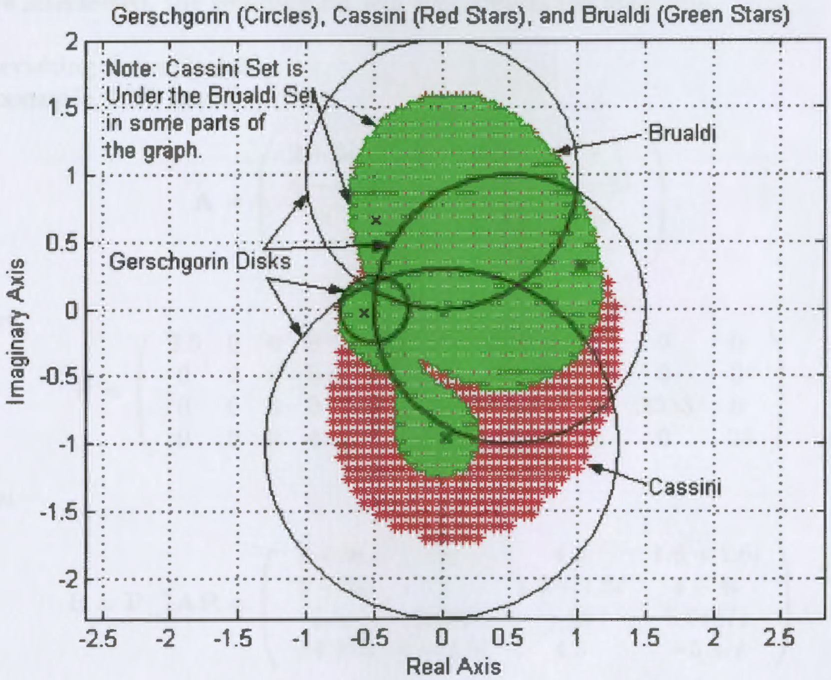


Figure 4.12B

In these figures, the eigenvalues are represented by the X's

The Brualdi sets are of great theoretical value and the ideas derived from the Brualdi theorem are likely to lead to other discoveries. However, in its present state, the Brualdi Theorem is not practical for large matrices because the number of cycles that must be considered increase **factorially** with the dimension of the matrix. This can lead to a huge number of calculations. Therefore, only limited use will be made of Brualdi's Theorem in this thesis.

Section 4.4 Minimal Gerschgorin set

Another approach to spectral inclusion is to take advantage of the fact that similarity transformations preserve the spectrum. Given a square, complex matrix A, the Gerschgorin disks of A, by Theorem 1.27, contain the spectrum of A. If a matrix B is similar to A, the Gerschgorin disks of B contain the spectrum of B and, by virtue of similarity, the spectrum of B is the spectrum of A. Now, since the Gerschgorin disks of A contain the spectrum of A and the Gerschgorin disks of B contain the spectrum of A, then the intersection of the Gerschgorin disks of A and the Gerschgorin disks of B contain the spectrum of A. Therefore, even if the Gerschgorin sets for a large number of similar matrices are intersected, the resulting set will still contain the spectrum.

Revisiting Example 1.29,
Example 4.13 Recall

$$\mathbf{A} = \begin{pmatrix} 2+3i & i & 4 & i+1 \\ 4-4i & 2 & 1+i & 2+2i \\ 3i & 4 & -4i & 5i \\ -7 & 2-5i & 6 & -5+i \end{pmatrix}.$$

Let

$$\mathbf{P} = \begin{pmatrix} 2.5 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix} \qquad \mathbf{P^{-1}} = \begin{pmatrix} .4 & 0 & 0 & 0 \\ 0 & .5 & 0 & 0 \\ 0 & 0 & .3333 & 0 \\ 0 & 0 & 0 & .25 \end{pmatrix}.$$

Let

$$\mathbf{B} = \mathbf{P^{-1}AP} = \begin{pmatrix} 2+3i & .8i & 4.8 & 1.6+1.6i \\ 5-5i & 2 & 1.5+1.5i & 4+4i \\ 2.5i & 2.667 & -4i & 6.6667i \\ -4.375 & 1-2.5i & 4.5 & -5+i \end{pmatrix}.$$

Now, if we continue to create new matrices P by changing the values of the diagonal elements (using only strictly positive values) and continue to find the Gerschgorin sets for the similar matrices associated with P and intersect these sets, we will get smaller and smaller inclusion sets. The set resulting from such intersections is shown in blue in figure 4.13. Notice how small this set is compared to the Gerschgorin disks for the original matrix A. The new set is even smaller than the Cassini set. This set is called the **Minimal Gerschgorin set** for the matrix A. (See Varga [87], [91], [89],[46] Johnson [39], and Levinger [47] for important work on the Minimal Gerschgorin Set).

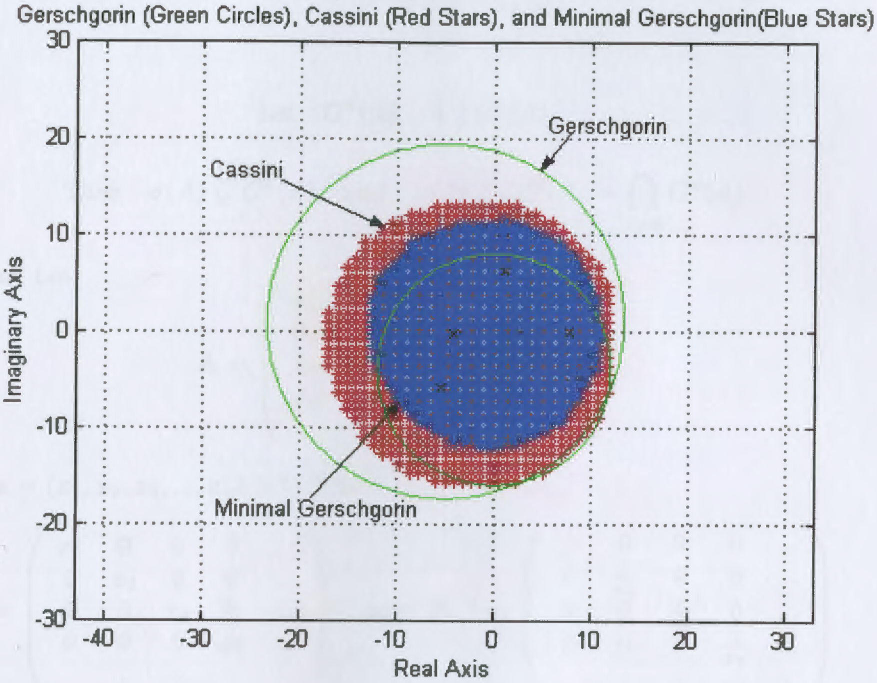


Figure 4.13

In these figures, the eigenvalues are represented by the X's

The idea used in the previous example is generalized in the next theorem.

Theorem 4.14 (Minimal Gerschgorin set) Let $A \in C^{nxn}$.
Let $\mathbf{x} = (x_1, x_2, x_3, ..., x_n) > 0$. Let

$$G_i^{\mathbf{x}}(A) = \{z \in C : |z - a_{ii}| \leq r_i^{\mathbf{x}}(A) = \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{|a_{ij}|x_j}{x_i}\} \quad mbox for \quad 1 \leq i \leq n.$$

$$\text{Let} \quad G^{\mathbf{x}}(A) = \bigcup_{i=1}^{n} G_i^{\mathbf{x}}(A).$$

$$\text{Then} \quad \sigma(A) \subseteq G^{\mathbf{x}}(A) \quad \text{and} \quad \sigma(A) \subseteq G^{\mathbf{R}}(A) = \bigcap_{\mathbf{x}>0} G^{\mathbf{x}}(A).$$

Proof Let

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & ... \\ a_{21} & a_{22} & a_{23} & a_{24} & ... \\ a_{31} & a_{32} & a_{33} & a_{34} & ... \\ a_{41} & a_{42} & a_{43} & a_{44} & ... \\ . & . & . & . & . \end{pmatrix}.$$

Let $\mathbf{x} = (x_1, x_2, x_3, ..., x_n) > 0$. Then

$$\mathbf{P} = \begin{pmatrix} x_1 & 0 & 0 & 0 & ... \\ 0 & x_2 & 0 & 0 & ... \\ 0 & 0 & x_3 & 0 & ... \\ 0 & 0 & 0 & x_4 & ... \\ . & . & . & . & . \end{pmatrix} \quad \text{and} \quad \mathbf{P^{-1}} = \begin{pmatrix} \frac{1}{x_1} & 0 & 0 & 0 & ... \\ 0 & \frac{1}{x_2} & 0 & 0 & ... \\ 0 & 0 & \frac{1}{x_3} & 0 & ... \\ 0 & 0 & 0 & \frac{1}{x_4} & ... \\ . & . & . & . & . \end{pmatrix}.$$

Let

$$\mathbf{B} = \mathbf{P^{-1}AP} = \begin{pmatrix} \frac{a_{11}x_1}{x_1} & \frac{a_{12}x_2}{x_1} & \frac{a_{13}x_3}{x_1} & \frac{a_{14}x_4}{x_1} & ... \\ \frac{a_{21}x_1}{x_21} & \frac{a_{22}x_2}{x_2} & \frac{a_{23}x_3}{x_2} & \frac{a_{24}x_4}{x_2} & ... \\ \frac{a_{31}x_1}{x_3} & \frac{a_{32}x_2}{x_3} & \frac{a_{33}x_3}{x_3} & \frac{a_{34}x_4}{x_3} & ... \\ \frac{a_{41}x_1}{x_4} & \frac{a_{42}x_2}{x_4} & \frac{a_{43}x_3}{x_4} & \frac{a_{44}x_4}{x_4} & ... \\ . & . & . & . & . \end{pmatrix}.$$

Then B is similar to A. Since similar matrices have the same spectrum, the Gerschgorin set of B contains the spectrum of A. The Gerschgorin set of B is given by:

$$G^{\mathbf{x}}(A) = G(B) = \bigcup_{i=1}^{n} G_i^{\mathbf{x}}(A).$$

Where

$$G_i^{\mathbf{x}}(A) = G_i(B) = \{z \in C : |z - a_{ii}| \leq r_i^{\mathbf{x}}(A) = \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{|a_{ij}|x_j}{x_i}\} \quad \text{for} \quad 1 \leq i \leq n.$$

So, the above is the Gerschgorin set for B. That is, the Gerschgorin set for a particular similarity transformation based on a particular $\mathbf{x}$. Note that as $\mathbf{x}$ is changed, new similar matrices will be formed. The Gerschgorin set of each of these similar matrices contain the spectrum of A. That is,

$$\sigma(A) \subseteq G^{\mathbf{x}}(A).$$

Therefore, the intersection of all of these Gerschgorin sets contain the spectrum of A:

$$\sigma(A) \subseteq G^{\mathbf{R}}(A) = \bigcap_{\mathbf{x}>\mathbf{0}} G^{\mathbf{x}}(A).$$

••••

Example 4.15 Let

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & -1 & 2 \\ 1 & 2 & 1 & -1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & .5 & -1 \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{-1.79, 2.17, .81 + 341i, .81 - 341i\}.$$

The minimal Gerschgorin set for this matrix is plotted in figure 4.15. For comparison purposes, the Brauer-Cassini set and the basic Gerschgorin set are also shown. The minimal Gerschgorin set consists of points represented by the blue stars. The Cassini set consists of the points represented by the red stars. The basic Gerschgorin set is the set enclosed by the union of the green circles. Notice that the minimal Gerschgorin set is a subset of Brauer-Cassini. (It is true in general that the minimal Gerschgorin set is a subset of Brauer-Cassini. See Varga [88] for the proof.)

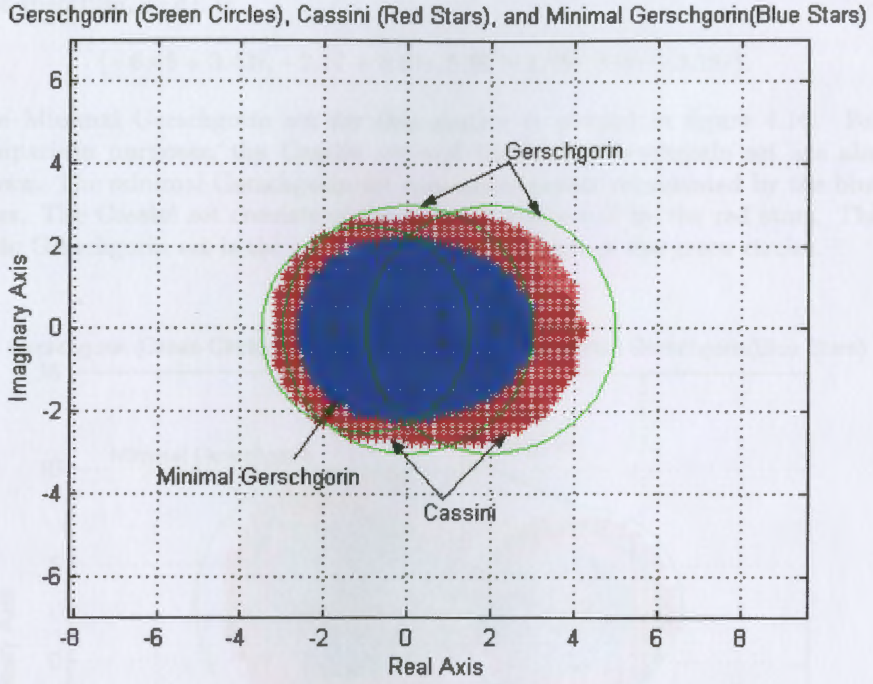


Figure 4.15

In these figures, the eigenvalues are represented by the X's

Revisiting Example 4.5,
Example 4.16 Recall

$$\mathbf{A} = \begin{pmatrix} -4+4i & -2 & 3 & -1 \\ 1 & 6+2i & -2 & 2 \\ 2 & 2 & -6+3i & 3 \\ .5 & .5 & -4 & 7+3i \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{-6.46+3.43i, -2.17+3.61i, 5.67+1.78i, 5.96+3.18i\}.$$

The Minimal Gerschgorin set for this matrix is plotted in figure 4.16. For comparison purposes, the Cassini set and the basic Gerschgorin set are also shown. The minimal Gerschgorin set consists of points represented by the blue stars. The Cassini set consists of the points represented by the red stars. The basic Gerschgorin set is the set enclosed by the union of the green circles.

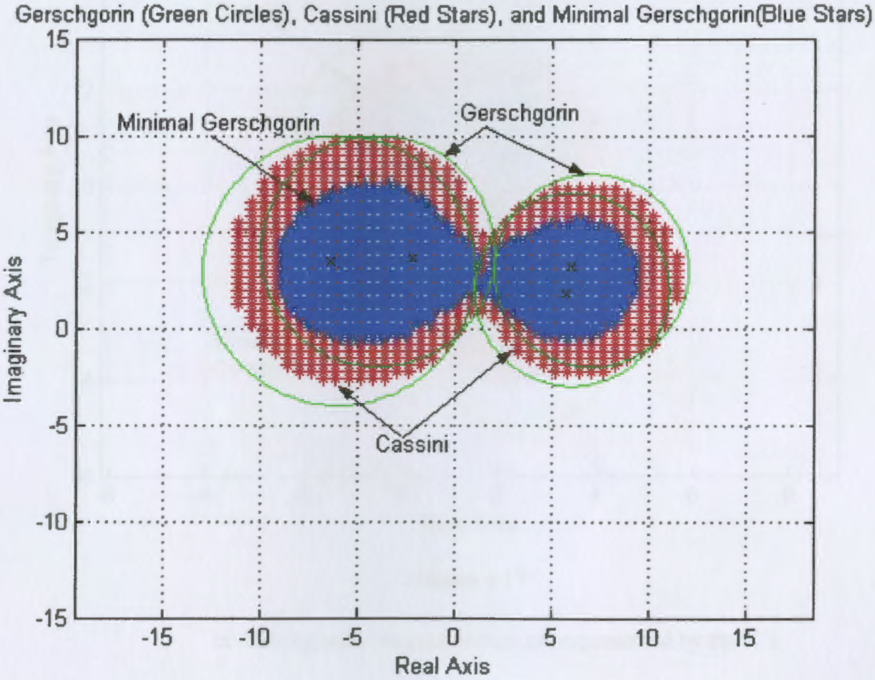


Figure 4.16

In these figures, the eigenvalues are represented by the X's

Revisiting Example 4.6,

Example 4.17 Recall

$$\mathbf{A} = \begin{pmatrix} 2.75+i & 1 & .75 & -.5 \\ -.75 & -i & .5 & .75 \\ .75 & .5 & .1+i & 1 \\ 1 & .5i & -.5i & 2.5-3i \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{2.41 - 3.03i, 2.75 + 1.12i, .36 - 1.08i, -.17 + .99i\}.$$

The minimal Gerschgorin set for this matrix is shown in figure 4.17. The Cassini set consists of the points enclosed by the red stars. The Gerschgorin set consists of the points enclosed by the green circles.

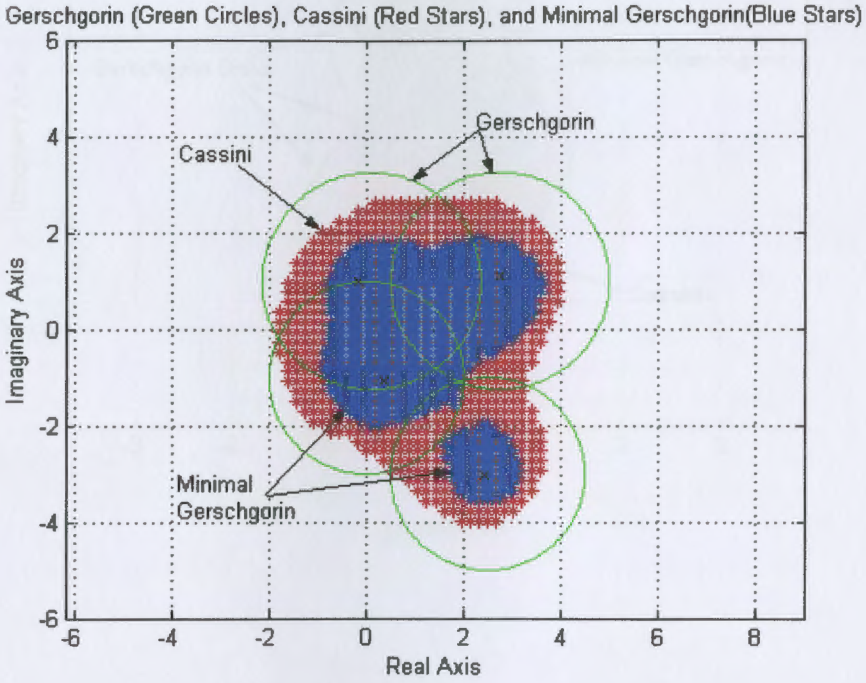


Figure 4.17

In these figures, the eigenvalues are represented by the X's

In each of these cases, the minimal Gerschgorin set was smaller than the Cassini set. As noted earlier, Brualdi's theorem will not be used much in this thesis. However, one example is in order to illustrate the finding of Varga [88] that the minimal Gerschgorin set is a subset of not only Cassini but also of the Brualdi set.

Example 4.18 Let A be the matrix of example 4.17. figure 4.18 compares Gerschgorin, Cassini, Brualdi, and minimal Gerschgorin for this matrix. Note that the minimal Gerschgorin set (represented by the blue stars) is a subset of the other three.

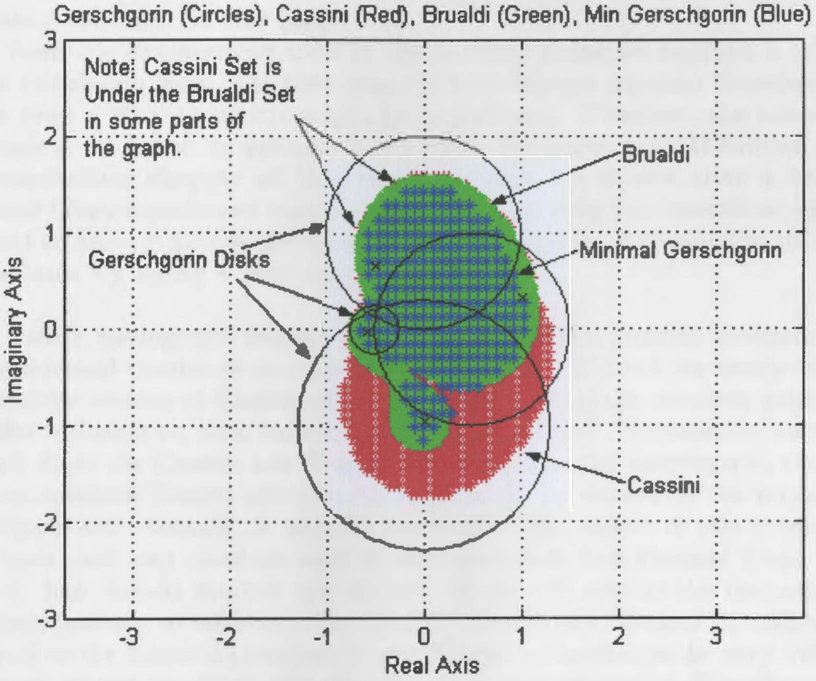


Figure 4.18

So, the minimal Gerschgorin set is a relatively small inclusion set but its production does not come without a price! First of all, to produce a truly *minimal* Gerschgorin set, in general, requires an *infinite* number of similarity transformations. Since this is not possible, some concessions must be made. That is, some increment must be chosen that will be used in the production a finite number similar matrices. If the increments chosen are too large, the set produced may be as large or larger than the Cassini set. On the other hand, when an increment is chosen so that the set produced is a subset of Cassini, the calculation time may be very long.

Even the 4x4 matrices used in the previous examples required a great deal of calculation time. The time required to produce a minimal Gerschgorin set for even a 10 x 10 matrix might be prohibitive. Therefore, the minimal Gerschgorin set is not, in general, a practical, numerical tool. **However, in the concluding chapter of this thesis, it will be shown that a *truly* minimal Gerschgorin set can be produced for Toeplitz matrices, that is equal to the set produced by an infinite number of similarity transformations, by using a new approach.**

Before leaving this chapter we should answer the obvious question: is there a minimal Cassini set or a minimal Brualdi set? Will not the intersection of an infinite number of Cassini sets or Brualdi sets of similar matrices produce a smaller inclusion set than the minimal Gerschgorin set? The intuitive answer is 'Yes': Since the Cassini and Brualdi sets are subsets of Gerschgorin, therefore, the minimal Cassini and Brualdi sets should be subsets of the minimal Gerschgorin set. Actually, it was the intention of the author of this thesis to investigate that very question until it was discovered that Richard Varga (of course!) had already studied the matter. Varga [88] proved the fascinating fact that, contrary to intuition, the minimal Cassini and minimal Brualdi sets are equal to the minimal Gerschgorin set! **Varga's conclusion is very valuable and shows us that, for the moment, the minimal Gerschgorin set is the smallest spectral inclusion set that can be produced with Gerschgorin-type analysis.**

5 The Numerical Range

The numerical range apparently was discovered early in the twentieth century. However, it has only been since the emergence of computers and efficient algorithms that the numerical range has become a viable method for practical applications. Extensive research during the past 10-15 years has made the numerical range even more attractive for Numerical applications. People such as Christiane Tretter, Markus Wagenhofer [83], K.E. Gustafson, D.K.M. Rao [29], and Anne Greenebaum [28] have developed algorithms that significantly sharpen the numerical range inclusion set.

Still, the numerical range cannot be considered a 'simple' method for producing spectral inclusion sets of a matrix or operator because it may take several minutes to calculate its inclusion set for large matrices or operators. Therefore, for the purposes of this thesis, the numerical range will be considered one of the more 'involved' methods for producing spectral inclusion sets.

In a limited number of applications, the numerical range is much better than most other methods. The numerical range is particularly powerful when applied to normal matrices. As will be shown, when applied to normal matrices, the numerical range inclusion set is small and very often, a straight line!

The definition of the numerical range and its spectral inclusion Theorem are stated as follows:

Let $A \in C^{nxn}$. Then the **Numerical Range** is defined as $W(A) = \{\langle A\mathbf{x}, \mathbf{x}\rangle : \mathbf{x} \in C^n \text{ and } ||x|| = 1\}$

Theorem 5.1 (The Numerical Range is a spectral inclusion set) Let $A \in C^{nxn}$. Then $\sigma(A) \subset W(A)$.

Proof Let $\lambda \in \sigma(A)$. Let $\mathbf{x} \in C^n$ such that $||\mathbf{x}|| = 1$. Then $\langle A\mathbf{x}, \mathbf{x}\rangle = \langle \lambda\mathbf{x}, \mathbf{x}\rangle = \lambda\langle \mathbf{x}, \mathbf{x}\rangle = \lambda$. Therefore, $\lambda \in W(A)$ and $\sigma(A) \subset W(A)$.
● ● ●●

A number of examples are given below. The numerical range is calculated in all of these examples using a Matlab program written by Carl C. Cowen (Purdue University) and Elad Harel. The Matlab code may be found in the appendix of this thesis.

Example 5.2 Let

$$\mathbf{A} = \begin{pmatrix} 3i & 3-2i & -5 & -7+4i \\ 3-2i & -7i & 2-11i & -8 \\ -5 & 2-11i & 11i & 5+5i \\ -7+4i & -8 & 5+5i & -5i \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$\{-2.4879 + 12.4501i, 5.3584 - 11.5066i, -7.6868 - 2.043i, 4.8164 + 3.0994i\}$.

The numerical range of this matrix is given in figure 5.2.

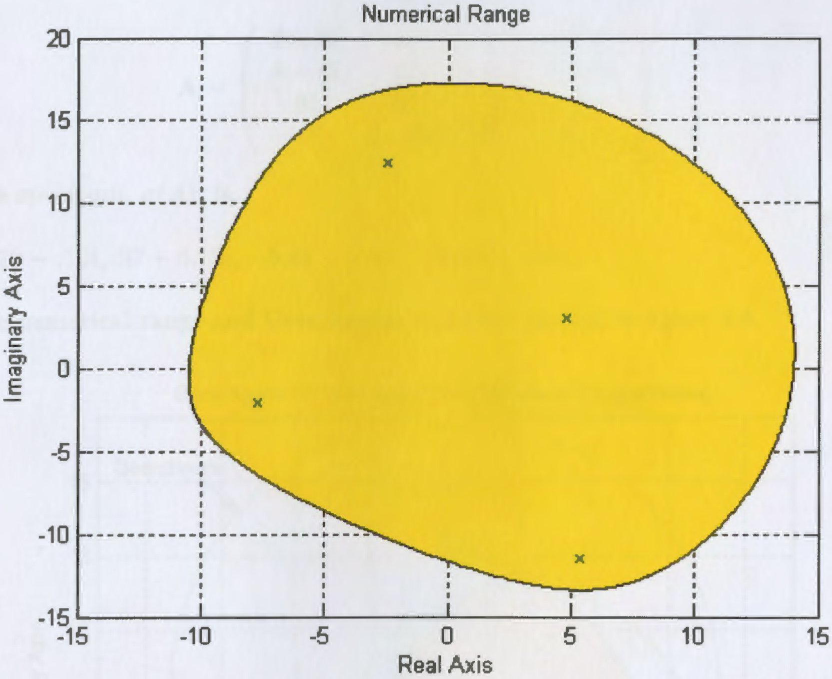


Figure 5.2

The Matlab code to generate the numerical range in this figure was written by Cowen and Harel

Notice that the inclusion set is convex. In fact, **the numerical range always produces a convex set.** This fact is stated in the next theorem.

Theorem 5.3 (Toeplitz-Hausdorff) The numerical range of a matrix is a convex set.

Proof The formal proof can be found in Gustafson and Rao [29].

A number of matrices that were studied in previous chapters are examined here using the numerical range. In each of the following examples, the numerical range is plotted along with the Gerschgorin disks. A wide range of examples are presented here in order to show that sometimes the numerical range will perform better than the Gerschgorin disks and at other times the Gerschgorin disks will perform better than the numerical range.

Revisiting Example 1.29,
Example 5.4 Recall

$$\mathbf{A} = \begin{pmatrix} 2+3i & i & 4 & i+1 \\ 4-4i & 2 & 1+i & 2+2i \\ 3i & 4 & -4i & 5i \\ -7 & 2-5i & 6 & -5+i \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$\{7.79 - .12i, .97 + 6.19i, -5.47 - 5.89i, -4.29 - .18i\}$.

The numerical range and Gerschgorin disks are plotted in figure 5.4.

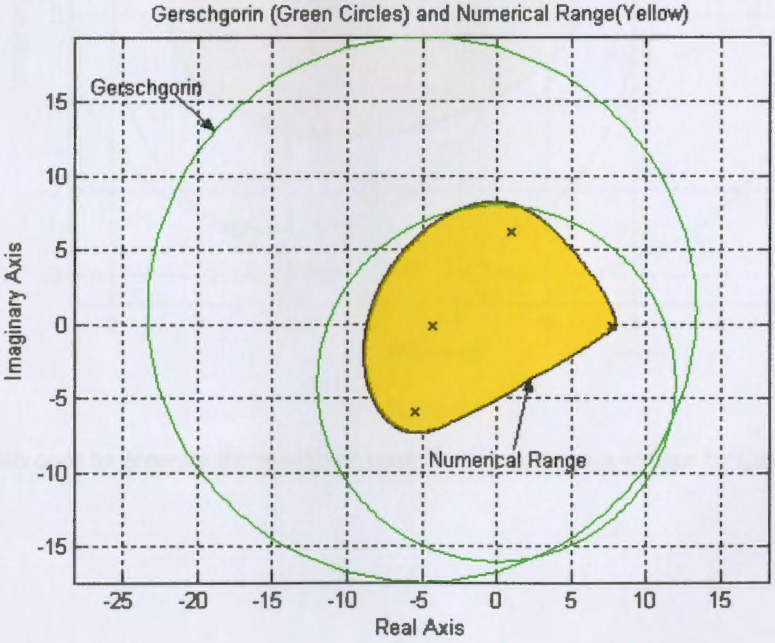


Figure 5.4

The Matlab code to generate the numerical range in this figure was written by Cowen and Harel

Revisiting Example 1.2,
Example 5.5 Let

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & -1 & 2 \\ 1 & 2 & 1 & -1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & .5 & -1 \end{pmatrix}.$$

The spectrum of this matrix is:

$\{-1.79, 2.17, .81 + 341i, .81 - 341i\}$.

The numerical range and Gerschgorin disks are plotted in figure 5.5.

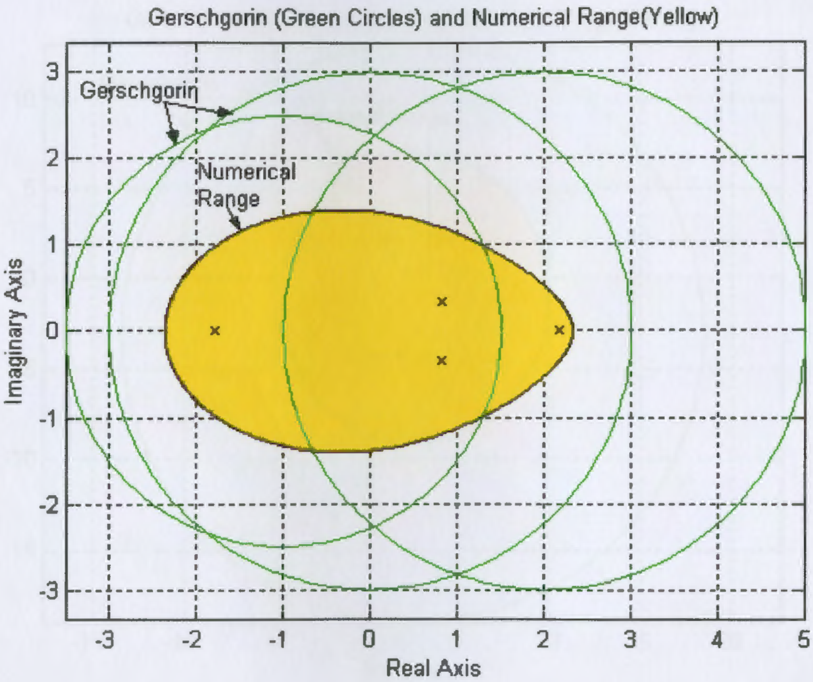


Figure 5.5

The Matlab code to generate the numerical range in this figure was written by Cowen and Harel

Revisiting Example 4.6,
Example 5.6 Recall

$$\mathbf{A} = \begin{pmatrix} 2.75+i & 1 & .75 & -.5 \\ -.75 & -i & .5 & .75 \\ .75 & .5 & .1+i & 1 \\ 1 & .5i & -.5i & 2.5-3i \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$\{2.41 - 3.03i, 2.75 + 1.12i, .36 - 1.08i, -.17 + .99i\}$.

The numerical range and Gerschgorin disks are plotted in figure 5.6.

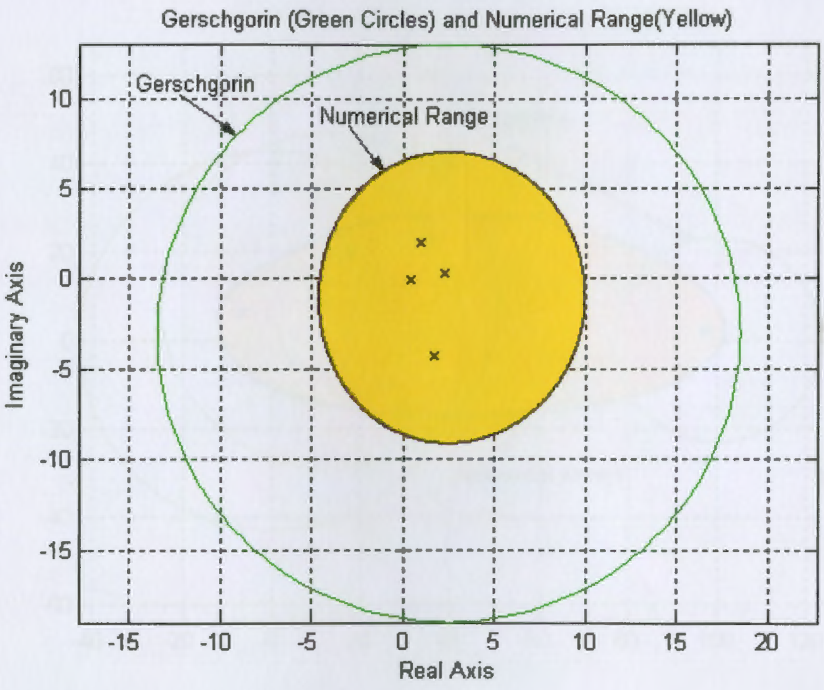


Figure 5.6

The Matlab code to generate the numerical range in this figure was written by Cowen and Harel

Revisiting Example 1.4,

Example 5.7 Recall

$$\mathbf{A} = \begin{pmatrix} 100 & 2 & -6 & 15i \\ 7-6i & 15i & -7+3i & 8+5i \\ 19 & 8-4i & 13+9i & 10 \\ 16+9i & 15-2i & -9+3i & 0 \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$\{97.33 + 2.07i, 17.51 + 20.11i, 5.14 - 4.56i, -7.0 + 6.38i\}$.

The numerical range and Gerschgorin disks are plotted in figure 5.7.

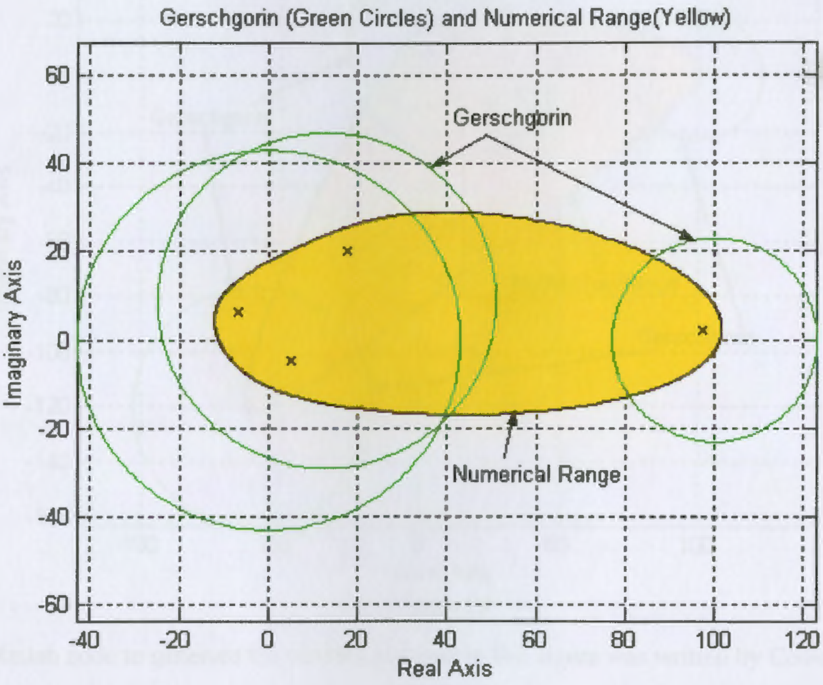


Figure 5.7

The Matlab code to generate the numerical range in this figure was written by Cowen and Harel

Revisiting Example 1.5,

Example 5.8 Recall

$$\mathbf{A} = \begin{pmatrix} 100 & 2 & -6 & 15i \\ 7-6i & 15i & -7+3i & 8+5i \\ 19 & 8-4i & 13+9i & 10 \\ 16+9i & 15-2i & -9+3i & -60-120i \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$\{98.87 + 1.12i, -60.76 - 121.3i, 3.32 + 9.03i, 11.56 + 15.16i\}$.

The numerical range and Gerschgorin disks are plotted in figure 5.8.

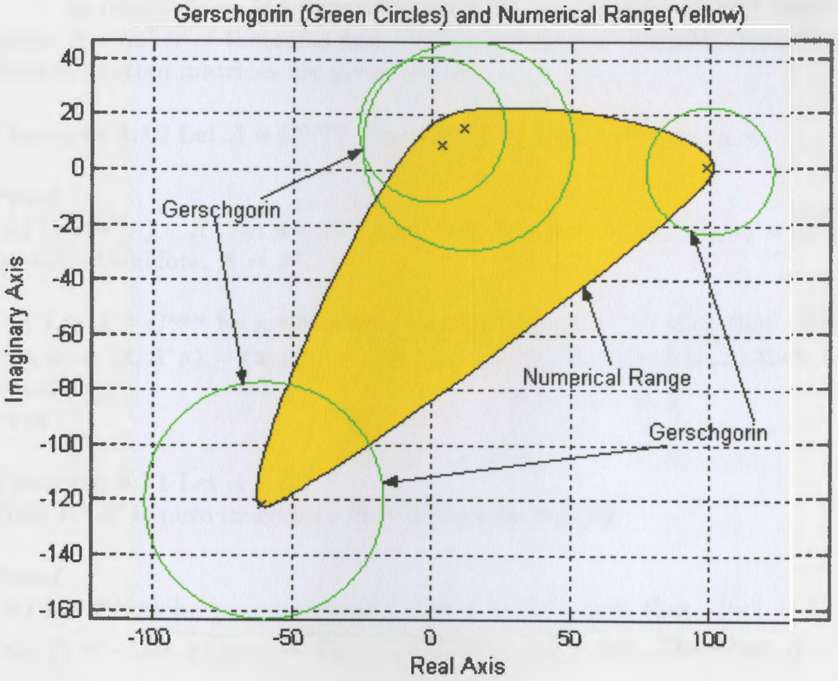


Figure 5.8

The Matlab code to generate the numerical range in this figure was written by Cowen and Harel

A careful study of the graphs presented thus far in this chapter will reveal that the numerical range does produce a convex set. However, the set produced is *not necessarily the convex hull of the spectrum.* (The convex hull of a set of points is the smallest convex set that contains those points). This is particularly noticeable in example 5.2 which was presented earlier.

One of the more interesting theorems related to the numerical range is the following:

Theorem 5.9 Let $A, B \in C^{nxn}$. Then $\sigma(A+B) \subset W(A) + W(B)$.

Proof Let $A, B \in C^{nxn}$. Let $\mathbf{x} \in C^n$ such that $\|\mathbf{x}\| = 1$. Then by Theorem 5.1, $\sigma(A+B) \subset W(A+B)$. By definition, $W(A+B) = \{\langle (A+B)\mathbf{x}, \mathbf{x}\rangle : \mathbf{x} \in C^n$ and $\|x\| = 1\}$. Now, $\langle (A+B)\mathbf{x}, \mathbf{x}\rangle = \langle (A\mathbf{x} + B\mathbf{x}, \mathbf{x}\rangle = \langle A\mathbf{x}, \mathbf{x}\rangle + \langle B\mathbf{x}, \mathbf{x}\rangle = W(A) + W(B)$. Therefore, $\sigma(A+B) \subset W(A) + W(B)$.

● ● ●●

Theorems related to the numerical range inclusion set

As noted above, the numerical range performs very well with normal matrices. A number of theorems and examples related to normal, Hermitian, and Skew-Hermitian matrices are given below:

Theorem 5.10 Let $A \in C^{nxn}$. Then $W(A) \in R$ iff A is Hermitian.

Proof
(⇒) Let $W(A) \in R$. Let $\mathbf{x} \in C^n$ such that $\|\mathbf{x}\| = 1$. Then $\langle A\mathbf{x}, \mathbf{x}\rangle = \overline{\langle \mathbf{x}, A\mathbf{x}\rangle} = \langle \mathbf{x}, A\mathbf{x}\rangle$. Therefore, $A = A^*$.

(⇐) Let $A \in C^{nxn}$ be a Hermitian matrix. Let $\mathbf{x} \in C^n$ such that $\|\mathbf{x}\| = 1$. $\langle A\mathbf{x}, \mathbf{x}\rangle = \langle \mathbf{x}, A^*\mathbf{x}\rangle = \langle \mathbf{x}, A\mathbf{x}\rangle = \overline{\langle A\mathbf{x}, \mathbf{x}\rangle}$. Since $\langle A\mathbf{x}, \mathbf{x}\rangle = \overline{\langle A\mathbf{x}, \mathbf{x}\rangle}$ then $\langle A\mathbf{x}, \mathbf{x}\rangle$ is real.

● ● ●●

Theorem 5.11 Let $A \in C^{nxn}$.
Then $W(A)$ is pure imaginary iff A is skew-Hermitian.

Proof
(⇒) Let $W(A)$ be pure imaginary. Let $\mathbf{x} \in C^n$ such that $\|\mathbf{x}\| = 1$. Then $\langle A\mathbf{x}, \mathbf{x}\rangle = -\overline{\langle A\mathbf{x}, \mathbf{x}\rangle} = -\overline{\overline{\langle \mathbf{x}, A\mathbf{x}\rangle}} = -\langle \mathbf{x}, A\mathbf{x}\rangle = \langle \mathbf{x}, -A\mathbf{x}\rangle$. Therefore, $A = -A^*$.

(⇐) Let $A \in C^{nxn}$ be a Skew-Hermitian matrix. Let $\mathbf{x} \in C^n$ such that $\|\mathbf{x}\| = 1$. $\langle A\mathbf{x}, \mathbf{x}\rangle = \langle \mathbf{x}, A^*\mathbf{x}\rangle = \langle \mathbf{x}, -A\mathbf{x}\rangle = \overline{\langle -A\mathbf{x}, \mathbf{x}\rangle} = -\overline{\langle A\mathbf{x}, \mathbf{x}\rangle}$. Since $\langle A\mathbf{x}, \mathbf{x}\rangle = -\overline{\langle A\mathbf{x}, \mathbf{x}\rangle}$ then $\langle A\mathbf{x}, \mathbf{x}\rangle$ is pure Imaginary.

● ● ●●

Theorem 5.12 (From Gustafson and Rao [29]) Let $A \in C^{nxn}$ be a normal matrix. Then the extreme points of $W(A)$ are eigenvalues of A.

The numerical range for a normal matrix is, very often, a straight line. In fact, only a normal matrix can produce a numerical range with a straight line as the next theorem states.

Theorem 5.13 Let $A \in C^{nxn}$ such that $W(A)$ is a straight line. Then A is normal.

Proof Let $W(A)$ be a straight line in the complex plane. Let θ be the angle between $W(A)$ and the Real axis. Let $z \in W(A) = \{\langle A\mathbf{x}, \mathbf{x}\rangle : \mathbf{x} \in C^n$ and $\|x\| = 1\}$. Since z is on the line with angle θ, then $e^{-i\theta}z$ is on the Real axis. Therefore, $\{\langle e^{-i\theta}A\mathbf{x}, \mathbf{x}\rangle : \mathbf{x} \in C^n$ and $\|x\| = 1\}$ is a set on the Real axis. That is, $W(e^{-i\theta}A)$ is real.

Now,

$$\langle e^{-i\theta}zI\mathbf{x}, \mathbf{x}\rangle = e^{-i\theta}z\langle I\mathbf{x}, \mathbf{x}\rangle = e^{-i\theta}z\langle \mathbf{x}, \mathbf{x}\rangle = e^{-i\theta}z\langle \mathbf{x}, I\mathbf{x}\rangle = \langle \mathbf{x}, \overline{e^{-i\theta}z}I\mathbf{x}\rangle = \langle \mathbf{x}, e^{-i\theta}zI\mathbf{x}\rangle$$

(The very last equality on the right holds because $e^{-i\theta}z$ is real).
Since $\langle e^{-i\theta}zI\mathbf{x}, \mathbf{x}\rangle = \langle \mathbf{x}, e^{-i\theta}zI\mathbf{x}\rangle$, then $e^{-i\theta}zI$ is Hermitian.
Then by Theorem 5.10, $W(e^{-i\theta}zI)$ is Real.

So, we have shown that $W(e^{-i\theta}A)$ and $W(e^{-i\theta}zI)$ are both real. Therefore, $W(e^{-i\theta}A) + W(e^{-i\theta}zI)$ is real.
Then $W(e^{-i\theta}A - e^{-i\theta}zI)$ is real. $W(e^{-i\theta}(A - zI))$ is real.
Therefore, by Theorem 5.10, $e^{-i\theta}(A - zI)$ is Hermitian. That means, $e^{-i\theta}(A - zI)e^{i\theta}(A^* - \overline{z}I) = e^{i\theta}(A^* - \overline{z}I)e^{-i\theta}(A - zI)$.
Therefore, $(A^* - \overline{z}I) = (A^* - \overline{z}I)(A - zI)$.
Therefore, $AA^* - \overline{z}A - zA^* + z\overline{z} = A^*A - zA^* - \overline{z}A + \overline{z}z$.
Therefore, $AA^* = A^*A$.

Therefore A is normal.
● ● ●●

Note that **the converse of Theorem 5.13 is not always true!**

Example 5.14 Consider the following matrix. Notice that this matrix is normal but not Hermitian.

$$\mathbf{A} = \begin{pmatrix} 1+i & 0 \\ 0 & 1-i \end{pmatrix}.$$

The numerical range of this matrix is shown in figure 5.14. The numerical range is a straight line. Notice that each end point of the line is an eigenvalue of the matrix. Also notice that the eigenvalues are complex (not pure imaginary or pure real).

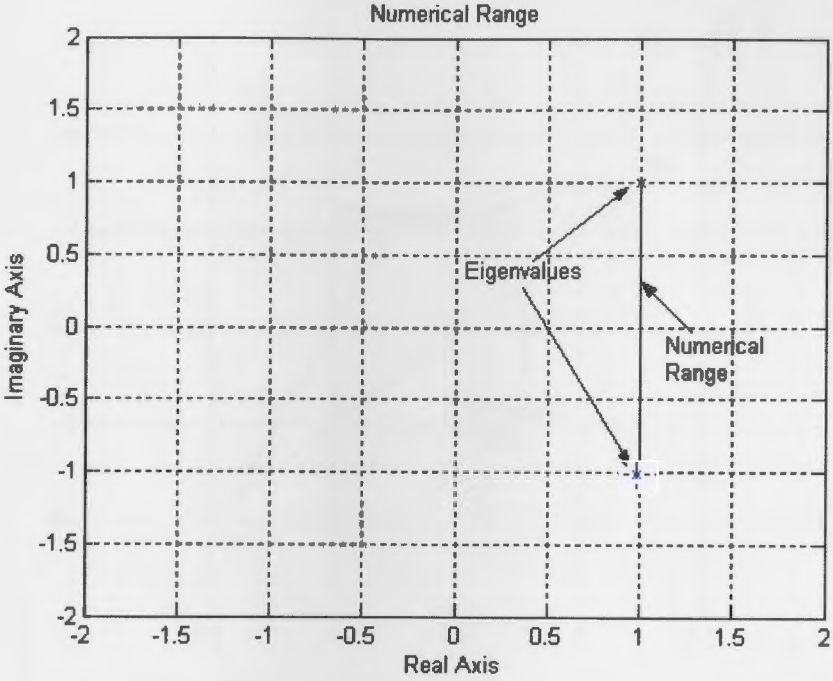


Figure 5.14

Example 5.15 Consider the following normal matrix:

$$\mathbf{A} = \begin{pmatrix} 1 & i \\ 1 & 2+i \end{pmatrix}.$$

The numerical range of this matrix is shown in figure 5.15. This time, the numerical range is a straight line that is not vertical nor horizontal. Again, each end point of the line is an eigenvalue of the matrix.

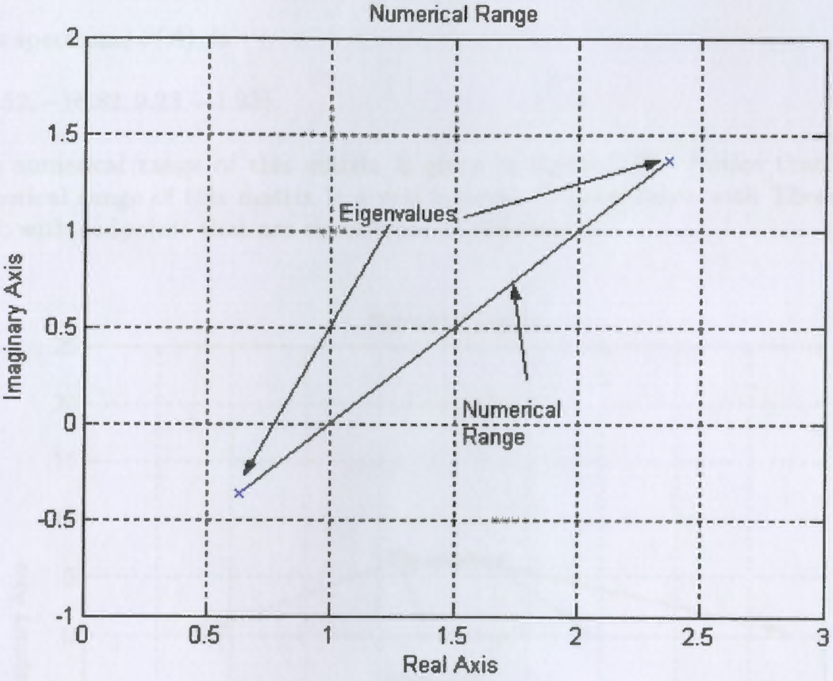


Figure 5.15

Theorem 5.16 (From Gustafson and Rao [29]) The numerical range of a symmetric matrix A is the real interval [m,M] where m and M are the least and greatest eigenvalues of A.

Example 5.17 Consider following Hermitian matrix. Recall that the eigenvalues of a Hermitian matrix are all real.

$$\mathbf{A} = \begin{pmatrix} 3 & 3-2i & -5 & -7+4i \\ 3+2i & -4 & 2-11i & -8 \\ -5 & 2+11i & 1 & 5+5i \\ -7-4i & -8 & 5-5i & 13 \end{pmatrix}.$$

This spectrum, $\sigma(A)$, is

$\{22.52, -16.82, 9.23, -1.93\}$.

The numerical range of this matrix is given in figure 5.17. Notice that the numerical range of this matrix is a real interval, in accordance with Theorem 5.10, with endpoints that are eigenvalues of the matrix.

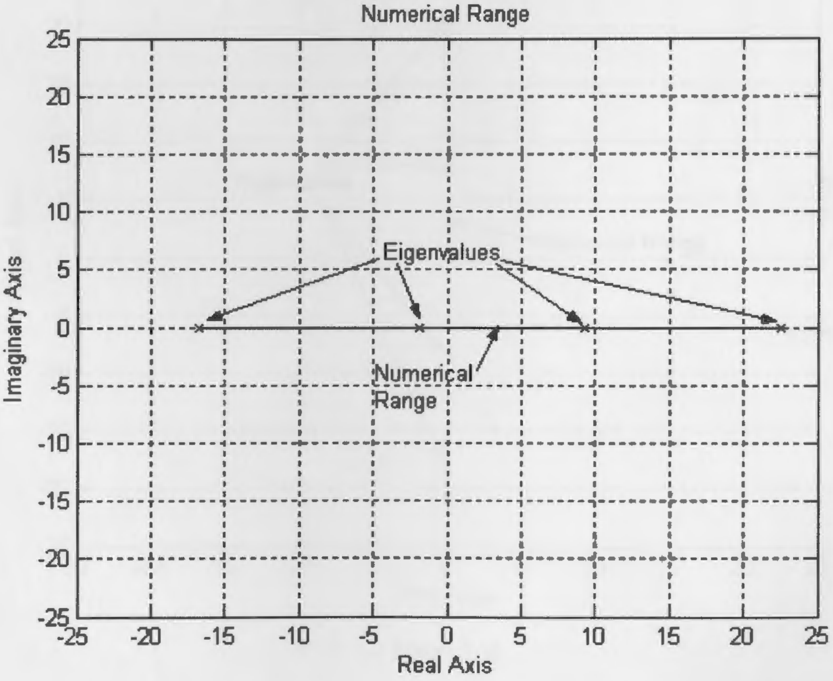


Figure 5.17

Revisiting Example 5.2,

Example 5.18 Consider following skew-Hermitian matrix. Recall that the eigenvalues of a skew-Hermitian matrix are all purely imaginary.

$$\mathbf{A} = \begin{pmatrix} 3i & 3-2i & -5 & 7+4i \\ -3-2i & -7i & 2-11i & 8 \\ 5 & -2-11i & 11i & -5+5i \\ -7+4i & -8 & 5+5i & -5i \end{pmatrix}.$$

The spectrum, $\sigma(A)$ is

$\{20.68i, 7.87i, -18.07i, -8.48i\}$.

The numerical range of this matrix is given in figure 5.18. Notice that the numerical range of this matrix is a vertical line on the imaginary axis, in accordance with Theorem 5.11, with endpoints that are eigenvalues of the matrix.

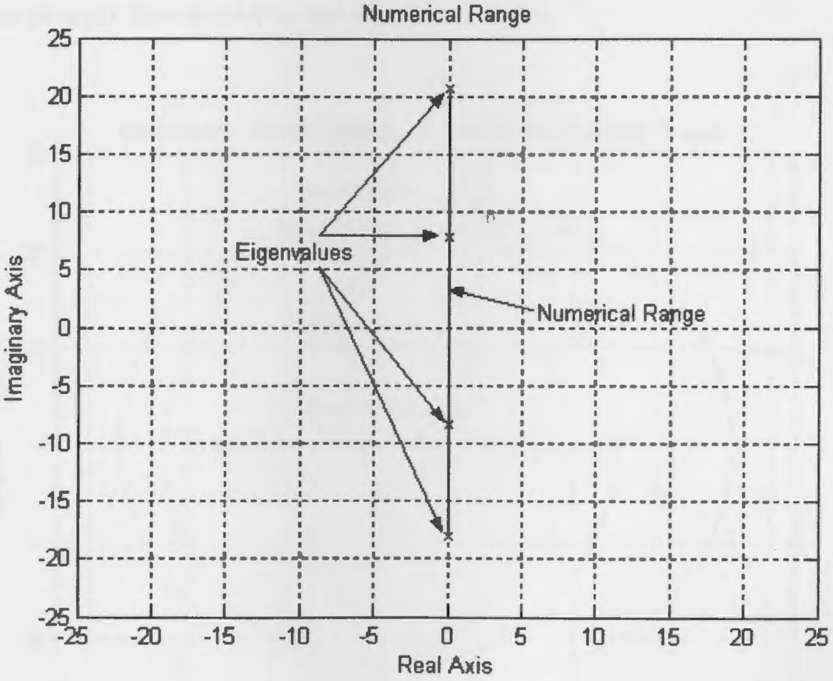


Figure 5.18

The Strength of the Numerical Range

The last series of examples demonstrated the efficiency of the numerical range when applied to normal, Hermitian and skew Hermitian matrices. In the case of such matrices, the numerical range is a very small set and, often, a straight line. No other method of spectral estimation examined thus far comes even close to producing such a small set. A couple of examples will illustrate this fact.

Revisiting Example 5.17,
Example 5.19 Consider, once again, the Hermitian matrix,

$$\mathbf{A} = \begin{pmatrix} 3 & 3-2i & -5 & -7+4i \\ 3+2i & -4 & 2-11i & -8 \\ -5 & 2+11i & 1 & 5+5i \\ -7-4i & -8 & 5-5i & 13 \end{pmatrix}.$$

The numerical range (the black line) is shown along with the Gerschgorin disks in figure 5.19. The set created by the Gerschgorin disks looks massive compared to the straight line created by the numerical range.

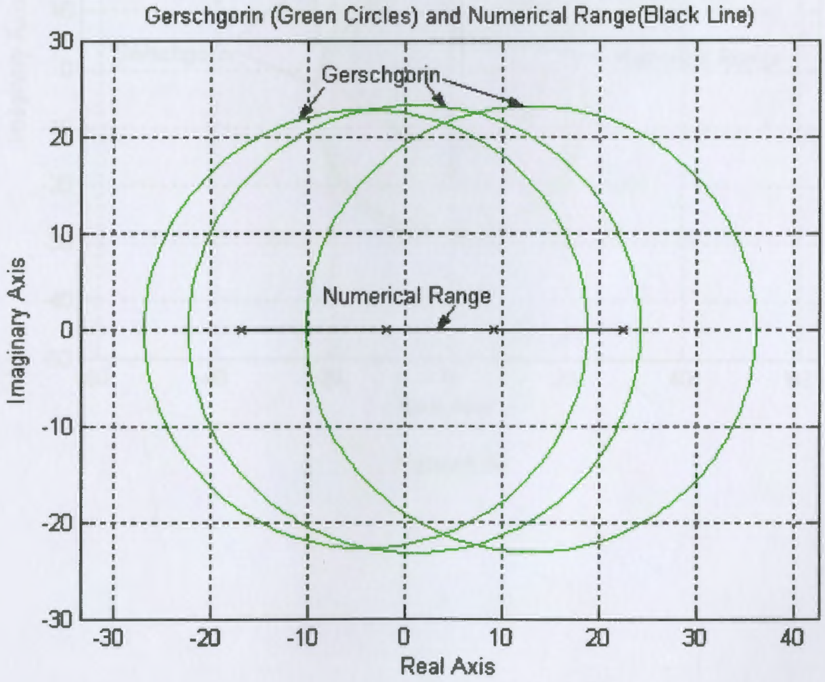


Figure 5.19

Revisiting Example 5.2,

Example 5.20 Take another look at the skew-Hermitian matrix,

$$\mathbf{A} = \begin{pmatrix} 3i & 3-2i & -5 & 7+4i \\ -3-2i & -7i & 2-11i & 8 \\ 5 & -2-11i & 11i & -5+5i \\ -7+4i & -8 & 5+5i & -5i \end{pmatrix}.$$

The numerical range (the black line) is shown along with the Gerschgorin disks in figure 5.20.

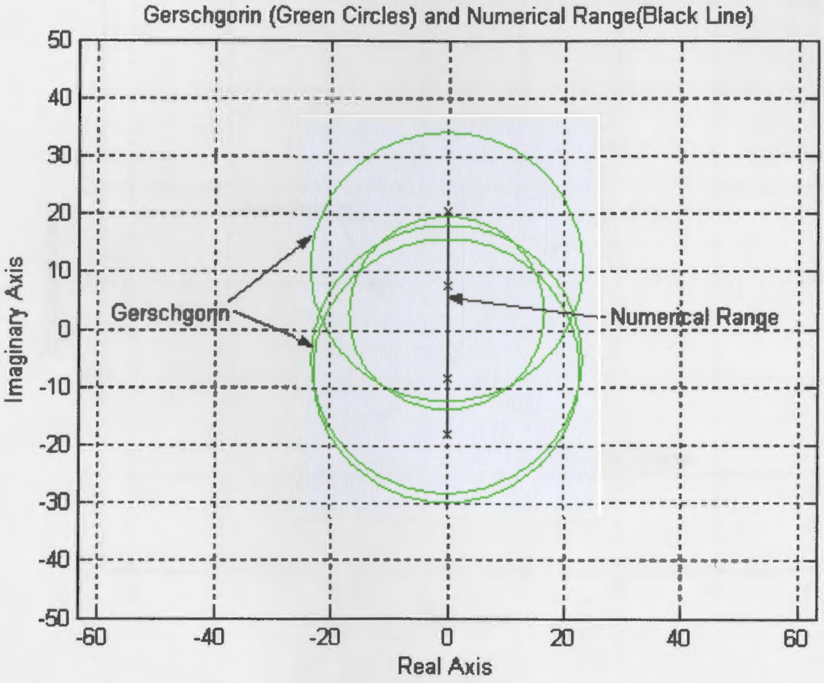


Figure 5.20

Example 5.21 Consider the normal matrix,

$$\mathbf{A} = \begin{pmatrix} 1 & i \\ 1 & 2+i \end{pmatrix}.$$

The numerical range (the black line) is shown along with the Gerschgorin disks in figure 5.21.

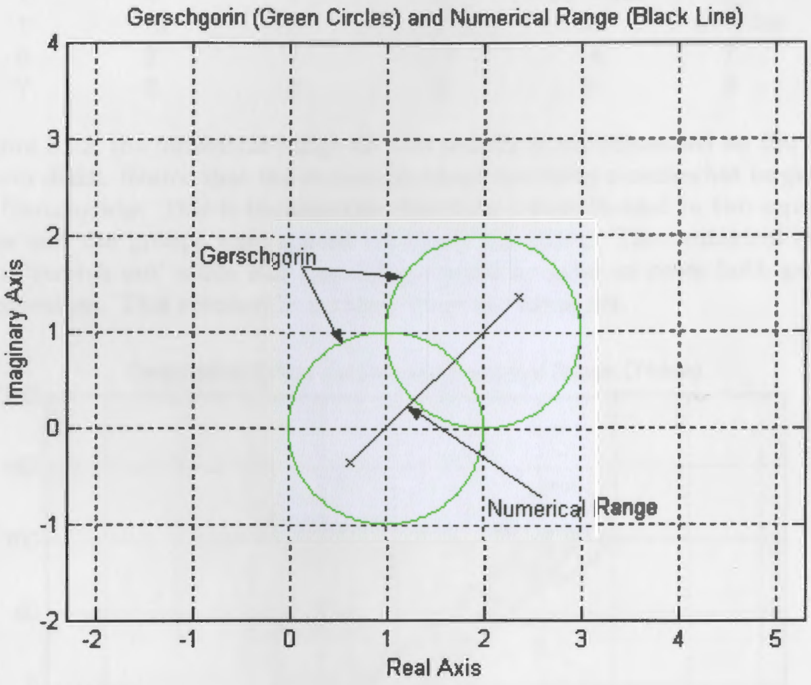


Figure 5.21

In all of these examples, the numerical range is nothing less than outstanding. In all of these cases, it produces a short, straight line while Gerschgorin produces relatively large disks. So, the numerical range works very efficiently with normal, Hermitian, and skew-Hermitian matrices.

The Weakness of the Numerical Range

When the matrix under consideration is not normal the performance of the numerical range is mixed ranging from excellent to very poor. The next two examples will illustrate some poorer performances by the numerical range.

Example 5.22 Consider the following matrix, A.

$$\begin{pmatrix} 95+77i & 3 & 2 & 6 & 8 & 2 & 4 & 7 \\ 1 & 105+97i & 4 & 5 & -6 & 3 & 2 & 8 \\ 3 & 4 & 110+105i & 8 & -2 & 5 & 9 & 1 \\ 5 & 7 & 8 & 115+108i & 4 & -2 & 2 & 4 \\ 2 & 4 & 5 & 5 & -55-60i & 1 & 3 & 2 \\ 1 & -3 & 6 & 7 & 2 & -65-59i & 4 & 5 \\ 6 & 2 & 4 & -6 & -4 & 7 & -72-65i & 4 \\ 7 & 3 & 3 & 4 & 6 & 8 & -9 & -80-62i \end{pmatrix}$$

In figure 5.22, the numerical range for this matrix is superimposed on the Gerschgorin disks. Notice that the numerical range produces a somewhat larger set than Gerschgorin. This is because the eigenvalues were located in two separate groups and the groups were spaced relatively far apart. The numerical range had to 'stretch out' while still remaining convex in order to cover both groups of eigenvalues. This resulted in a rather large inclusion set.

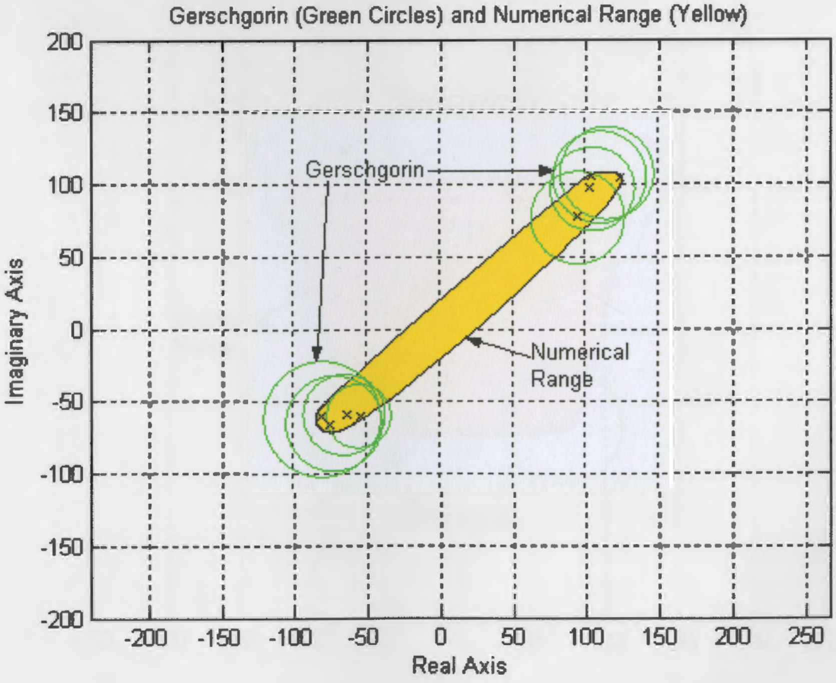


Figure 5.22

The Matlab code to generate the numerical range in this figure was written by Cowen and Harel

Things get even worse for the numerical range when the eigenvalues form three or more groups that are relatively far apart from each other. In such cases the numerical range produces an unreasonably large set.This is illustrated in the next example.

Example 5.23 Consider the following matrix, A.

$$\begin{pmatrix} 95+77i & 3 & 2 & 6 & 8 & 2 & 4 & 7 \\ 1 & -105+99i & 4 & 5 & -6 & 3 & 2 & 8 \\ 3 & 4 & 110+105i & 8 & -2 & 5 & 9 & 1 \\ 5 & 7 & 8 & -115+110i & 4 & -2 & 2 & 4 \\ 2 & 4 & 5 & 5 & -55-60i & 1 & 3 & 2 \\ 1 & -3 & 6 & 7 & 2 & 65-60i & 4 & 5 \\ 6 & 2 & 4 & -6 & -4 & 7 & 72-65i & 4 \\ 7 & 3 & 3 & 4 & 6 & 8 & -9 & -80-62i \end{pmatrix}$$

In figure 5.23 the numerical range for this matrix is superimposed on the Gerschgorin disks. This time the numerical range produces a huge set. The set is large due to the nature of the convex set - in order for the set to be convex and yet 'catch' all of the eigenvalues, it had to grow into the monstrosity shown in the figure.

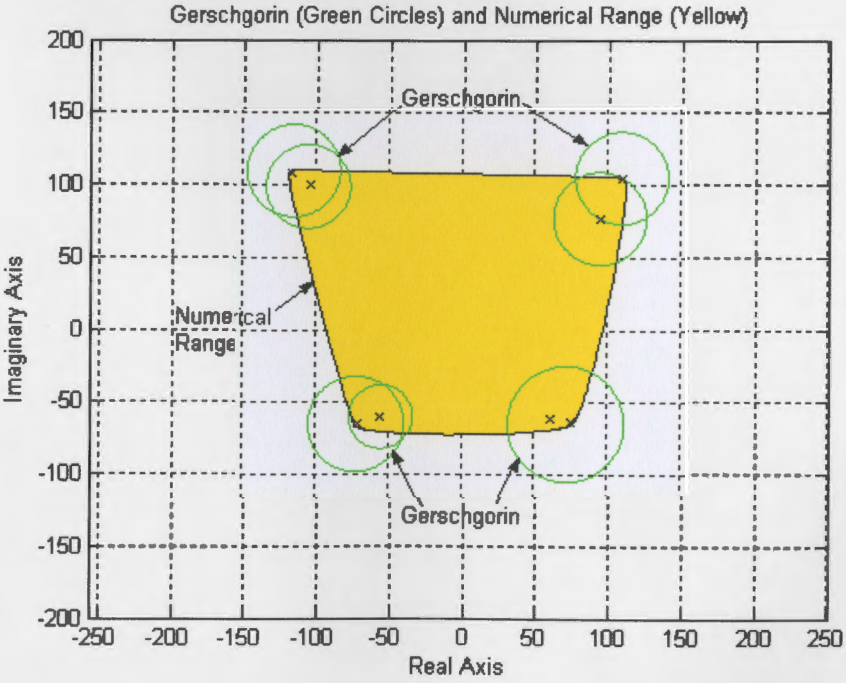


Figure 5.23

The Matlab code to generate the numerical range in this figure was written by Cowen and Harel

Summary

In general the numerical range produces a small spectral inclusion set. The method is very efficient when applied to normal matrices. In such cases the numerical range usually outperforms all of the other methods. When the matrix is not normal the results depend on the specific matrix. When the eigenvalues are grouped relatively close together, the numerical range continues to perform well. However, as the eigenvalues separate into distinct, widely separated groups, the sharpness of the numerical range's set decreases considerably.

The speed required to calculate the numerical range sometimes must be taken into account. In most cases calculation time is not an issue but if the matrix is extremely large it may be wiser to produce a spectral inclusion set by using one of the other methods.

6 Toeplitz Matrices

Toeplitz matrices will be used extensively in the next few chapters. In particular, Toepltiz matrices will be analyzed in the pseudospectra and 'Results' chapters of this thesis. Therefore, it will be useful to investigate some of the features of the spectrum Toepltiz matrices and operators before arriving at those later chapters.

A Toeplitz matrix is a matrix of the form:

$$\mathbf{A} = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & \cdots \\ a_{-1} & a_0 & a_1 & a_2 & \cdots \\ a_{-2} & a_{-1} & a_0 & a_1 & \cdots \\ a_{-3} & a_{-2} & a_{-1} & a_0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}.$$

Notice that within each diagonal all of the elements are the same. This type of matrix arises from many applications in Engineering and Physics. The coefficients of certain systems of Ordinary and Partial Differential Equations can sometimes take the form of a Toeplitz matrix.

As pointed out by Reichel and Trefethen [69], the spectrum for infinite dimensional Toeplitz operators is well understood while the same is not true for finite dimensional operators. In this chapter, both cases will be examined. Our main interest is in the finite dimensional case. However,in a later chapter, we will attempt to use the spectrum of the infinite dimensional operator to help us bound the spectrum of the finite dimensional operator.

Section 6.1 The spectrum for Finite Dimensional Toeplitz matrices

C.D. Meyer points out that the tridiagonal Toeplitz matrix is the only Toeplitz matrix that has explicit formula for its spectrum. The equations quoted by Meyer [58] are noted here:

$$\mathbf{A} = \begin{pmatrix} a_0 & a_1 & 0 & 0 & 0 & 0 & \cdots \\ a_{-1} & a_0 & a_1 & 0 & 0 & 0 & \cdots \\ 0 & a_{-1} & a_0 & a_1 & 0 & 0 & \cdots \\ 0 & 0 & a_{-1} & a_0 & a_1 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

with $a_1 \neq 0$ and $a_{-1} \neq 0$. Then the eigenvalues and eigenvectors are given by,

$$\lambda_j = a_0 + 2a_1\sqrt{\frac{a_{-1}}{a_1}}\cos(\frac{j\pi}{n+1}), \quad \text{and} \quad \mathbf{x_j} = \begin{pmatrix} (\frac{a_{-1}}{a_1})^{\frac{1}{2}}\sin(\frac{1j\pi}{n+1}) \\ (\frac{a_{-1}}{a_1})^{\frac{2}{2}}\sin(\frac{2j\pi}{n+1}) \\ \cdot \\ \cdot \\ \cdot \\ (\frac{a_{-1}}{a_1})^{\frac{n}{2}}\sin(\frac{nj\pi}{n+1}) \end{pmatrix},$$

$$\text{for } j = 1, 2, \ldots n.$$

The following appears to be typical of the available information on other types of Toeplitz matrices:

G. Geldenhuys and C. Sippel [25] were able to come up with a bound on the real eigenvalues for one specific type of Toeplitz matrix, the cascade matrix:

$$\mathbf{A} = \begin{pmatrix} a_0 & -a_1 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots \\ 0 & a_0 & -a_1 & 0 & 0 & 0 & 0 & 0 & \ldots \\ -a_{-2} & 0 & a_0 & -a_1 & 0 & 0 & 0 & 0 & \ldots \\ 0 & -a_{-2} & 0 & a_0 & -a_1 & 0 & 0 & 0 & \ldots \\ 0 & 0 & -a_{-2} & 0 & a_0 & -a_1 & 0 & 0 & \ldots \\ 0 & 0 & 0 & -a_{-2} & 0 & a_0 & -a_1 & 0 & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{pmatrix}.$$

where $a_1, a_{-2} > 0$.

Geldenhuys and Sippel were able to show that any *real* eigenvalue η of A will satisfy

$$a_0 - 3(a_1^2 a_{-2}/4)^{\frac{1}{3}} < \eta \leq a_0.$$

So Geldenhuys and Sippel results are limited to this specific type of Toeplitz matrix and bound only the real eigenvalues of the matrix as the next example shows.

Example 6.1

Consider the 12 x 12 matrix:

$$\mathbf{A} = \begin{pmatrix} 6 & -5 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots \\ 0 & 6 & -5 & 0 & 0 & 0 & 0 & 0 & \ldots \\ -8 & 0 & 6 & -5 & 0 & 0 & 0 & 0 & \ldots \\ 0 & -8 & 0 & 6 & -5 & 0 & 0 & 0 & \ldots \\ 0 & 0 & -8 & 0 & 6 & -5 & 0 & 0 & \ldots \\ 0 & 0 & 0 & -8 & 0 & 6 & -5 & 0 & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{pmatrix}.$$

According to Geldenhuys and Sippel, all real eigenvalues for this matrix will satisfy

$$6 - 3(5^2(8)/4)^{\frac{1}{3}} < \eta \leq 6 \text{ or } -5.052 < \eta \leq 6.$$

The spectrum, $\sigma(A)$, is

$$\{11.22 + 9i, 11.22 - 9i, 10.33 + 7.5i, 10.33 - 7.5i, 8.92 + 5.06i,$$

$$8.92-5.06i, 7.1+1.9i, 7.1-1.9i, -4.44, -2.66, .152, 3.79\}.$$

Notice that the *real* eigenvalues fall within the bounds predicted by the equations. However, Geldenhuys and Sippel's inequality tells us nothing about the complex eigenvalues. This seems to be typical of what is available for predicting the spectrum of finite dimensional Toeplitz matrices - limited results coming from very specific applications. Therefore, there is much work to be done on estimating the spectrum of finite dimensional Toeplitz matrices.

Section 6.2 Spectrum of Infinite Dimensional Toeplitz Operators

Essential for the study of the spectrum of infinite dimensional Toeplitz operators is the function called the *symbol* of the operator.

Given a Toeplitz matrix,

$$\mathbf{A}=\begin{pmatrix} a_0 & a_1 & a_2 & a_3 & \dots \\ a_{-1} & a_0 & a_1 & a_2 & \dots \\ a_{-2} & a_{-1} & a_0 & a_1 & \dots \\ a_{-3} & a_{-2} & a_{-1} & a_0 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

The **symbol** of the matrix is defined as,

$$f(z)=\sum_{i=-1}^{-N} a_i z^i + \sum_{i=1}^{N} a_i z^i.$$

So, we have

$$f(z)=\dots+a_{-2}z^{-2}+a_{-1}z^{-1}+a_0+a_1z+a_2z^2+\dots.$$

Determining symbols for Toeplitz operators is rather easy, as the following examples demonstrate.

Example 6.2 The symbol for the matrix

$$\mathbf{A}=\begin{pmatrix} 0 & 0 & 1 & .7 & 0 & 0 & \dots \\ 2i & 0 & 0 & 1 & .7 & 0 & \dots \\ 0 & 2i & 0 & 0 & 1 & .7 & \dots \\ 0 & 0 & 2i & 0 & 0 & 1 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

is

$$\begin{aligned} f_A &= 2iz^{-1}+0z^0+0z^1+1z^2+.7z^3 \\ &= 2iz^{-1}+z^2+.7z^3. \end{aligned}$$

The following two examples show how symbols of more complicated operators may be put into a compact form by using the geometric series.

Example 6.3 The symbol for the matrix

$$\mathbf{A} = \begin{pmatrix} 1 & \frac{3}{4} & \frac{3}{8} & \frac{3}{16} & \frac{3}{32} & \frac{3}{128} & \cdots \\ 0 & 1 & \frac{3}{4} & \frac{3}{8} & \frac{3}{16} & \frac{3}{32} & \cdots \\ 0 & 0 & 1 & \frac{3}{4} & \frac{3}{8} & \frac{3}{16} & \cdots \\ 0 & 0 & 0 & 1 & \frac{3}{4} & \frac{3}{8} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

is

$$\begin{aligned} f_A &= 1z^0 + \frac{3}{4}z^1 + \frac{3}{8}z^2 + \frac{3}{16}z^3 + \frac{3}{32}z^4 + ... \\ &= 1z^0 + \frac{3}{4}(z^1 + \frac{1}{2}z^2 + \frac{1}{2^2}z^3 + \frac{1}{2^3}z^4 + ...) \\ &= \frac{1+\frac{z}{4}}{1-\frac{z}{2}}. \end{aligned}$$

Example 6.4 The symbol for the matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 2 & 2 & 2 & 2 & \cdots \\ 0 & 1 & 2 & 2 & 2 & 2 & \cdots \\ 0 & 0 & 1 & 2 & 2 & 2 & \cdots \\ 0 & 0 & 0 & 1 & 2 & 2 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

is

$$\begin{aligned} f_A &= 1z^0 + 2z^1 + 2z^2 + 2z^3 + 2z^4 + ... \\ &= 1 + 2(z + z^2 + z^3 + ...) \\ &= \frac{1+z}{1-z}. \end{aligned}$$

Section 6.3 Using the Symbol to calculate the spectrum

Reichel and Trefethen point out that the symbol of an infinite dimensional Toeplitz operator is defined as follows:

Let A be an infinite dimensional Toeplitz operator, S the unit circle in the complex plane centered at the origin, $f(z)$ the symbol of the operator A. Define,

$$I(f(S),\lambda) = \frac{1}{2\pi i}\int_S \frac{f'(z)}{f(z)-\lambda}dz, \quad \text{with} \quad \lambda \notin f(S).$$

Then the spectrum of A is $\Lambda(A) = f(S) \cup \{\lambda \in C : I(f(S),\lambda) \neq 0\}$.

$$\text{Assuming that } \{\lambda \in C : I(f(S), \lambda) \neq 0\} \subseteq f(S).$$

$$\text{Then } \Lambda(A) = f(S).$$

The above implies that by using the unit circle (S) as the domain of the symbol, the range of the symbol will be the spectrum of the infinite dimensional operator. That is, by using, $z = \cos\theta + i\sin\theta$ where $0 \leq \theta \leq 2\pi$ as the domain for the symbol function, the range will be the spectrum.

The following example illustrates this.

Example 6.5 Consider the operator

$$\mathbf{A} = \begin{pmatrix} 0 & 2 & 0 & 0 & 0 & 0 & \dots \\ 1 & 0 & 2 & 0 & 0 & 0 & \dots \\ 0 & 1 & 0 & 2 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & 2 & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

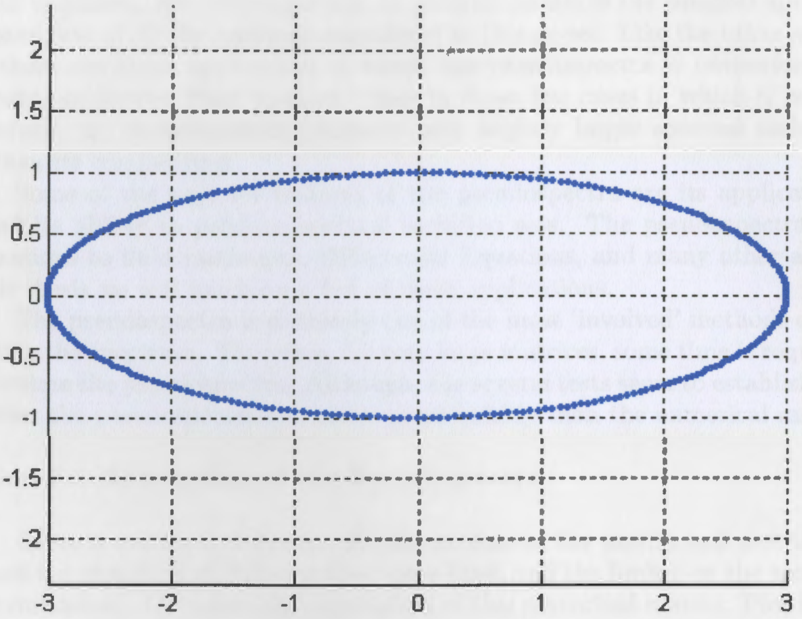
The symbol for this operator is $f_A = 2z + \frac{1}{z}$. The spectrum can be calculated by

$$f(S) = 2(\cos\theta + i\sin\theta) + \frac{1}{\cos\theta + i\sin\theta} \quad \text{for } 0 \leq \theta \leq 2\pi.$$

Calculating f(S) for various values of θ and plotting gives the following table:

θ	f(S)	θ	f(S)
0	3	π	-3
$\frac{\pi}{8}$	2.772+.3827i	$\frac{9\pi}{8}$	-2.7716-.3827i
$\frac{\pi}{4}$	2.1213+.7071i	$\frac{5\pi}{4}$	-2.1213-.7071i
$\frac{3\pi}{8}$	1.1481+.9239i	$\frac{11\pi}{8}$	-1.1481-.9239i
$\frac{\pi}{2}$	i	$\frac{3\pi}{2}$	-i
$\frac{5\pi}{8}$	-1.1481+.9239i	$\frac{13\pi}{8}$	1.1481-.9239i
$\frac{3\pi}{4}$	-2.1213+.7071i	$\frac{7\pi}{4}$	2.1213-.7071i
$\frac{7\pi}{8}$	-2.7716+.3827i	$\frac{15\pi}{8}$	2.7716-.3827i

This is plotted in figure 6.5.



The spectrum of the infinite dimensional operator of example 6.5
This spectrum was generated using the symbol of the operator.
Figure 6.5

Conclusions

A number of facts concerning Toeplitz matrices will be used in the rest of this thesis:

(1) The symbol of the Toeplitz matrix can be used to determine the spectrum of the related infinite dimension operator.

(2) The spectrum of a finite dimensional Toepitz operator as $n \to \infty$ is not equal to the spectrum of the related infinite dimensional operator. (See Reichel and Trefethen [69]).

(3) The symbol of the Toeplitz matrix serves as a spectral inclusion set for the related finite dimensional Toeplitz matrix.

7 Pseudospectra

As will be shown, the pseudospectra, in general, produces the smallest spectral inclusion sets of all the methods considered in this paper. Like the other methods, there are some application in which the pseudospectra is outperformed. However, unlike the other methods, even in those few cases in which it is outperformed, the pseudospectra produces only slightly larger spectral inclusion sets than its competition.

Some of the superior features of the pseudospectra are its applications beyond its ability to produce spectral inclusion sets. The pseudospectra has applications to fluid mechanics, Differential Equations, and many other areas. In this thesis we will touch on a few of these applications.

The pseudospectra is definitely one of the most 'involved' methods of estimating the spectrum. Therefore, for very large matrices, some time is required to calculate the pseudospectra. Although, the several tests seem to establish the fact that the pseudospectra calculates more quickly than the numerical range.

Section 7.1 Description of the Pseudospectra

Given a matrix A, calculate the eigenvalues of the matrix and plot them. Perturb the matrix A slightly (setting some limit, call the limit ε on the amount of perturbation). Calculate the eigenvalues of this perturbed matrix. Plot these eigenvalues. Perturb the original matrix once again using a different perturbation but staying within the value ε. Plot the eigenvalues of this matrix. Continue this process until all possible perturbations within ε are covered. The resulting plot of the eigenvalues of all of these perturbed matrices make up the **pseudospectra.**

The preceding description is made precise in the following definitions.

Pseudospectra - Definition 1

Let A $\in C^{nxn}$. Let $\varepsilon > 0$. Then the pseudospectra, $\sigma_\varepsilon(A)$, is defined as

$\sigma_\varepsilon(A) = \{z \in C : z \in \sigma(A + \Delta A) : ||\Delta A|| \leq \varepsilon\}$

Pseudospectra - Definition 2

Let A $\in C^{nxn}$. Let $\varepsilon > 0$. Then the pseudospectra, $\sigma_\varepsilon(A)$, is defined as

$\sigma_\varepsilon(A) = \{z \in C : ||(A - zI)^{-1}|| \geq \varepsilon^{-1}\}$

The sets produced according to Definitions 1 and 2 are the same. As the following shows.

Theorem 7.0 (Definitions 1 and 2 are equivalent)

Let A $\in C^{nxn}$. Let $\varepsilon > 0$. Then,

$\{z \in C : z \in \sigma(A + \Delta A) : ||\Delta A|| \leq \varepsilon\} = \{z \in C : ||(A - zI)^{-1}|| \geq \varepsilon^{-1}\}$

Proof (From Bottcher and Silbermann [2])

Let A $\in C^{nxn}$. Let $\varepsilon > 0$.

Show $\{z \in C : z \in \sigma(A + \Delta A) : ||\Delta A|| \leq \varepsilon\} = \{z \in C : ||(A - zI)^{-1}|| \geq \varepsilon^{-1}\}$

First Containment Direction $\subset$

Let $\lambda \in \{z \in C : z \in \sigma(A + \Delta A) : ||\Delta A|| \leq \varepsilon\}$

Choose ΔA such that $||\Delta A|| \leq \varepsilon$ and $A + \Delta A - \lambda I$ is not invertible

Case 1 $(A - \lambda I)$ is not invertible.

In this case, $\lambda \in \sigma(A)$.

Therefore, $\lambda \in \sigma(A + \Delta A) = \sigma(A)$ when $||\Delta A|| = 0$.

Therefore, $\lambda \in \{z \in C : z \in \sigma(A + \Delta A) : ||\Delta A|| \leq \varepsilon\}$

Case 2 $(A - \lambda I)$ is invertible.

Note:

$(A + \Delta A - \lambda I) = (A - \lambda I)(I + (A - \lambda I)^{-1}\Delta A)$
$= (I + (A - \lambda I)^{-1}\Delta A)(A - \lambda I)$

and

$(A - \lambda I)(I + (A - \lambda I)^{-1}\Delta A) =$

$A - \lambda I + (A - \lambda I)(A - \lambda I)^{-1}\Delta A =$

$A - \lambda I + \Delta A =$

$A + \Delta A - \lambda I$

So, we have

$(A - \lambda I)(I + (A - \lambda I)^{-1}\Delta A) = A + \Delta A - \lambda I$

Now since we have selected $A + \Delta A - \lambda I$ to be not invertible and $(A - \lambda I)$ is invertible then $I + (A - \lambda I)^{-1}\Delta A$ cannot be invertible. By the power series argument, $||(A - \lambda I)^{-1}\Delta A|| \geq 1$

and $||(A - \lambda I)^{-1}||||\Delta A|| \geq ||(A - \lambda I)^{-1}\Delta A|| \geq 1$

$||(A - \lambda I)^{-1}|| \geq 1/||\Delta A||$

But since we chose $||\Delta A|| \leq \varepsilon$,

$||(A - \lambda I)^{-1}|| \geq 1/\varepsilon$

Therefore,

$\lambda \in \{z \in C : ||(A - zI)^{-1}|| \geq \varepsilon^{-1}\}$.

Second Containment Direction $\supset$

Let $S_1 = \{z \in C : z \in \sigma(A + \Delta A) : ||\Delta A|| \leq \varepsilon\}$. Let $S_2 = \{z \in C : ||(A - zI)^{-1}|| \geq \varepsilon^{-1}\}$. We want to show $S_1 \supset S_2$. This will be done by contradiction. Assume that there exists $\lambda \in S_2 - S_1$. Since $\lambda \notin S_1$ then $\lambda \notin \sigma(A + \Delta A)$ for all $||\Delta A|| \leq \varepsilon$. That is $A + \Delta A - \lambda I$ is invertible for all $||\Delta A|| \leq \varepsilon$. This means that $A + \Delta A - \lambda I$ is invertible for all $||\Delta A|| = 0$. Therefore, $A - \lambda I$ is invertible.

$$\text{Set } \Delta A = \mu(A^* - \overline{\lambda} I)^{-1} \quad \text{where} \quad 0 < |\mu| \leq \frac{\varepsilon}{||(A^* - \overline{\lambda} I)^{-1}||}.$$

Taking the norm of both sides, $||\Delta A|| = ||\mu(A^* - \overline{\lambda} I)^{-1}||$. Therefore, $||\Delta A|| = |\mu| \; ||(A^* - \overline{\lambda} I)^{-1}||$. Therefore,

$$|\mu| = \frac{||\Delta A||}{||(A^* - \overline{\lambda} I)^{-1}||} \leq \frac{\varepsilon}{||(A^* - \overline{\lambda} I)^{-1}||}.$$

Therefore, $||\Delta A|| \leq \varepsilon$.

$$\begin{aligned} A - \lambda I + \Delta A &= A - \lambda I + \mu(A^* - \overline{\lambda} I)^{-1} \\ &= \mu(A - \lambda I)[\mu^{-1} I + (A - \lambda I)^{-1}(A^* - \overline{\lambda} I)^{-1}] \end{aligned}$$

is invertible. Therefore,

$\mu^{-1} I + (A - \lambda I)^{-1}(A^* - \overline{\lambda} I)^{-1}$ is invertible for all μ, $0 < |\mu| \leq \dfrac{\varepsilon}{||(A^* - \overline{\lambda} I)^{-1}||}$,

$$\mu^{-1} \notin \sigma((A-\lambda I)^{-1}(A^*-\overline{\lambda} I)^{-1}) \text{ for all } \mu,\ 0<|\mu| \leq \frac{\varepsilon}{\|(A^*-\overline{\lambda} I)^{-1}\|},$$

$$\text{and } \mu^{-1} \notin \sigma((A-\lambda I)^{-1}(A^*-\overline{\lambda} I)^{-1}) \text{ for all } \mu,\ \mu^{-1} \geq \frac{\|(A^*-\overline{\lambda} I)^{-1}\|}{\varepsilon}.$$

Therefore,

$$\begin{aligned} \rho((A-\lambda I)^{-1}(A^*-\overline{\lambda} I)^{-1}) &= \max\{|\lambda| : \lambda \in \sigma((A-\lambda I)^{-1}(A^*-\overline{\lambda} I)^{-1})\} \\ &< \frac{\|(A^*-\overline{\lambda} I)^{-1}\|}{\varepsilon}. \end{aligned}$$

Note that, $(A-\lambda I)^{-1}(A^*-\overline{\lambda} I)^{-1}$ is self-adjoint. Therefore,

$$\begin{aligned} \|(A-\lambda I)^{-1}\|^2 &= \|(A-\lambda I)^{-1}(A^*-\overline{\lambda} I)^{-1}\| \\ &= \rho((A-\lambda I)^{-1}(A^*-\overline{\lambda} I)^{-1}) \\ &< \frac{\|(A^*-\overline{\lambda} I)^{-1}\|}{\varepsilon} \\ &= \frac{\|(A-\lambda I)^{-1}\|}{\varepsilon}. \end{aligned}$$

Therefore, $\|(A-\lambda I)^{-1}\| < \frac{1}{\varepsilon}$. However, this is a contradiction since $\lambda \in S_2$.

Therefore, $S_1 \supset S_2$.

● ● ●●

Illustration

An an example is in order that illustrates how one might calculate the pseudospectra. Be aware that this is not the most efficient way of calculating the pseudospectra nor is this the way that the pseudospectra is calculated in most computer programs (most program employ 'singular value decomposition' to produce the pseudospectra. On the other hand, the method in the following example provides a good 'feel' as to what is happening in the pseudospectra.

Example 7.1 Given the matrix,

$$A = \begin{pmatrix} 7 & 3 \\ 2 & 6 \end{pmatrix}.$$

Find the pseudospectra for this matrix with $\varepsilon = .1$.

Solution

According to the Definition, the pseudospectra is produced by first choosing a matrix, (ΔA), whose norm is less than ε. (In this case, $\varepsilon = .1$). Add this matrix ΔA to A and find the eigenvalues of $(A + \Delta A)$. Repeat this process by choosing a different (ΔA). This process is repeated until every ΔA, whose norm is less than ε, has been chosen. (Of course, this would mean choosing an infinite number of matrices. So, for practical purposes, some finite increment is used when choosing the ΔA's). A sampling of the process is presented below:

$$A + \Delta A = \begin{pmatrix} 7 & 3 \\ 2 & 6 \end{pmatrix} + \begin{pmatrix} -.07i & -.07 \\ -.07 & -.07i \end{pmatrix} = \begin{pmatrix} 7 - .0707i & 2.9293 \\ 1.9293 & 6 - .0707i \end{pmatrix}$$

$$\lambda_1 = 8.9293 - .0707i \qquad \lambda_2 = 4.0707 - .0707i$$

$$A + \Delta A = \begin{pmatrix} 7 & 3 \\ 2 & 6 \end{pmatrix} + \begin{pmatrix} .08i & .05 \\ .06 & .07i \end{pmatrix} = \begin{pmatrix} 7 + .08i & 3.05 \\ 2.06 & 6 + .07i \end{pmatrix}$$

$$\lambda_1 = 9.056 + .076i \qquad \lambda_2 = 3.944 + .74i$$

$$A + \Delta A = \begin{pmatrix} 7 & 3 \\ 2 & 6 \end{pmatrix} + \begin{pmatrix} -.08i & -.05 \\ .06 & .07i \end{pmatrix} = \begin{pmatrix} 7 - .08i & 2.95 \\ 1.94 & 6 - .07i \end{pmatrix}$$

$$\lambda_1 = 8.944 - .076i \qquad \lambda_2 = 4.056 - .074i$$

$$A + \Delta A = \begin{pmatrix} 7 & 3 \\ 2 & 6 \end{pmatrix} + \begin{pmatrix} .08i & .06 \\ .06 & .08i \end{pmatrix} = \begin{pmatrix} 7 + .08i & 3.06 \\ 2.06 & 6 + .08i \end{pmatrix}$$

$$\lambda_1 = 9.06 + .08i \qquad \lambda_2 = 3.94 + .08i$$

$$A + \Delta A = \begin{pmatrix} 7 & 3 \\ 2 & 6 \end{pmatrix} + \begin{pmatrix} .07i & .07 \\ .07 & .07i \end{pmatrix} = \begin{pmatrix} 7 + .0707i & 3.0707 \\ 2.0707 & 6 + .0707i \end{pmatrix}$$

$$\lambda_1 = 9.0707 + .0707i \qquad \lambda_2 = 3.9293 + .0707i$$

$$A + \Delta A = \begin{pmatrix} 7 & 3 \\ 2 & 6 \end{pmatrix} + \begin{pmatrix} .1 & 0 \\ 0 & .1 \end{pmatrix} = \begin{pmatrix} 7.1 & 3 \\ 2 & 6.1 \end{pmatrix}$$

$$\lambda_1 = 9.1 \qquad \lambda_2 = 4.1$$

$$A + \Delta A = \begin{pmatrix} 7 & 3 \\ 2 & 6 \end{pmatrix} + \begin{pmatrix} 0 & .1 \\ .1 & 0 \end{pmatrix} = \begin{pmatrix} 7 & 3.1 \\ 2.1 & 6 \end{pmatrix}$$

$$\lambda_1 = 9.1 \qquad \lambda_2 = 3.9$$

$$A + \Delta A = \begin{pmatrix} 7 & 3 \\ 2 & 6 \end{pmatrix} + \begin{pmatrix} 0 & .1i \\ .1i & 0 \end{pmatrix} = \begin{pmatrix} 7 & 3 + .1i \\ 2 + .1i & 6 \end{pmatrix}$$

$$\lambda_1 = 9 + .1i \qquad \lambda_2 = 4 - .1i$$

$$A + \Delta A = \begin{pmatrix} 7 & 3 \\ 2 & 6 \end{pmatrix} + \begin{pmatrix} .1i & 0 \\ 0 & .1i \end{pmatrix} = \begin{pmatrix} 7 + .1i & 3 \\ 2 & 6 + .1i \end{pmatrix}$$
$$\lambda_1 = 9 + .1i \qquad \lambda_2 = 4 + .1i$$

$$A + \Delta A = \begin{pmatrix} 7 & 3 \\ 2 & 6 \end{pmatrix} + \begin{pmatrix} -.1 & 0 \\ 0 & -.1 \end{pmatrix} = \begin{pmatrix} 6.9 & 3 \\ 2 & 5.9 \end{pmatrix}$$
$$\lambda_1 = 8.9 \qquad \lambda_2 = 3.9$$

$$A + \Delta A = \begin{pmatrix} 7 & 3 \\ 2 & 6 \end{pmatrix} + \begin{pmatrix} .09i & .05 \\ .05 & .09i \end{pmatrix} = \begin{pmatrix} 7 + .09i & 3.05 \\ 2.05 & 6 + .09i \end{pmatrix}$$
$$\lambda_1 = 9.05 + .09i \qquad \lambda_2 = 3.95 + .09i$$

$$A + \Delta A = \begin{pmatrix} 7 & 3 \\ 2 & 6 \end{pmatrix} + \begin{pmatrix} .09i & -.05 \\ -.05 & .09i \end{pmatrix} = \begin{pmatrix} 7 + .09i & 2.95 \\ 1.95 & 6 + .09i \end{pmatrix}$$
$$\lambda_1 = 8.95 + .09i \qquad \lambda_2 = 4.05 + .09i$$

$$A + \Delta A = \begin{pmatrix} 7 & 3 \\ 2 & 6 \end{pmatrix} + \begin{pmatrix} -.09i & -.05 \\ -.05 & -.09i \end{pmatrix} = \begin{pmatrix} 7 - .09i & 2.95 \\ 1.95 & 6 - .09i \end{pmatrix}$$
$$\lambda_1 = 8.95 - .09i \qquad \lambda_2 = 4.05 - .09i$$

$$A + \Delta A = \begin{pmatrix} 7 & 3 \\ 2 & 6 \end{pmatrix} + \begin{pmatrix} -.09i & .05 \\ .05 & -.09i \end{pmatrix} = \begin{pmatrix} 7 - .09i & 3.05 \\ 2.05 & 6 - .09i \end{pmatrix}$$
$$\lambda_1 = 9.05 - .09i \qquad \lambda_2 = 3.95 - .09i$$

$$A + \Delta A = \begin{pmatrix} 7 & 3 \\ 2 & 6 \end{pmatrix} + \begin{pmatrix} .033i & .051 \\ .6i & .053 \end{pmatrix} = \begin{pmatrix} 7 + .033i & 3.051 \\ 2 + .06i & .053 \end{pmatrix}$$
$$\lambda_1 = 9.042 + .056i \qquad \lambda_2 = 4.011 + .023i$$

$$A + \Delta A = \begin{pmatrix} 7 & 3 \\ 2 & 6 \end{pmatrix} + \begin{pmatrix} -.033i & .051 \\ -.6i & .053 \end{pmatrix} = \begin{pmatrix} 7 - .033i & 3.051 \\ 2 - .06i & .053 \end{pmatrix}$$
$$\lambda_1 = 9.042 - .056i \qquad \lambda_2 = 4.011 + .023i$$

...This process is continues until a sufficient number of points have been calculated...

The eigenvalues of all of these 'perturbed' matrices are then plotted as shown in figures 7.1A and 7.1B. This is the pseudospectra.

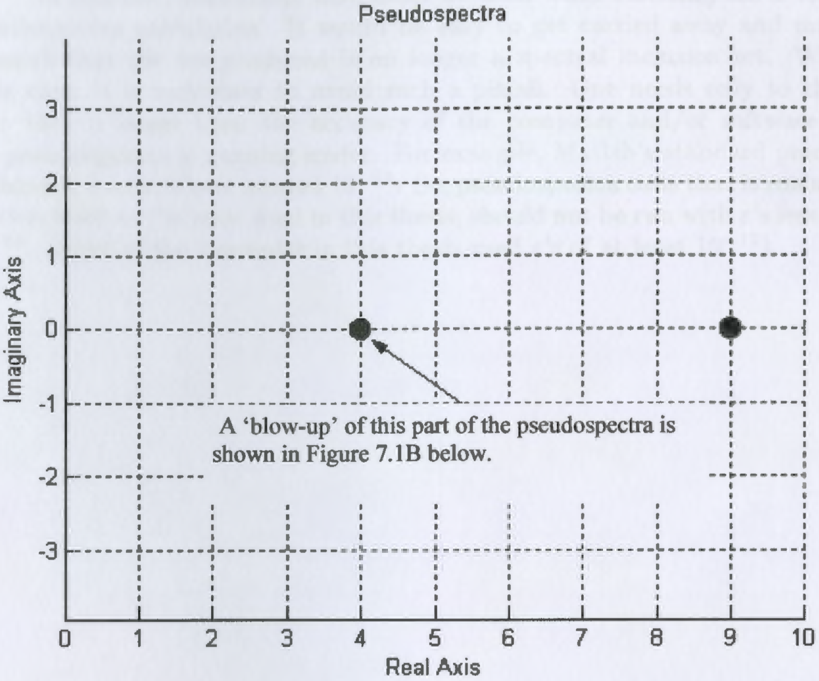


Figure 7.1A

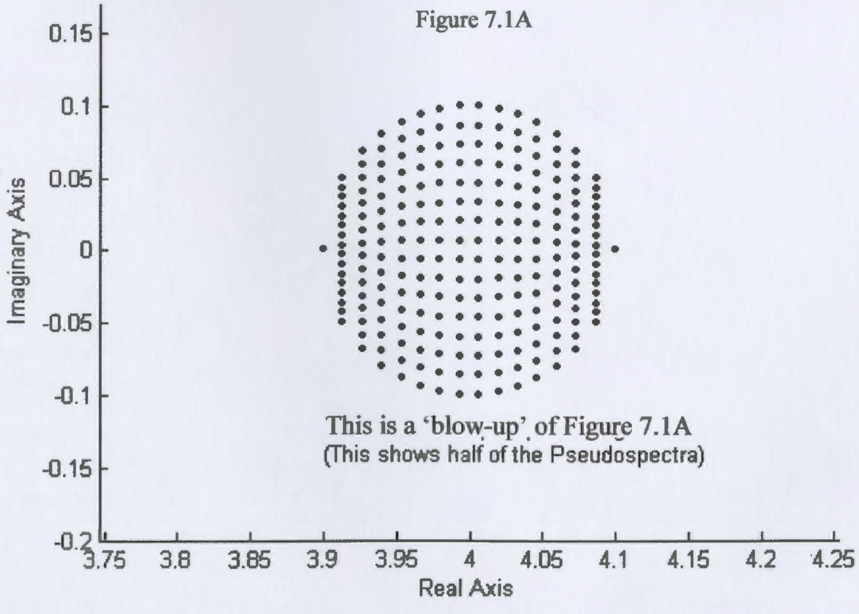


Figure 7.1B

Again, the method used to calculate the pseudospectra in the last example is not very efficient. That method was used here because it helps to understand the pseudospectra. The most efficient way to calculate the pseudospectra is by *singular value decomposition.*

Section 7.1.1 Appropriate selection of ε for the pseudospectra

In practice, reasonable care must be used when selecting the ε for the pseudospectra calculation. It would be easy to get carried away and make ε so small that the set produced is no longer a spectral inclusion set. With a little care, it is very easy to avoid such a pitfall. One needs only to choose an ε that is larger than the accuracy of the computer and/or software that the pseudospectra is running under. For example, Matlab's standard precision arithmetic is somewhere around 10^{-16}. So, pseudospectra code that is run under Matlab, such as the code used in this thesis, should not be run with ε's less than 10^{-15}. (Most of the examples in this thesis used ε's of at least 10^{-12}).

Section 7.2 spectral inclusion sets

The pseudospectra is superior to any other method in its ability to produce sharp spectral inclusion sets. Just a few examples will illustrate the point.

Note that in all of these examples, the pseudospectra was calculated using the program 'eigtool' written by Tom Wright.

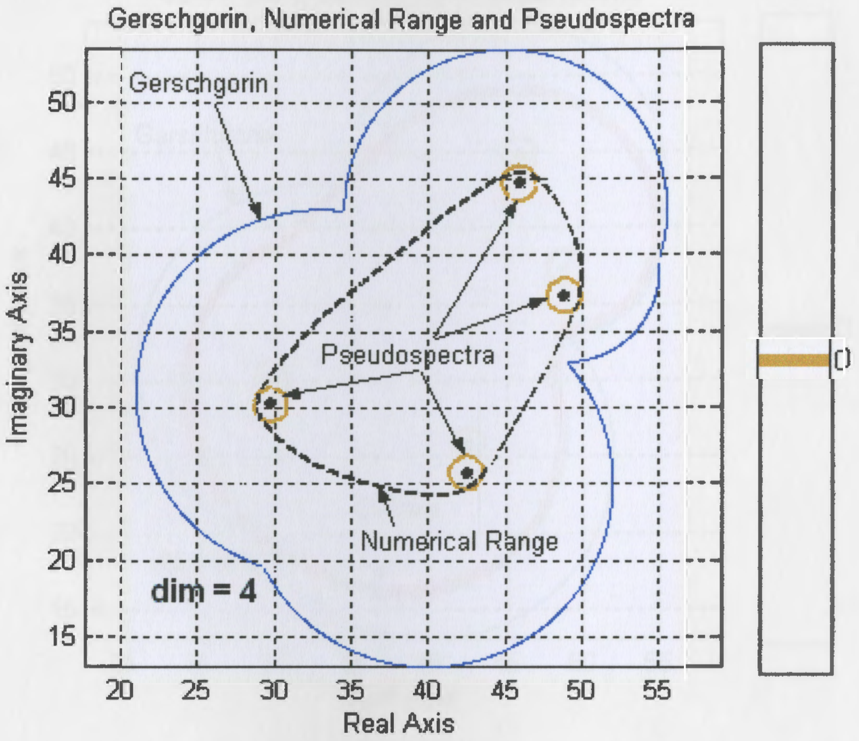
Example 7.2 Let

$$\mathbf{A} = \begin{pmatrix} 40+25i & 5 & 4 & -3 \\ 6 & 33+31i & -4 & 2 \\ 1 & -2 & 49+39i & 3 \\ 3 & 3.5 & -4 & 45+43i \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{29.74 + 30.21i, 42.50 + 25.71i, 45.91 + 44.76i, 48.85 + 37.31i\}.$$

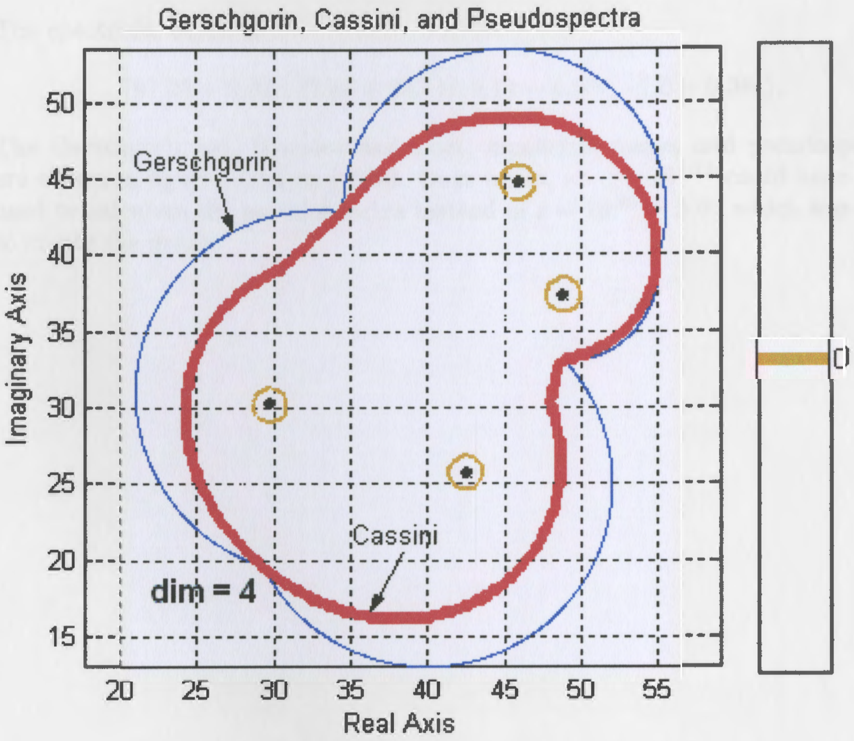
The Gerschgorin set, Brauer-Cassini set, numerical range, and pseudospectra are shown in figures 7.2A and 7.2B. Notice how small the pseudospectra set is compared to all the others. In this case, the pseudospectra was calculated using an $\varepsilon = 10^0 = 1$. A much smaller ε could have been used but, in that case, the set would have been too small to see! For example, an $\varepsilon = 10^{-12}$ could have been used.



Notice that the Numerical Range's Inclusion Set is large and the Gerschgorin Set is even larger while the Pseudospectra Set consists only of four tiny circles.

Figure 7.2A

In these figures, the eigenvalues are reprented by the black dots



The Cassini Set is slightly smaller than Gerschgorin and larger than the Numerical Range (see figure at top of page). → 7.24

Figure 7.2B

In these figures, the eigenvalues are reprented by the black dots

Revisiting Example 1.4,

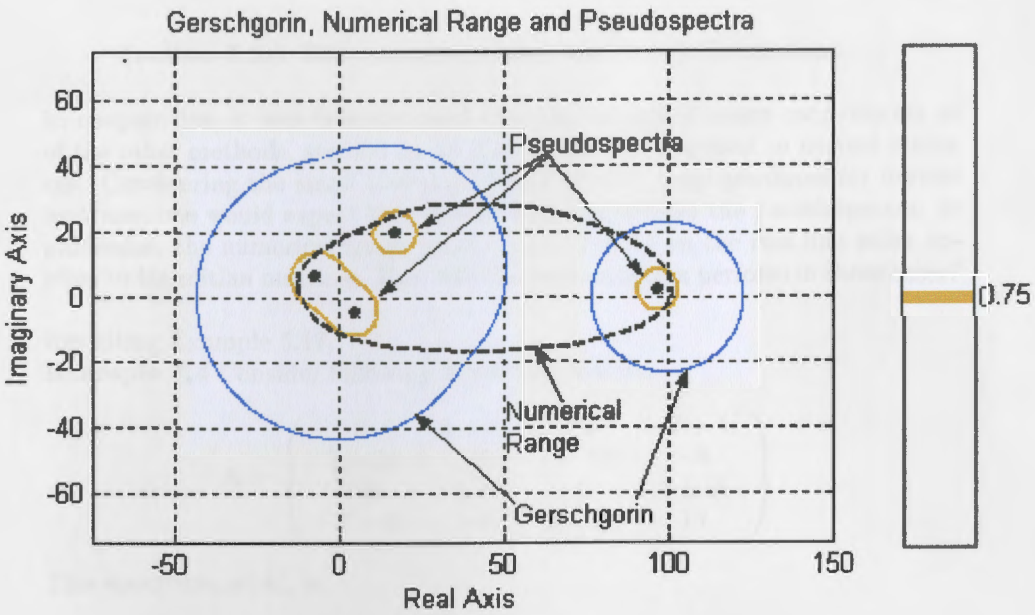
Example 7.3 Recall

$$\mathbf{A} = \begin{pmatrix} 100 & 2 & -6 & 15i \\ 7-6i & 15i & -7+3i & 8+5i \\ 19 & 8-4i & 13+9i & 10 \\ 16+9i & 15-2i & -9+3i & 0 \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

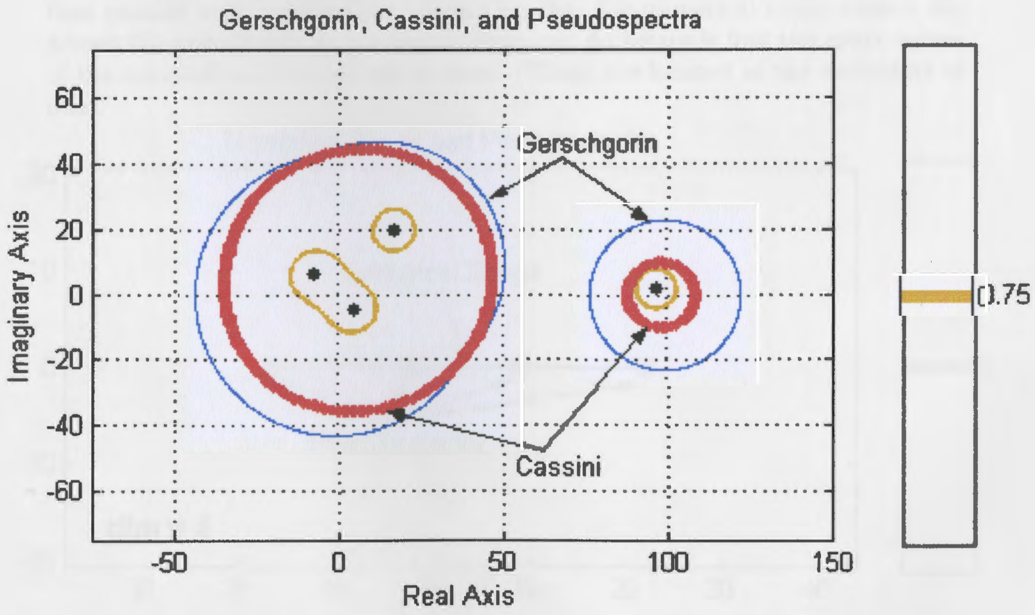
$$\{97.33 + 2.07i, 17.51 + 20.11i, 5.14 - 4.56i, -7.0 + 6.38i\}.$$

The Gerschgorin set, Brauer-Cassini set, numerical range, and pseudospectra are shown in figures 7.3A and 7.3B. Once again, an $\varepsilon = 10^{-12}$ could have been used to calculate the pseudospectra instead of $\varepsilon = 10^{.75} = 5.62$ which was used to create the graph.



Notice that the Numerical Range and the Gerschgorin sets are large while the Pseudospectra Set consists only of the intersection of three small regions.

Figure 7.3A



The Cassini Set is smaller than Gerschgorin but much larger than the Pseudospectra

Figure 7.3B

For illustrative purposes, the Pseudospectra was calculated with an e = 5.62. A much smaller e could have been used which would have further reduced the size of the Pseudospectra Set.

Section 7.2.1 The pseudospectra and normal matrices

In chapter five, it was demonstrated that the numerical range outperforms all of the other methods, studied up to that point, when applied to normal matrices. Considering the small sets that the numerical range produces for normal matrices, one would expect that it would far outperform the pseudospectra. In particular, the numerical range produces an interval on the real line when applied to Hermitian matrices. How will the pseudospectra perform in those cases?

Revisiting Example 5.17,
Example 7.4 Consider following Hermitian matrix.

$$\mathbf{A} = \begin{pmatrix} 3 & 3-2i & -5 & -7+4i \\ 3+2i & -4 & 2-11i & -8 \\ -5 & 2+11i & 1 & 5+5i \\ -7-4i & -8 & 5-5i & 13 \end{pmatrix}.$$

This spectrum, $\sigma(A)$, is

$$\{22.52, -16.82, 9.23, -1.93\}.$$

The numerical range and pseudospectra are shown in figure 7.4. Even with this Hermitian matrix, the pseudospectra's performance is at least as good as the numerical range's. The pseudospectra is able to separate and isolate the regions that contain each eigenvalue - something that the numerical range cannot do. About the only thing the numerical range can do better is find the *exact* values of the smallest and largest eigenvalues. (These are located at the endpoints of line).

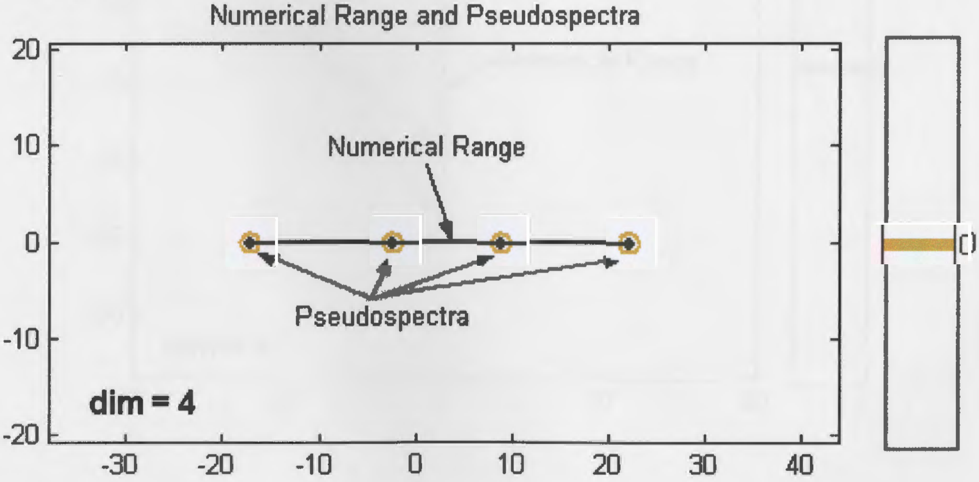


Figure 7.4

The Pseudospectra in these graphs was produced using the software program 'eigtool' developed by Tom Wright.

Revisiting Example 5.2,
Example 7.5 Consider following skew-Hermitian matrix.

$$\mathbf{A} = \begin{pmatrix} 3i & 3-2i & -5 & 7+4i \\ -3-2i & -7i & 2-11i & 8 \\ 5 & -2-11i & 11i & -5+5i \\ -7+4i & -8 & 5+5i & -5i \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{20.68i, 7.87i, -18.07i, -8.48i\}.$$

The numerical range and pseudospectra are shown in figure 7.5. Again, the pseudospectra's performance is outstanding.

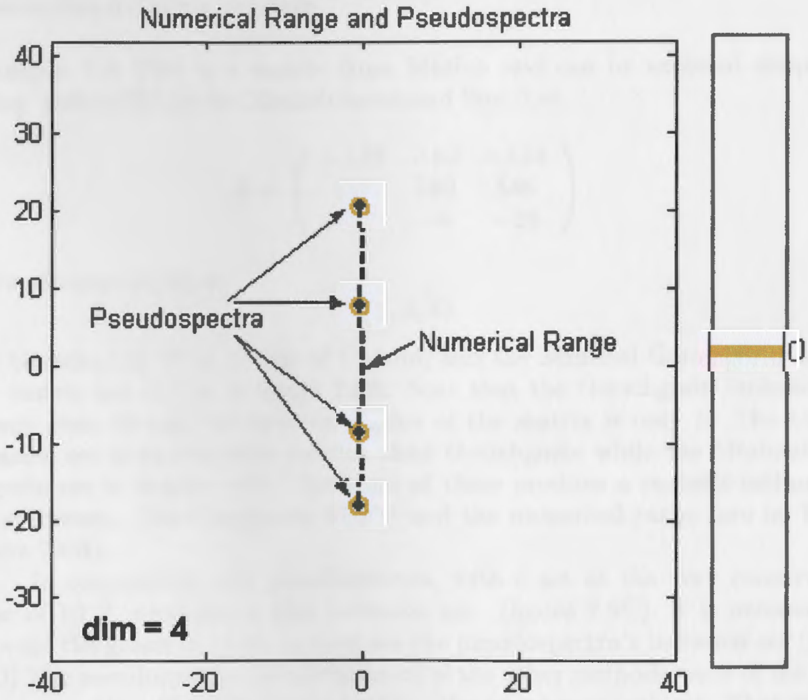


Figure 7.5

The Pseudospectra in these graphs were produced using the software program 'eigtool' developed by Tom Wright.

Even though the pseudospectra was excellent in the preceding examples, its peformance is even more outstanding in difficult situations as the next section will show.

Section 7.2.2 The pseudospectra and Ill-Conditioned matrices

Of course, one of the main reasons for using a spectral estimation method is to handle ill-conditioned matrices. Ill-Conditioned matrices are very sensitive to perturbations. Therefore, it is very difficult to accurately calculate the eigenvalues of such matrices because any roundoff error will cause inaccurate results.

All of the spectral-estimation methods that we have considered so far will produce sets that include the spectrum of any matrix even an ill-conditioned one. However, in some cases, the inclusion sets that are produced by these methods are so large that they are practically unusable. **On the other hand, the pseudospectra will produce relatively small inclusion sets even with ill-conditioned matrices.**

Consider the following example,

Example 7.6 This is a matrix from Matlab and can be accessed simply by typing 'gallery(3)' at the Matlab command line. Let

$$\mathbf{A} = \begin{pmatrix} -149 & -50 & -154 \\ 537 & 180 & 546 \\ -27 & -9 & -25 \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{1, 2, 3\}.$$

The Gerschgorin disks, Ovals of Cassini, and the Minimal Gerschgorin set for this matrix are shown in figure 7.6A. Note that the Gerschgorin inclusion set is huge even though the spectral radius of the matrix is only 3! The Cassini inclusion set is significantly smaller than Gerschgorin while the Minimal Gerschgorin set is smaller still. Yet none of these produce a realistic estimate of the spectrum. The Composite BBFP and the numerical range fare no better (figure 7.6B).

In comparison, the pseudospectra, with ε set at the very conservative value of 10^{-3}, produces a tiny inclusion set. (figure 7.6C). It is necessary to 'blowup' the graph in order to even see the pseudospectra's inclusion set (figure 7.6D) The pseudospectra outperforms all of the other methods, each of the other sets is more than 10,000 times larger than the pseudospectra's set. That is, each of the other sets covers more than 10,000 times the area on the Complex plane than the pseudospectra's set.

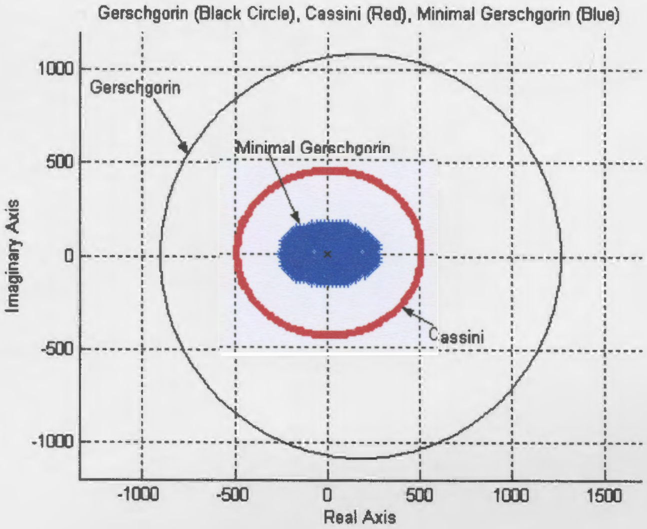


Figure 7.6A

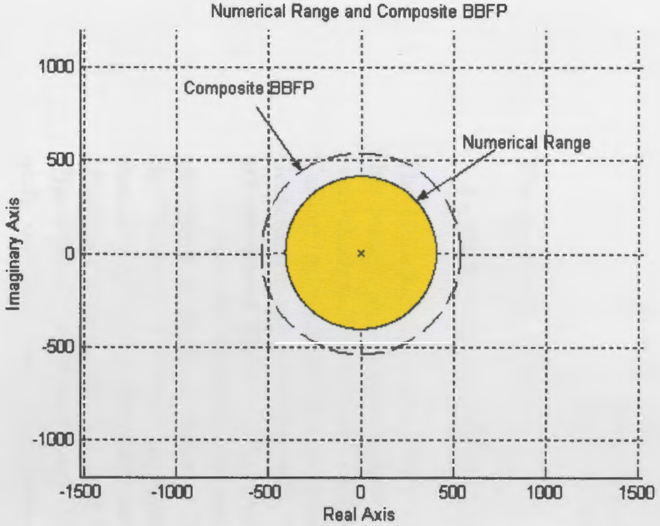
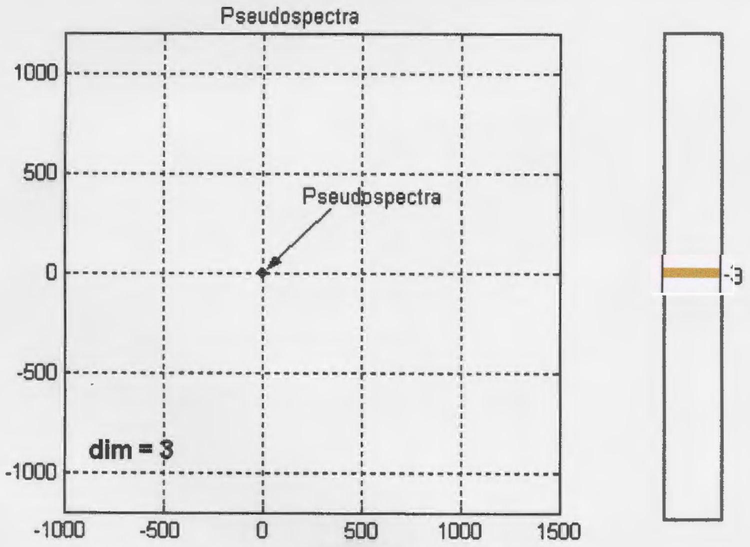
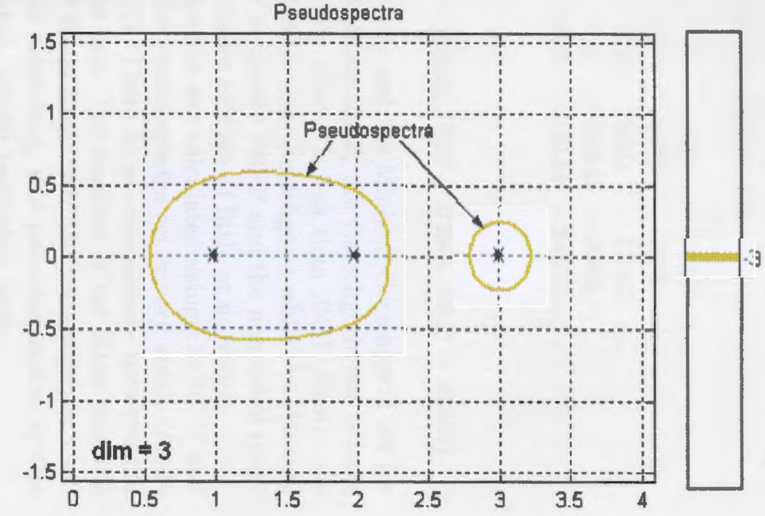


Figure 7.6B



Notice how small the Psuedospectra Set is compared to the sets in Figures 7.6A and 7.6B above.

Figure 7.6C



This Figure is a 'blow-up' of figure 7.6C

Figure 7.6D

Example 7.7 This is a matrix from Matlab and can be accessed simply by typing 'gallery(5)' at the Matlab command line. Let

$$\mathbf{A} = \begin{pmatrix} -9 & 11 & -21 & 63 & -252 \\ 70 & -69 & 141 & -421 & 1684 \\ -575 & 575 & -1149 & 3451 & -13801 \\ 3891 & -3891 & 7782 & -23345 & 93365 \\ 1024 & -1024 & 2048 & -6144 & 24572 \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{-.0408, -.0119 + .0386i, -.0119 - .0386i, .0323 + .0230i, .0323 - .0230i\}.$$

The Gerschgorin disks, Ovals of Cassini, and the Minimal Gerschgorin set for this matrix are shown in figure 7.7A. Once again, the Gerschgorin set is huge even though the spectral radius of this matrix is less than .041! Again, the Cassini inclusion set is significantly smaller than Gerschgorin while the Minimal Gerschgorin set is smaller still. The Composite BBFP and the numerical range are enormous (figure 7.7B). Such inclusion sets are of little or no value.

For this example, the pseudospectra was calculated using $\varepsilon = 10^{-12}$ and the result is shown in figure 7.7C. The pseudospectra set is very small. Again the 'blow-up' is shown in figure 7.7D. There is no comparison between the pseudospectra and the other inclusion sets. The smallest of the other sets, the Minimal Gerschgorin, is over 4×10^{10} times larger than the pseudospectra's set! **Therefore, for these very difficult matrices, the pseudospectra is the only method that produces reliable, sharp inclusion sets.**

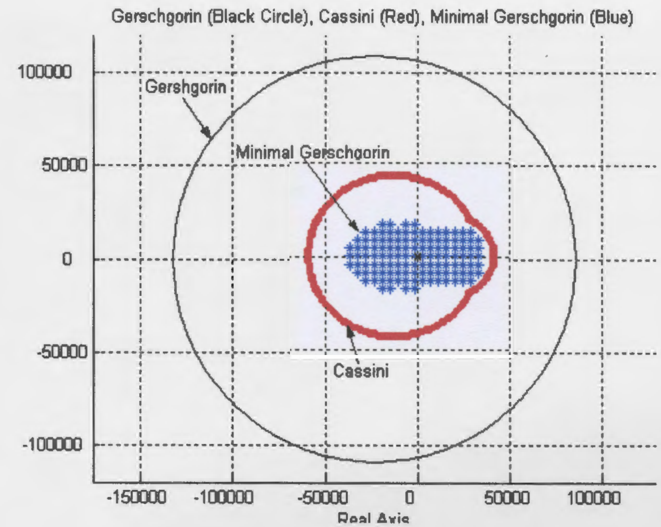


Figure 7.7A

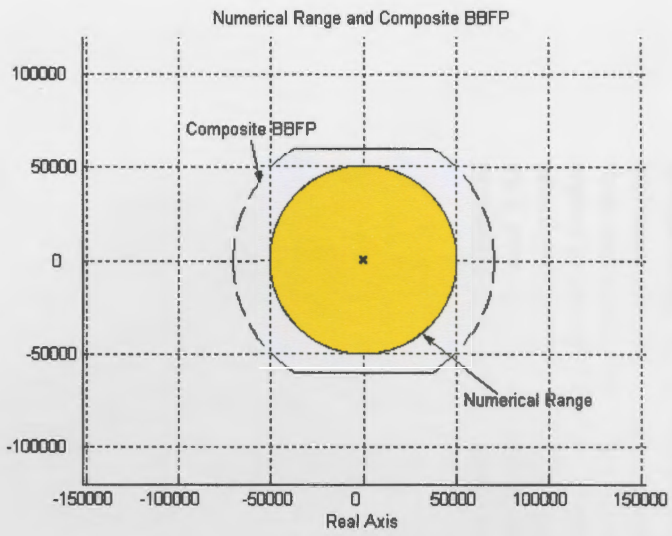
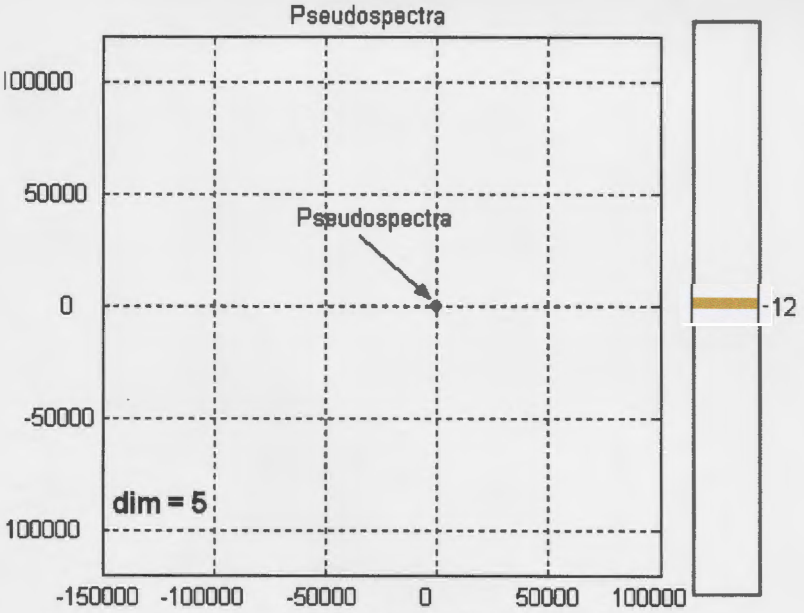
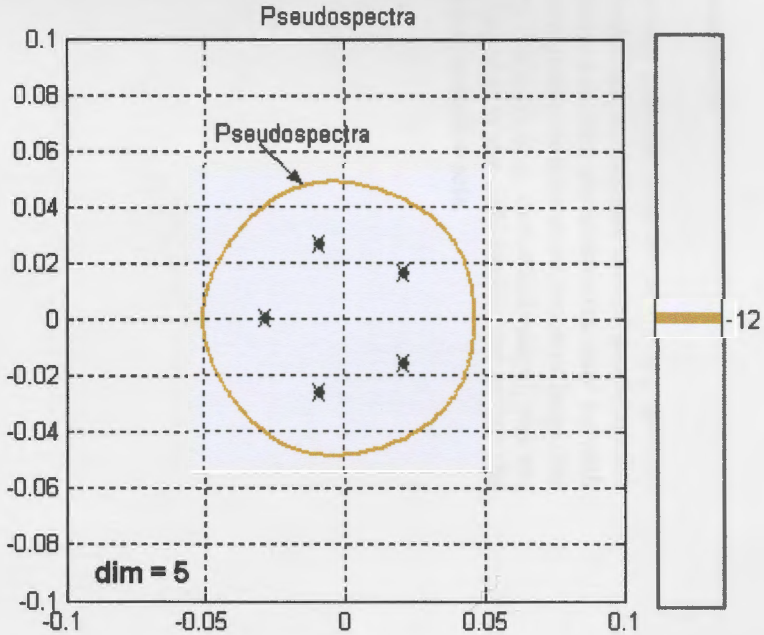


Figure 7.7B



Notice how small the Psuedospectra Set is compared to the sets in Figures 7.7A and 7.7B above.

Figure 7.7C



This Figure is a 'blow-up' of figure 7.7C

Figure 7.7D

Conclusions for spectral inclusions

In conclusion, the pseudospectra is a very powerful tool for producing spectral inclusion sets. This method produces sets that are, almost always, much smaller than any other sets. The only drawback with the pseudospectra may be with regard to calculation time. If the matrix under consideration becomes large, the pseudospectra, like the other 'complex' methods (e.g. numerical range), can use up a great deal of calculation time. Yet, this is the *only* reason not to use the pseudospectra a way to produce spectral inclusion sets.

8 Comparison of Various Inclusion Sets

This is the first of three 'Results' chapters that bring together analysis presented earlier in the thesis. Throughout this thesis, a number of spectral estimation methods have been examined as to their speed, sharpness and unique characteristics. In this chapter, an attempt will be made to order the different methods according to speed and sharpness.

Section 8.1 spectral inclusion sets

As demonstrated in chapter five, when the matrix under consideration is normal, the numerical range will produce a very small, often linear, spectral inclusion set. Only the pseudospectra can produce spectral inclusion sets of comparable size for normal matrices. When applied to normal matrices, the pseudospectra will produce slightly smaller inclusion sets than the numerical range. On the other hand, the numerical range will determine the exact values of the extreme eigenvalues of a normal matrix. So, each method has its advantages over the other when considering normal matrices. No other method comes even close to the pseudospectra and numerical range when applied to normal matrices.

On the other hand, when considering finite-dimensional, non-normal matrices the pseudospectra is far and away the best method available. In particular, pseudospectra is to be preferred over the numerical range. This observation is significant because the numerical range continues to be the subject of numerous journal articles. Improvements in numerical range methods and algorithms continue to be put forth. Yet, the sharpness and versatility of the pseudospectra make it clearly superior to the numerical range for non-normal matrices. Furthermore, the pseudospectra can be calculated as fast if not faster than the the numerical range. Therefore, if the pseudospectra is available, the numerical range need not be considered when analyzing non-normal matrices.

If the pseudospectra is better than all other methods, where does that leave the 'simpler' methods? The pseudospectra does produce the best possible inclusion sets for non-normal matrices as long as time and computer memory are not an issue. On the other hand, if the matrix is so large that pseudospectra calculation time becomes prohibitively long, then the simpler methods are a reasonable alternative.

For very large matrices, the Composite BBFP (developed in chapter One) or the Gerschgorin methods produce reasonably small spectral inclusion set while using relatively little computer time. In many instances, these 'simpler' methods will outperform even the numerical range. Recall from chapter five that the numerical range always produces a convex set. This can be an advantage if the eigenvalues are bunched close together. However, when the eigenvalues are separated, the numerical range is forced to 'stretch out' in order to enclose the spectrum. The resulting set becomes much larger than the Gerschgorin or Composite BBFP sets.

Illustrations

The preceding discussion can be illustrated by the use of three examples. The Gerschgorin disks are fast simple and easily constructed. Theses disks produce a reasonable inclusion set for any application. The great strength of the Gerschgorin disks manifests itself when the eigenvalues are separated into groups on the complex plane. When the eigenvalues are separated into groups, the Gerschgorin disks will usually take the form of two or more separate groups of disks in order to cover the eigenvalues. In such applications, all of the pre-Gerschgorin methods and the numerical range will cover the eigenvalues by forming one large convex inclusion set. A set that is almost always much larger than the Gerschgorin disks.

The next three examples will illustrate the strengths and weaknesses of the of the various methods.

Example 8.1 Let

$$\mathbf{A} = \begin{pmatrix} -31-18i & -3 & 8i & -4 & 7 & 6 & -12 & 10 \\ 2 & -20+21i & 3 & 5 & -4 & 13i & -8 & 6 \\ -5 & 6 & 25+33i & 3 & 6 & -9 & 10 & -11 \\ 1 & -2 & 2 & -40-23i & 7 & 5 & -3 & 8 \\ 6i & 3 & 9i & -5 & 30+26i & 7 & 6 & 5 \\ 5 & 3i & 2i & 4 & 8 & -25+19i & 3 & 4 \\ 7 & 4 & 3 & 4 & 12 & 8 & 30-24i & 6 \\ 8 & 2 & 7 & 0 & -6 & 7 & 3 & 25-25i \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{31.2 - 17.8i, -39.4 - 23.2i, 25.1 - 26i, 30.7 - 22.5i,$$

$$-22.6 + 25.7i, -23.97 + 13.5i, 23.5 + 27.9i, 31.9 + 31.4i\}.$$

The Gerschgorin disks and numerical range for this matrix are shown in figure 8.1A. Notice that the Gerschgorin disks form one connected set - none of the disks are separated from the others. When the Gercshgorin disks form a connected set, the numerical range will usually perform well. This case is no exception, the numerical range produces an inclusion set that is somewhat smaller than the set produced by the Gerschgorin disks. Notice that the convex nature of the numerical range tends to produce a boundary that is close to the actual eigenvalues. The numerical range also outperforms the Ovals of Cassini (figure 8.1B). It can be seen that the Composite BBFP performs slightly worse than the Gerschgorin disks (figure 8.1C). On the other hand, the pseudospectra outperforms all of the other methods (figure 8.1D). Notice that its inclusion set is tiny compared to the others.

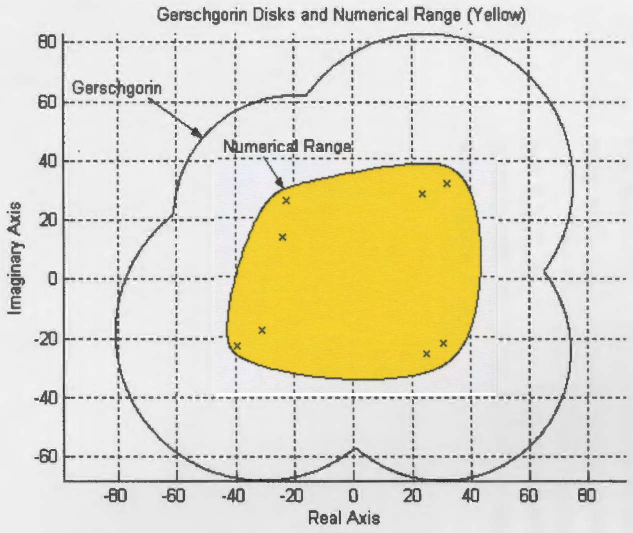


Figure 8.1A

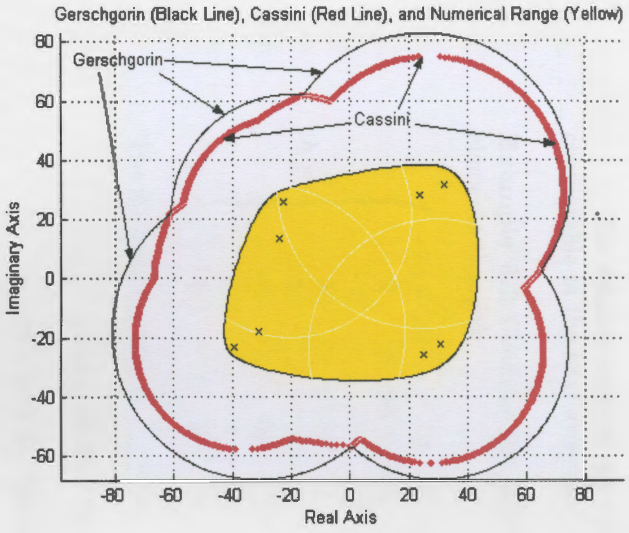


Figure 8.1B

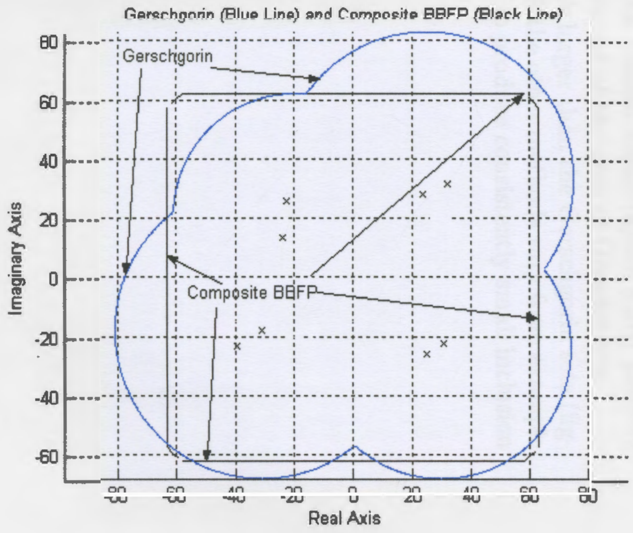


Figure 8.1 C

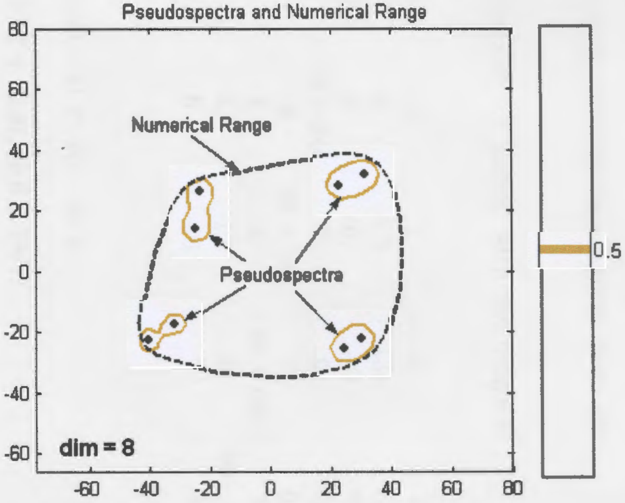


Figure 8.1D

All of the graphs on this page are for Example 8.1

The relative sizes of the numerical range and the Gerschgorin disks are comparable in the next example.

Example 8.2 Consider the matrix of example 8.1 except with the diagonal elements changed:

$$\mathbf{A} = \begin{pmatrix} -74-55i & -3 & 8i & -4 & 7 & 6 & -12 & 10 \\ 2 & -65+68i & 3 & 5 & -4 & 13i & -8 & 6 \\ -5 & 6 & -79+77i & 3 & 6 & -9 & 10 & -11 \\ 1 & -2 & 2 & -79-84i & 7 & 5 & -3 & 8 \\ 6i & 3 & 9i & -5 & 88+74i & 7 & 6 & 5 \\ 5 & 3i & 2i & 4 & 8 & -86+68i & 3 & 4 \\ 7 & 4 & 3 & 4 & 12 & 8 & 90-83i & 6 \\ 8 & 2 & 7 & 0 & -6 & 7 & 3 & 95-63i \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{-74-54.7i, -78.8-84i, 89-81.7i, 96-64.4i,$$
$$-84.5+67.6i, -67+68.2i, 77.7+72.8i, 89.6+78i\}.$$

The Gerschgorin disks and numerical range for this matrix are shown in figure 8.2A. Notice that the Gerschgorin set is made up of a union of disjoint sets. Also note that even though the disks are separated, they are not too far apart compared to their size. This is an indication that the numerical range's performance will be moderate. As can be seen in the figure, the area covered by the numerical range is only slightly less than the area covered by the Gerschgorin disks. The numerical range 'stretched out' to cover the sets. On the other hand, the Gerschgorin disks separated to cover the eigenvalues thereby keeping the total area covered to a minimum. The Cassini Ovals (figure 8.2B) perform very well covering only about seventy percent of the area of Gerschgorin.

The Composite BBFP is noticeable larger than the Gerschgorin sets (figure 8.2C). As the performance levels of all the other methods vary from example to example, the pseudospectra continues to produce consistently small inclusion sets (figure 8.2D).

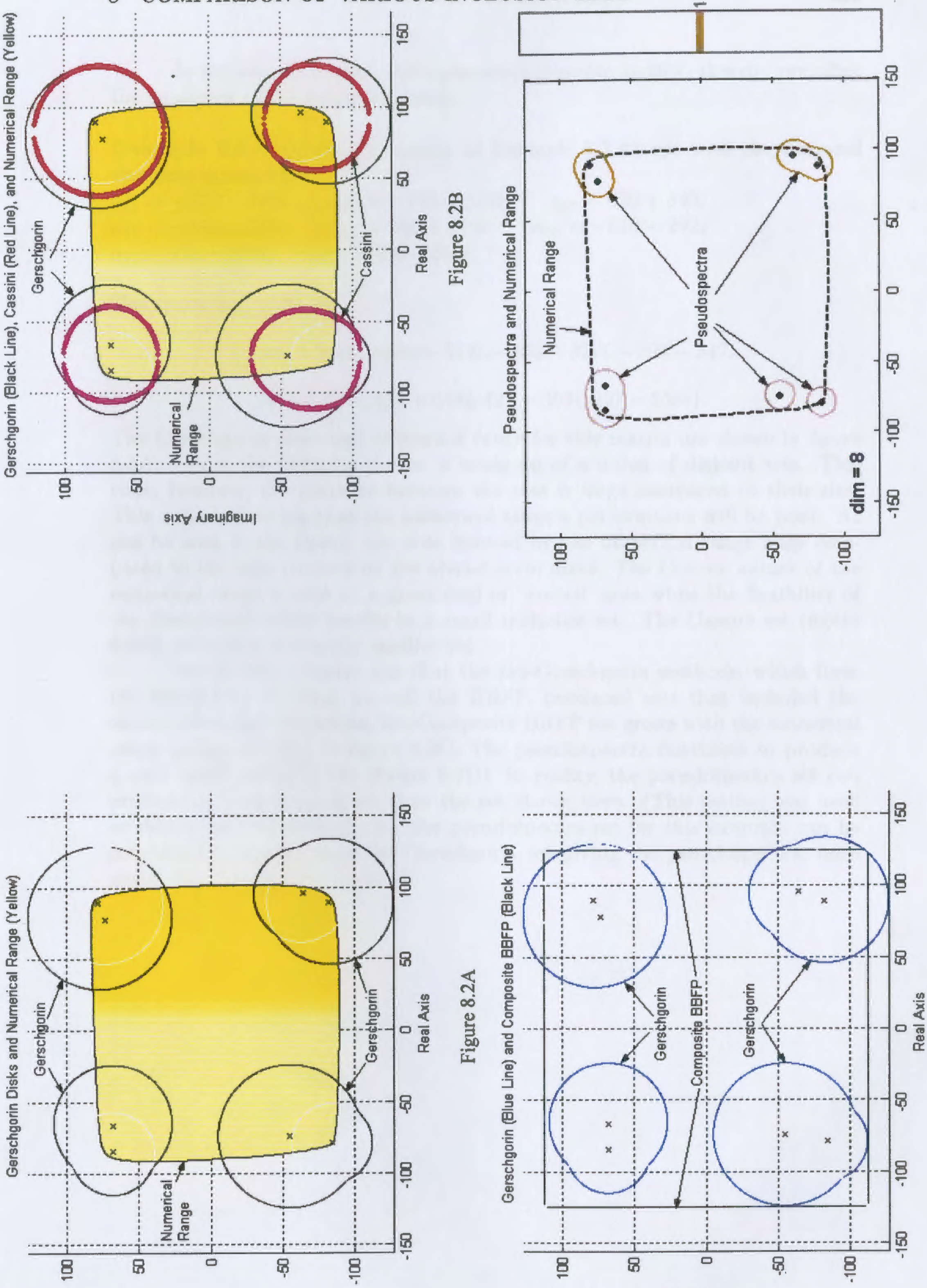


Figure 8.2A

Figure 8.2B

Figure 8.2C

Figure 8.2D

All of the graphs on this page are for Example 8.2

In the next examples, the eigenvalues separate further, thereby revealing the weakness of the numerical range.

Example 8.3 Consider the matrix of Example 8.2 except with the diagonal elements changed to:
$a_{11} = -303 - 327i$ $a_{22} = -499 + 310i$ $a_{33} = 436 + 545i$
$a_{44} = -298 - 347i$ $a_{55} = 455 + 519i$ $a_{66} = -510 + 292i$
$a_{77} = 421 - 363i$ $a_{88} = 408 - 358i$

The spectrum, $\sigma(A)$, is

$$\{-509 + 290i, -500 + 312i, -303 - 327i, -298 - 347i,$$

$$454 + 520i, 437 + 544i, 422 - 363i, 407 - 358i\}.$$

The Gerschgorin disks and numerical range for this matrix are shown in figure 8.3A. Again, the Gerschgorin set is made up of a union of disjoint sets. This time, however, the distance between the sets is large compared to their size. This is an indication that the numerical range's performance will be poor. As can be seen in the figure, the area covered by the numerical range huge compared to the area covered by the Gerschgorin disks. The Convex nature of the numerical range results in a great deal of 'wasted' area while the flexibility of the Gerschgorin disks results in a small inclusion set. The Cassini set (figure 8.3B), of course, is slightly smaller yet.

Recall from chapter one that the pre-Gerschgorin methods, which form the foundation of what we call the BBFP, produced sets that included the numerical range. Therefore, the Composite BBFP set grows with the numerical range as can be seen in figure 8.3C. The pseudospectra continues to produce a very small inclusion set (figure 8.3D). In reality, the pseudospectra set can produce an even smaller set than the set shown here. (This setting was used so that it can be visible). So, the pseudospectra set for this example can be considerably smaller than the Gerschgorin set giving the pseudospectra, once again, the sharpest inclusion set.

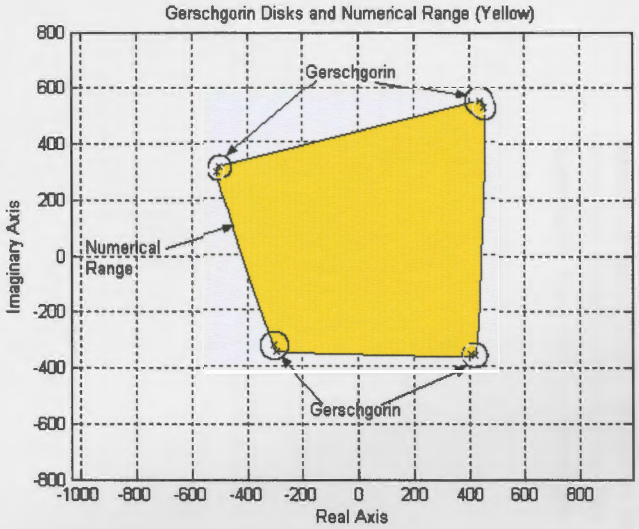


Figure 8.3A

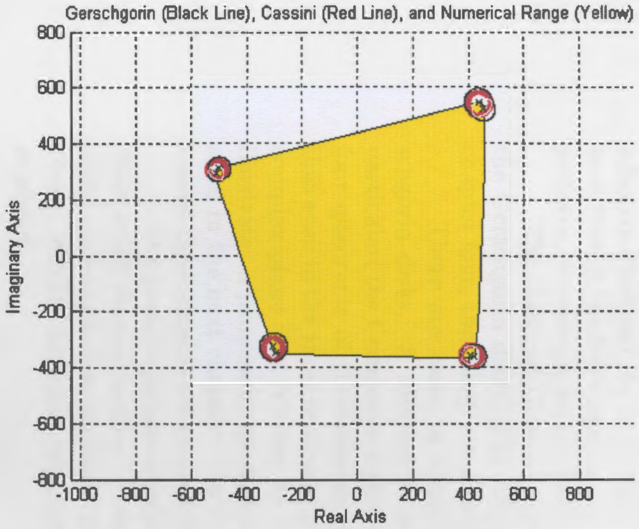


Figure 8.3B

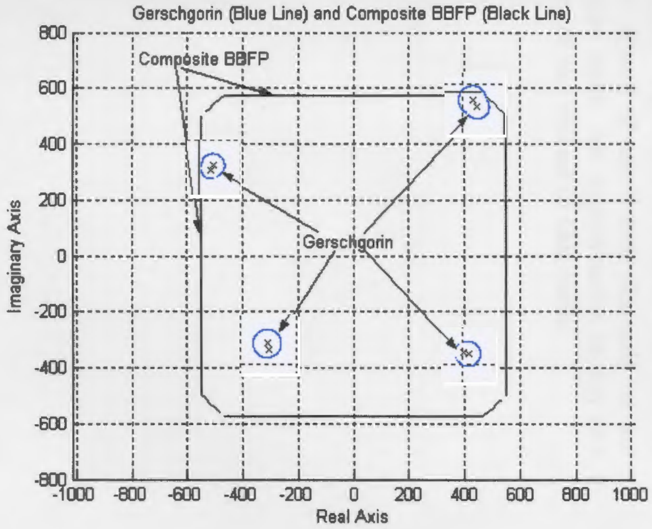


Figure 8.3C

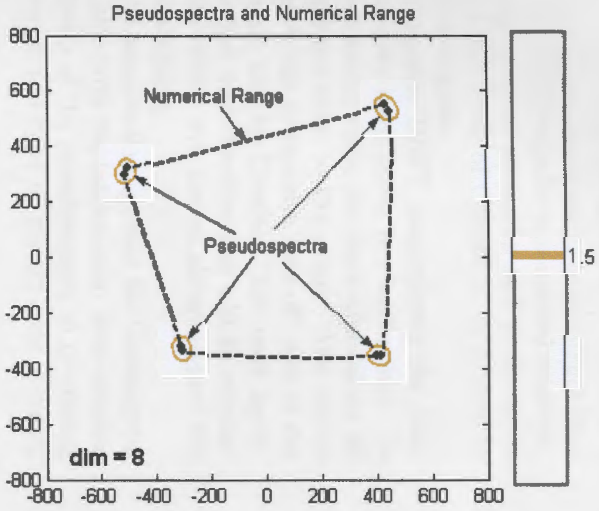


Figure 8.3D

All of the graphs on this page are for Example 8.3

The preceding examples show that the relative performance of the Gerschgorin disks and the numerical range depend upon the application. Therefore, even though the Gerschgorin disks are simpler and much easier to construct they perform, on *average*, just as well as the numerical range for non-normal matrices.

The Ovals of Cassini have the same general strengths and weakness as the Gerschgorin disks. The advantage of the Ovals of Cassini is, as proven in chapter four, that they are a subset of Gerschgorin.

There are cases in which the Composite BBFP outperforms the Gerschgorin disks. A number of such examples were given in chapter one. In particular, the Composite BBFP performs better when the eigenvalues cover all four quadrants of the complex plane and are close to the origin. Also notice that even in the three examples given above, the Composite 'cuts off' part of the complex plane that is cover by Gerschgorin and by Cassini. This, once again, illustrates a principal expounded in the first two chapters and will be emphasized in the next chapter - the Composite BBFP will almost always reduce the size of the inclusion set produced by Gerschgorin.

The pseudospectra outperforms the numerical range and the Gerschgorin disks in all three examples. Combining this with the information from chapter seven once again demonstrates the efficiency of the pseudospectra in producing sharp, reliable inclusion sets irrespective of the type of matrix that is being analyzed.

In the tables below, an attempt will be made to rate each of the spectral inclusion methods according to speed and sharpness. It should be stressed that, even though a great deal of analysis and computer calculations went into producing this table, the numbers in these tables are very subjective. For as has been demonstrated, the relative sharpness of many of the inclusion sets is dependent upon the matrix being analyzed. The same is true for the relative speed of the various inclusion methods. Speed of calculation depends on not only the matrix but also on the algorithm used. An improvement in any one algorithm may change the order of many of the entries in this table.

Method	**Relative Speed**	
Gerschgorin	1	Fastest
Browne	2	
Brauer	2	
Farnell's First	2	
Parker's First	2.5	
Farnell's Second	3	
Parker's Second	3	
Numerical Range	20	
Pseudospectra	20	
Cassini	25	
Minimal Gerschgorin	100	Slowest

Method	Relative Sharpness	
Pseudospectra	1	Sharpest
Minimal Gerschgorin	4	
Numerical Range	7	
Cassini	8	
Gerschgorin	10	
Parker's Second	15	
Farnell's Second	17	
Brauer	25	
Parker's First	28	
Farnell's First	35	
Browne	40	Least Sharp

9 The Intersection Method

This chapter will bring to completion an idea that has been utilized throughout this thesis - the intersection of spectral inclusion sets. In the past, some use has been made of this idea - Meyer [58] intersected the Gerschgorin set for a matrix and the Gerschgorin set of its transpose. Also, the Minimal Gerschgorin set is produced by intersecting the Gerschgorin sets of similar matrices. Beyond this, however, no use has been made of this idea of the Intersection of inclusion sets. On the other hand, it was shown in chapter one that the intersection of four or five different inclusion sets produce some surprisingly good results. In this chapter, this idea will be extended to include even more sets.

Intersection of sets

In chapter one a new theorem was presented that is based on the intersection of the inclusion sets of Browne, Parker, Farnell and Brauer. The inclusion set that results from this theorem is called the Composite BBFP. In chapter two, it was shown that the size of this inclusion set can be reduced farther by intersecting the Composite BBFP with Parker's Second (1948) Theorem. In this chapter, two more sets will be included in this intersection: the Gerschgorin set and the Gerschgorin set of the transpose of the matrix.

Definition 9A - Intersection of sets

Let $A \in C^{nxn}$. Let

$$B = \frac{A + A^*}{2}, \quad \text{and} \quad C = \frac{A - A^*}{2i}.$$

Let $R_{(A)i}$,$R_{(B)i}$,and $R_{(C)i}$,be the sums of the absolute values of the elements in the i^{th} row of the matrices A,B, and C, respectively. Let T_i be the sum of the absolute values of the elements in the i^{th} column of A. Define:

$$R_A = max\{R_{(A)1}, ..., R_{(A)N}\}, \quad R_B = max\{R_{(B)1}, ..., R_{(B)N}\},$$
$$R_C = max\{R_{(C)1}, ..., R_{(C)N}\}, \quad \text{and} \quad T = max\{T_1, ..., T_N\},$$

Let

$$F_1 = min\{R_A, T\}.$$

Let U_i be the sum of the squares of the absolute values of the elements in the i^{th} row of A (i.e. $U_i = \sum_{j=1}^{n} |a_{ij}|^2\}$). Let V_i be the sum of the squares of the absolute values of the elements in the i^{th} column of A (i.e. $V_i = \sum_{k=1}^{n} |a_{ki}|^2\}$). Let

$$F_2 = \Big[\sum_{i=1}^{n} (U_i V_i)^{\frac{1}{2}}\Big]^{\frac{1}{2}} \quad \text{and} \quad F_3 = (R_A T)^{\frac{1}{2}}.$$

Let

$$F = min\{F_1, F_2, F_3\}.$$

Let

$$FS = \{(x,y) : \sqrt{x^2+y^2} \leq F\} \text{ and } BR = \{(x,y) : |x| \leq R_B \text{ and } |y| \leq R_C\}.$$

$$\text{Let } P_i = \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}| \quad \text{and} \quad Q_i = \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ji}| \quad \text{for} \quad 1 \leq i \leq n.$$

Let

$$\mu = \frac{1}{n}\sum_{i=1}^{n} a_{ii}.$$

Let

$$R_i = P_i + |a_{ii} - \mu|, \; T_i = Q_i + |a_{ii} - \mu| \text{ and } S_i = (R_i + T_i)/2 \text{ for } 1 \leq i \leq n.$$

Setting

$$S = max\{S_i, ..., S_n\} \text{ and } PK = \{z \in C : |z - \mu| \leq S\}.$$

Let

$$G_i(A) = \{z \in C : |z - a_{ii}| \leq r_i(A) = \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}|\} \text{ for } 1 \leq i \leq n.$$

Let

$$G(A) = \bigcup_{i=1}^{n} G_i(A).$$

Let

$$G_j^{\mathbf{T}}(A) = \{z \in C : |z - a_{jj}| \leq r_j(A) = \sum_{\substack{i=1 \\ i\neq j}}^{n} |a_{ij}|\} \text{ for } 1 \leq i \leq n.$$

Let

$$G^{\mathbf{T}}(A) = \bigcup_{j=1}^{n} G_j^{\mathbf{T}}(A).$$

Then

$$\sigma(A) \subseteq FS \cap BR \cap PK \cap G(A) \cap G^{\mathbf{T}}(A).$$

● ● ●●

The intersection of all of these sets can best be illustrated by examples.

Example 9.1 Let

$$\mathbf{A} = \begin{pmatrix} -35-75i & 2 & -6 & 15i \\ 7-6i & -35-35i & -7+3i & 8+5i \\ 19 & 8-4i & -75-35i & 10 \\ 16+19i & 15-2i & -9+3i & -75-75i \end{pmatrix}$$

The spectrum, $\sigma(A)$, is

$$\{-71.4 - 82.4i, -43.9 - 70.2i, -71.1 - 33.1i, -33.6 - 34.4i\}.$$

A number of spectral inclusion sets for this matrix are shown in figures 9.1A and 9.1B. Among the sets shown are Parker(1948), Gerschgorin, Gerschgorin for the transpose of the matrix, Browne, Brauer, Parker's First, Farnell's First, Farnell's Second and the 'rectangular box' that bounds the Real and Imaginary parts of the eigenvalues. (All of these sets were discussed in chapter one and two of this thesis.) With the exception of the inclusion sets of Browne and Parker's First, all of the inclusion sets are of different sizes and many of them are of different shapes.

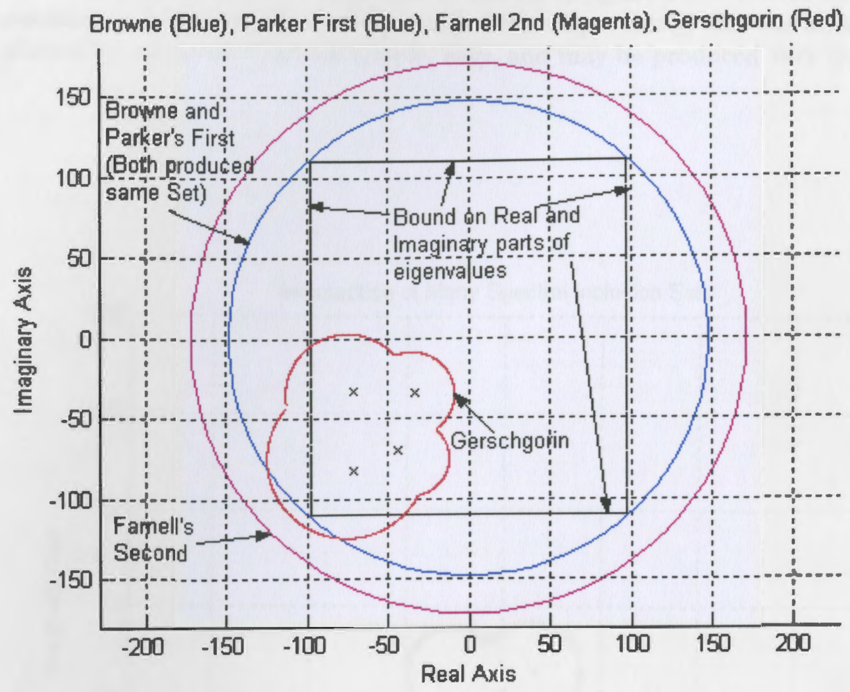


Figure 9.1A

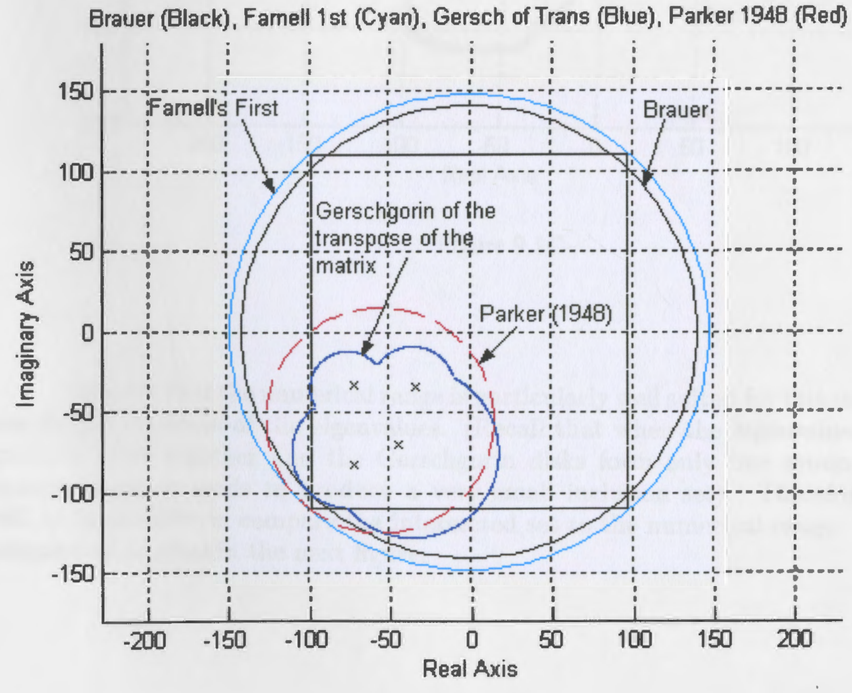


Figure 9.1B

The intersection of these sets is shown in figure 9.1C. Notice that the resulting inclusion set is relatively small. More importantly, this was all accomplished by using sets that are simple, easy, and may be produced very quickly.

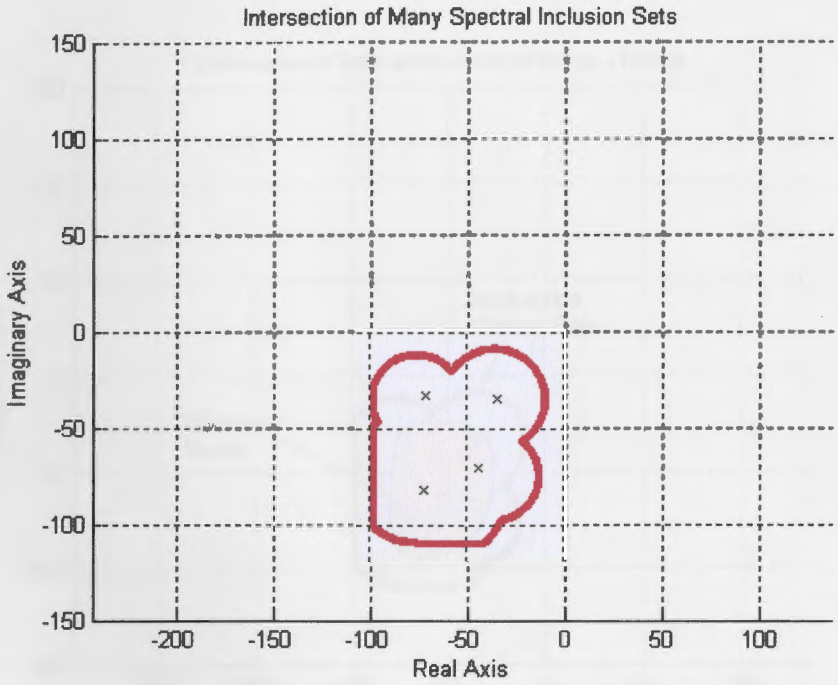


Figure 9.1C

Observe that the numerical range is particularly well suited for this matrix due to the closeness of the eigenvalues. (Recall that when the eigenvalues are grouped close together and the Gerschgorin disks form only one group, the numerical range tends to produce a very small inclusion set). Therefore, it will be instructive to compare the intersected set to the numerical range. This comparison is done in the next figure.

Figure 9.1D compares the intersected set with the numerical range. As expected, the numerical range did produce a smaller inclusion set than the intersected set. However, the intersected set is very 'competitive' being only a bit larger than the numerical range. This is very encouraging because the numerical range might be expected to be significantly 'sharper' in this particular case.

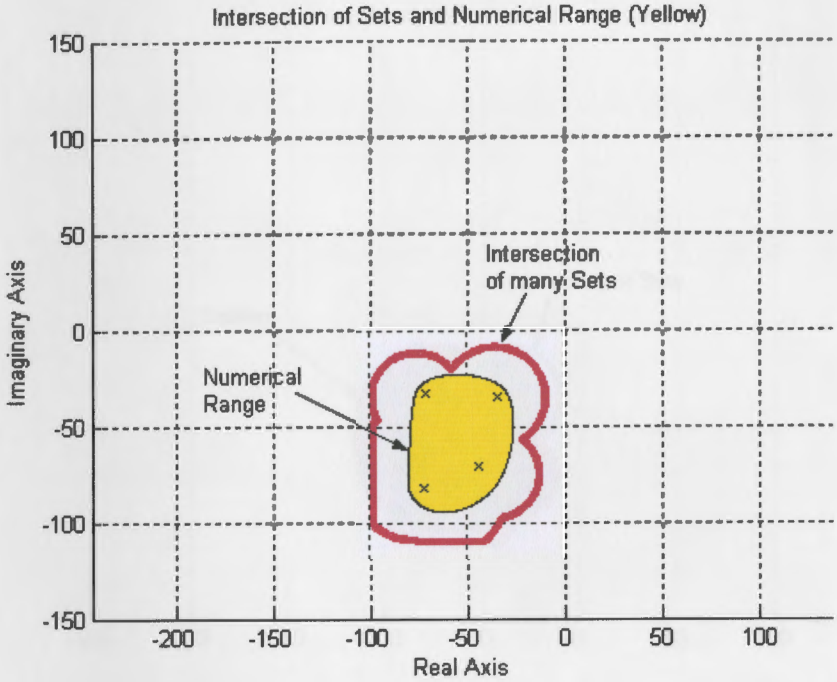


Figure 9.1D

The next figure, 9.1E, compares the intersected set to the Cassini set. The intersected set covers slightly less area in the complex plane than Cassini. Therefore, the intersected set, in this case, can be considered slightly 'sharper' than Cassini. This is significant because the Cassini set is almost always much sharper than any *one individual* set included in this intersection. Comparison was also made to the minimal Gerschgorin set (not shown). The minimal Gerschgorin set was comparable in size to the numerical range for this matrix. Therefore, the minimal Gerschgorin set was slightly smaller than the intersected set.

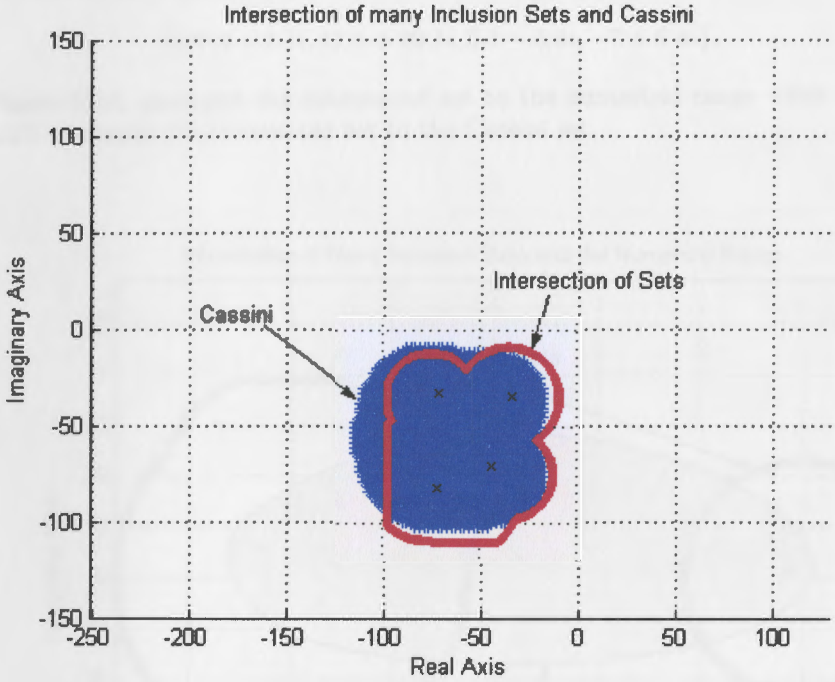


Figure 9.1E

The next two examples continue these comparisons.

Revisiting Example 1.4,
Example 9.2 Recall

$$\mathbf{A} = \begin{pmatrix} 100 & 2 & -6 & 15i \\ 7-6i & 15i & -7+3i & 8+5i \\ 19 & 8-4i & 13+9i & 10 \\ 16+9i & 15-2i & -9+3i & 0 \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{97.3 + 2.1i, 17.5 + 20.1i, 5.1 - 4.6i, -7 + 6.4i\}.$$

Figure 9.2A compares the intersected set to the numerical range while figure 9.2B compares the intersected set to the Cassini set.

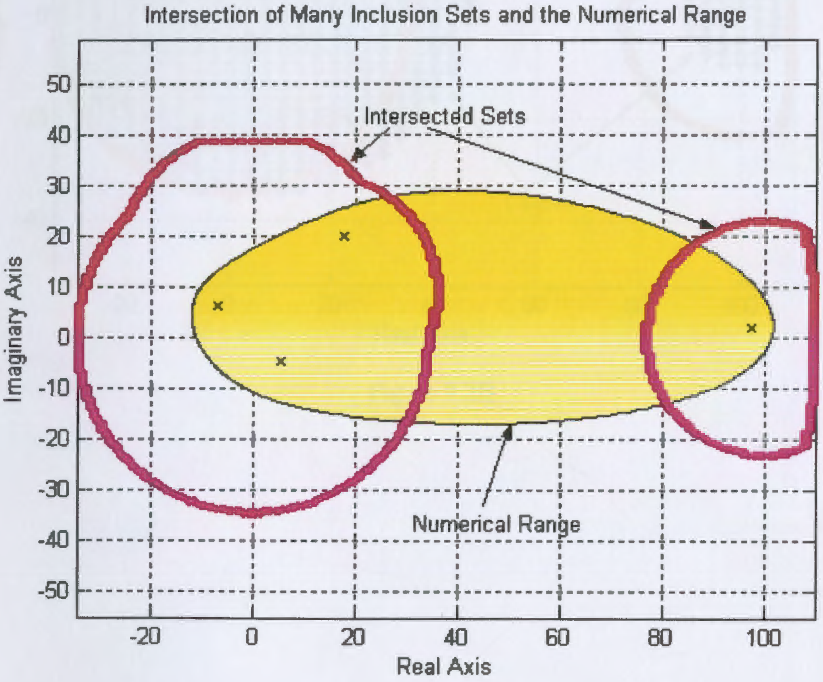


Figure 9.2A

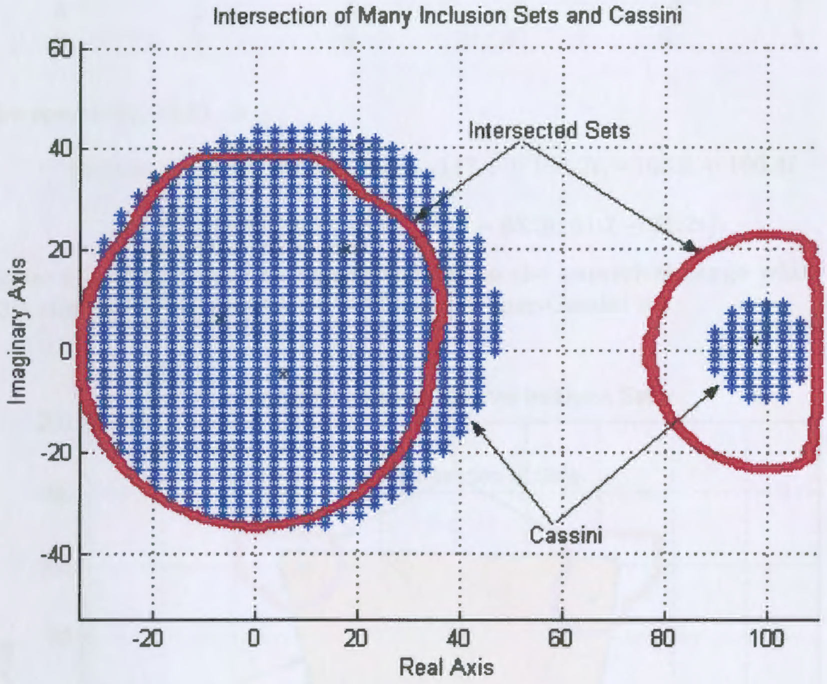

Intersection of Many Inclusion Sets and Cassini
Intersected Sets
Cassini
Imaginary Axis
Real Axis
60
40
20
0
-20
-40
-20
0
20
40
60
80
100

Figure 9.2B

Example 9.3 Let

$$\begin{pmatrix} 95+77i & 3 & 2 & 6 & 8 & 2 & 4 & 7 \\ 1 & -105+99i & 4 & 5 & -6 & 3 & 2 & 8 \\ 3 & 4 & 110+105i & 8 & -2 & 5 & 9 & 1 \\ 5 & 7 & 8 & -115+110i & 4 & -2 & 2 & 4 \\ 2 & 4 & 5 & 5 & -55-60i & 1 & 3 & 2 \\ 1 & -3 & 6 & 7 & 2 & 65-60i & 4 & 5 \\ 6 & 2 & 4 & -6 & -4 & 7 & 72-65i & 4 \\ 7 & 3 & 3 & 4 & 6 & 8 & -9 & -80-62i \end{pmatrix}$$

The spectrum, $\sigma(A)$, is

$$\{110.55+104.5i, 95.3+76.8i, -117.1+108.7i, -103.5+100.4i,$$

$$-71.4-65i, -56-60i, 75.3-63.2i, 61.7-61.2i\}.$$

Figure 9.3A compares the Intersected set to the numerical range while figure 9.3B compares the Intersected set to the Brauer-Cassini set.

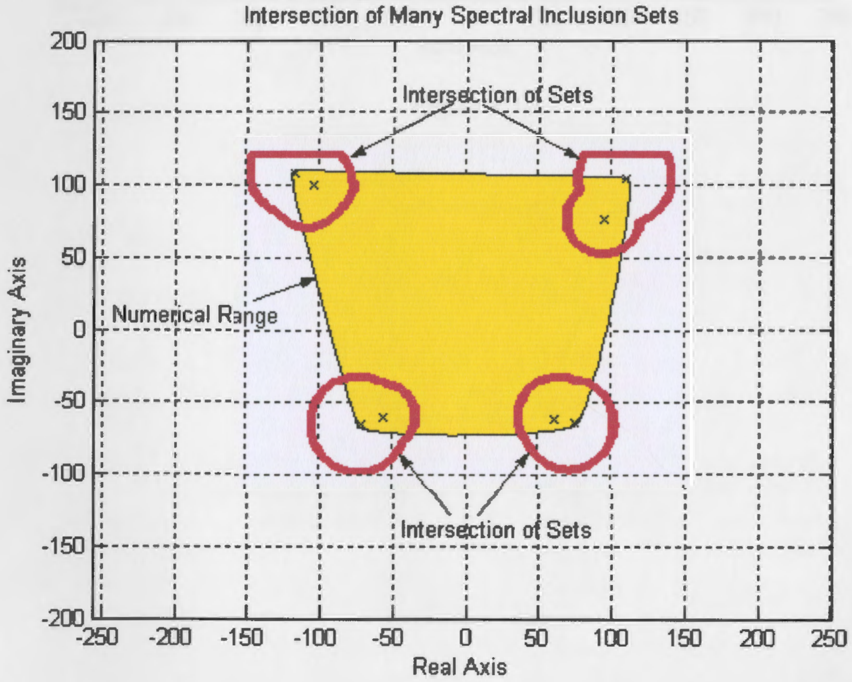


Figure 9.3A

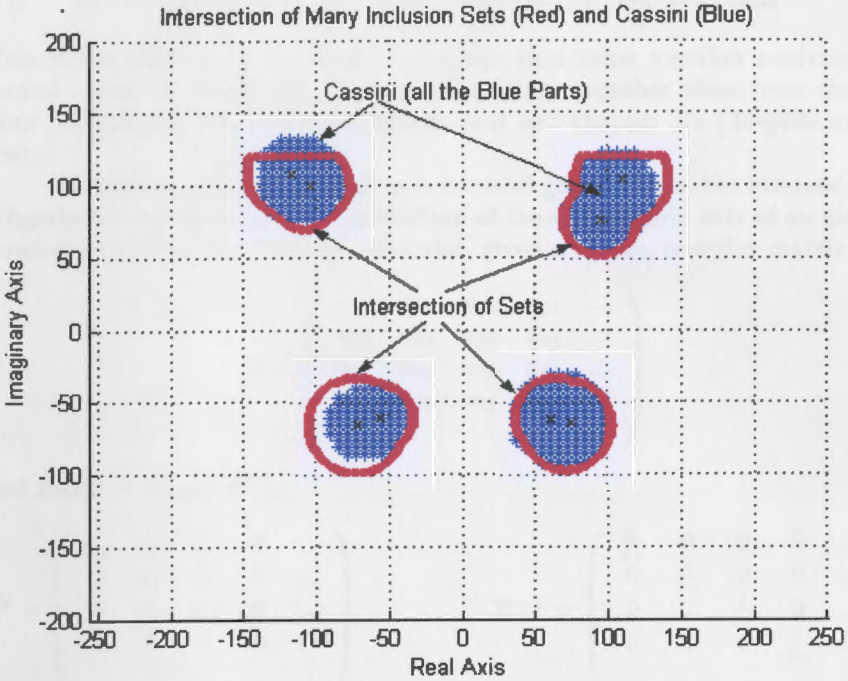

Intersection of Many Inclusion Sets (Red) and Cassini (Blue)
Cassini (all the Blue Parts)
Intersection of Sets
Imaginary Axis
Real Axis
200
150
100
50
0
-50
-100
-150
-200
-250
-200
-150
-100
-50
0
50
100
150
200
250

Figure 9.3B

10 A New Method for Toeplitz Matrices

This is the third of three 'Results' chapters that bring together analysis presented earlier in the thesis. This chapter brings together ideas from chapter Four (specifically Minimal Gerschgorin sets) and chapter Six (Toeplitz matrices).

Recall from chapter Four that,in its most general form, the Minimal Gerschgorin set comprises of the intersection of the Gerschgorin sets of an infinite number of similar matrices. In particular, given a square, complex matrix A,

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & \dots \\ a_{21} & a_{22} & a_{23} & a_{24} & \dots \\ a_{31} & a_{32} & a_{33} & a_{34} & \dots \\ a_{41} & a_{42} & a_{43} & a_{44} & \dots \\ . & . & . & . & . \end{pmatrix}$$

and matrices P and P^{-1},

$$\mathbf{P} = \begin{pmatrix} x_1 & 0 & 0 & 0 & \dots \\ 0 & x_2 & 0 & 0 & \dots \\ 0 & 0 & x_3 & 0 & \dots \\ 0 & 0 & 0 & x_4 & \dots \\ . & . & . & . & . \end{pmatrix} \qquad \mathbf{P^{-1}} = \begin{pmatrix} \frac{1}{x_1} & 0 & 0 & 0 & \dots \\ 0 & \frac{1}{x_2} & 0 & 0 & \dots \\ 0 & 0 & \frac{1}{x_3} & 0 & \dots \\ 0 & 0 & 0 & \frac{1}{x_4} & \dots \\ . & . & . & . & . \end{pmatrix}$$

with similarity transformations,

$$\mathbf{B = P^{-1}AP} = \begin{pmatrix} \frac{a_{11}x_1}{x_1} & \frac{a_{12}x_2}{x_1} & \frac{a_{13}x_3}{x_1} & \frac{a_{14}x_4}{x_1} & \dots \\ \frac{a_{21}x_1}{x_21} & \frac{a_{22}x_2}{x_2} & \frac{a_{23}x_3}{x_2} & \frac{a_{24}x_4}{x_2} & \dots \\ \frac{a_{31}x_1}{x_3} & \frac{a_{32}x_2}{x_3} & \frac{a_{33}x_3}{x_3} & \frac{a_{34}x_4}{x_3} & \dots \\ \frac{a_{41}x_1}{x_4} & \frac{a_{42}x_2}{x_4} & \frac{a_{43}x_3}{x_4} & \frac{a_{44}x_4}{x_4} & \dots \\ . & . & . & . & . \end{pmatrix}.$$

Then the minimal Gerschgorin set of A is equal to the intersection of the Gerschgorin disks of all similar matrices B such that $\mathbf{x} = (x_1, x_2, x_3, ..., x_n) > 0$.

Richard Varga and others have done extensive research on and have greatly advanced the theory of Minimal Gerschgorin sets over the past fifty years. However, no practical method has been discovered to apply this theory to numerical applications. In order to produce a truly Minimal Gerschgorin set for a numerical application, it is necessary to calculate the Gerschgorin sets for an infinite number of similar matrices. This, of course, is impossible.

However, there appears to be a simpler way to produce the Minimal Gerschgorin set for real or complex matrices in which all the diagonal elements are the same. The following two theorems address this situation. Even though these theorems are simple and rather obvious, it does not appear that they have ever been stated or used previously.

Theorem 10.1 Let A be an nxn complex matrix, such that $a_{ii} = a_{jj}$ for $1 \leq i, j \leq n$. Then, the Gerschgorin set is equal to the set included in the single disk in the complex plane centered at a_{11} with radius:

$$r = \sup_{i} \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|$$

i.e.

$$G(A) = \{z \in C : |z - a_{11}| \leq \sup_{i} \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|\}.$$

Proof

Let A be an nxn complex matrix such that $a_{ii} = a_{jj}$ for $1 \leq i, j \leq n$. The ith Gerschgorin disk for this matrix is defined as:

$$G_i(A) = \{z \in C : |z - a_{ii}| \leq r_i(A) = \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|\}.$$

According to the Gerschgorin theorem, the inclusion set for this matrix is:

$$\begin{aligned} G(A) &= \bigcup_{i=1}^{n} G_i(A) \\ &= \bigcup_{i=1}^{n} \{z \in C : |z - a_{ii}| \leq r_i(A) = \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|\} \end{aligned}$$

since the diagonal elements are equal we can say,

$$\begin{aligned} &= \bigcup_{i=1}^{n} \{z \in C : |z - a_{11}| \leq r_i(A) = \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|\} \\ &= \{z \in C : |z - a_{11}| \leq \sup_{i} r_i(A) = \sup_{i} \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|\}. \end{aligned}$$

• • ••

Theorem 10.2 Let A be an nxn complex matrix, such that $a_{ii} = a_{jj}$ for $1 \leq i, j \leq n$. Then, there exists a real diagonal matrix P, with strictly positive diagonal elements such that:

$$B = P^{-1}AP,$$

and the Minimal Gerschgorin set of A is given by,

$$G^{R}(A) = \bigcap_{x > 0} G^{x}(A) = G(B)$$

$$= \{z \in C : |z - a_{11}| \leq \sup_i r_i(B) = \sup_i \sum_{\substack{j=1 \\ j \neq i}}^{n} |b_{ij}|\}.$$

Proof

Let A be an nxn complex matrix such that $a_{ii} = a_{jj}$ for $1 \leq i, j \leq n$. Let P be a diagonal, real matrix such that the diagonal elements are strictly positive. Then $B = P^{-1}AP$ is a similarity transformation of A. Since the diagonal entries of A are all equal to a_{11} then $b_{ii} = b_{jj} = a_{11}$ for $1 \leq i, j \leq n$.

By Theorem 10.1,

$$G(B) = \{z \in C : |z - b_{11}| \leq \sup_i r_i(B) = \sup_i r_i(P^{-1}AP) = \sup_i \sum_{\substack{j=1 \\ j \neq i}}^{n} |b_{ij}|\}.$$

Now, the Minimal Gerschgorin set of A is defined as the intersection of an infinite number of these transformations,

$$G^R(A) = \bigcap_{\mathbf{x}>0} \{z \in C : |z - a_{11}| \leq \sup_i r_i(B) = \sup_i r_i(P^{-1}AP) = \sum_{\substack{j=1 \\ j \neq i}}^{n} |b_{ij}|\}.$$

But since each set on the right side of the equation above is a circle centered at a_{11}, this becomes:

$$G^R(A) = \inf_P \{z \in C : |z - a_{11}| \leq \sup_i r_i(B) = \sup_i r_i(P^{-1}AP) = \sum_{\substack{j=1 \\ j \neq i}}^{n} |b_{ij}|\}.$$

Which is the set included within a single disk centered at a_{11}.

● ● ●●

In words, these theorems together say that, given a square, complex matrix in which all of the diagonal elements are the same, then there exists a **single** transformation such that the Gerschgorin set of that single similar matrix is equal to the Minimal Gerschgorin set (i.e. equal to the intersection of Gerschgorin sets of an infinite number of similar matrices)!

Is it possible to find this transformation? The answer is 'Yes, for certain types of Toeplitz matrices'.

Section 10.1 Practical Implications

Below, an algorithm will be presented to find the minimal Gerschgorin set for a certain type of Toeplitz matrix.

Section 10.1.1 Hessenberg - Toeplitz .

Consider the nxn complex Hessenberg matrix

$$\mathbf{A} = \begin{pmatrix} a_0 & a_1 & 0 & 0 & 0 & \dots \\ a_{-1} & a_0 & a_1 & 0 & 0 & \dots \\ a_{-2} & a_{-1} & a_0 & a_1 & 0 & \dots \\ a_{-3} & a_{-2} & a_{-1} & a_0 & a_1 & \dots \\ a_{-4} & a_{-3} & a_{-2} & a_{-1} & a_0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

The goal is to find that one transformation that will produce the Minimal Gerschgorin set. One way to set about finding this transformation is to consider the Gerschgorin disks of a typical similar matrix such as B.

$$\mathbf{B} = \mathbf{P}^{-1}\mathbf{AP} = \begin{pmatrix} a_0 & \frac{a_1 x_2}{x_1} & 0 & 0 & \dots \\ \frac{a_{-1} x_1}{x_2} & a_0 & \frac{a_1 x_3}{x_2} & 0 & \dots \\ \frac{a_{-2} x_1}{x_3} & \frac{a_{-1} x_2}{x_3} & a_0 & \frac{a_1 x_4}{x_3} & \dots \\ \frac{a_{-3} x_1}{x_4} & \frac{a_{-2} x_2}{x_4} & \frac{a_{-1} x_3}{x_4} & a_0 & \dots \\ . & . & . & . & . \end{pmatrix}$$

where, $x_1, ..., x_n > 0$.
The focus will be on the sum of the off-diagonal elements of each row of B. In order the reach that goal, it will be necessary to make the largest of these sums as small as possible. That is, choose x_1, x_2, x_3, etc. so that the sums of the absolute values of the off-diagonal elements of each row are small. Since it is not possible to explicitly solve for the x_i's in this way, it will be necessary to develop an iterative process to achieve the same end.

Algorithm 10.1

Notice that the sum for a typical row i is,

$$r_i = \frac{|a_1| x_{i+1}}{x_i} + \frac{|a_{-1}| x_{i-1}}{x_i} + \frac{|a_{-2}| x_{i-2}}{x_i} + ...$$

As stated above, the goal is to make the largest of the r_i's as small as possible. One way to do this is to pick a sum, call this sum T, solve for the x_i's based on T and see if all of the x_i's are greater than zero and the sum of the absolute values of the off-diagonal elements of the last row less than or equal to T.

$$T = \frac{|a_1| x_{i+1}}{x_i} + \frac{|a_{-1}| x_{i-1}}{x_i} + \frac{|a_{-2}| x_{i-2}}{x_i}.$$

$$\Rightarrow \quad T = \frac{|a_1| x_{i+1}}{x_i} + \sum_{j=1}^{i-1} \frac{|a_{-j}| x_{i-j}}{x_i} + ...$$

$$\Rightarrow \qquad Tx_i = |a_1|x_{i+1} + \sum_{j=1}^{i-1} |a_{-j}|x_{i-j}.$$

$$\Rightarrow \qquad x_{i+1} = (Tx_i - \sum_{j=1}^{i-1} |a_{-j}|x_{i-j})/|a_1|. \qquad (Eq.10.1)$$

If all of the x_i's turn out to be positive and the sum of the absolute values of the off-diagonal elements of the last row are less than or equal to T, then there is a transformation that will produce a matrix that is similar to A and that matrix will have a Gerschgorin set with radius T centered at a_{11}. (Notice that the reason for checking the last row is because the last row elements were not considered when the x_i's were calculated).

This process is repeated using smaller and smaller values of T. The smallest value of T that meets the criteria (i.e. all of the x_i's are be positive and the sum of the absolute values of the off-diagonal elements of the last row are less than or equal to T), is the radius of the Minimal Gerchgorin set of A centered at a_{11}.

This algorithm will be illustrated in the next example.

Example 10.3 Consider the 5x5 matrix

$$\mathbf{A} = \begin{pmatrix} 6 & 5 & 0 & 0 & 0 \\ 7 & 6 & 5 & 0 & 0 \\ 8 & 7 & 6 & 5 & 0 \\ 0 & 8 & 7 & 6 & 5 \\ 0 & 0 & 8 & 7 & 6 \end{pmatrix}.$$

Since all of the diagonal elements are the same (equal to 6) and the largest off-diagonal row sum is 20 (i.e. row three or row four is $|8| + |7| + |5| = 20$) then, by Theorem 10.1, the spectrum is included in the set bounded by a disk centered at (6,0) with radius = 20.

So, an attempt will be made to find a similar matrix that has a Gerschgorin disk with a radius less than 20. First, an attempt will be made to find a similar matrix with radius = 19. Equation 10.1 will be used with T=19. setting $x_1 = 1$,

Row 1, $i = 1$

$$x_2 = (19(1) - 0)/5 = 3.8.$$

Row 2, $i = 2$

$$x_3 = (19(3.8) - 7(1))/5 = 13.04.$$

Row 3, $i = 3$

$$x_4 = (19(13.04) - [7(3.8) + 8(1)])/5 = 42.632.$$

Row =4, $i = 4$

$$x_5 = (19(42.632) - [7(13.04) + 8(3.8)])/5 = 137.66.$$

So, values have been found for x_1, x_2, x_3, x_4, x_5 based on off-diagonal row sums of 19. This was done by using rows one through four but notice that the last row was not included in the calculations. So, the last row must be checked to be sure that the sum of the absolute values of the off-diagonal elements of row 5 of this similar matrix is less than or equal to 19:

$$\begin{aligned} r_5(B) &= a_{-4}\frac{|x_1|}{|x_5|} + a_{-3}\frac{|x_2|}{|x_5|} + a_{-2}\frac{|x_3|}{|x_5|} + a_{-1}\frac{|x_4|}{|x_5|} \\ &= (0)\frac{1}{137.66} + (0)\frac{3.8}{137.66} + (8)\frac{13.04}{137.66} + (7)\frac{42.632}{137.66} \\ &= 2.9255. \end{aligned}$$

Since this last sum is equal to or less than 19 *and* the $x_1, ..., x_5$ were strictly positive, there exists a similar matrix with a Gerschgorin set consisting of a disk centered at (6,0) with radius equal to 19.

Since 19 worked, perhaps a lower sum will also work. So, 18, 17, etc were attempted with the following results:

Off-diagonal row sum	**Result**
18	Works
17	Works
16	Works
15	Works
14	Works
13	Works
12	Fails (Off-diagonal elements of last row sum up to 22.479 which is larger than 12).

Notice that an off-diagonal row sum of 13 works but 12 fails. So, the 'search' can be refined further:

Off-diagonal row sum	**Result**
12.9	Works
12.8	Works
12.7	Works
12.6	Works
12.5	Fails

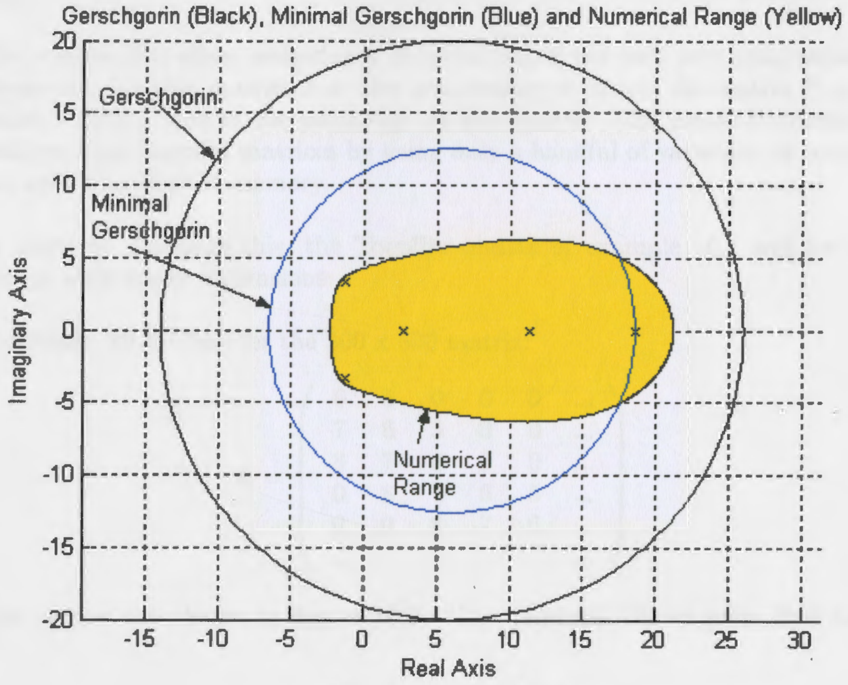
(Off-diagonal elements of last row sum to 12.99 which is larger than 12).

Continuing this process results in T=12.546 or a disk centered at (6,0) with a radius of 12.546. This is the Minimal Gerschgorin set of the matrix A (figure 10.3). Observe that the spectrum, $\sigma(A)$, is

$$\{18.5453, 12.3964, 2.7059, -1.27386 + 3.325i, -1.27386 - 3.325i\}.$$

Notice that the disk that forms the boundary of the Minimal Gerschgorin set practically intersects one of the eigenvalues. This example was particulary nice in that the Gerschgorin disk actually intersected one of the eigenvalues. This method will not produce such sharp results in all applications. However this method will *always* produce the minimal Gerschgorin set for the matrix.

It should also be noted that each of the 'failures' listed above were due to the off diagonal row sum of the last row being too large. However, it is often the case that the failure will be due to one of the x_i's going negative.



Toepltiz 5 x 5 Matrix
Figure 10.3

Optimizing the Algorithm

Notice in the algorithm that each x_{i+1} was calculated by using only the previous three or four x_i's. So, it was not necessary to have the whole vector **x** available when doing these calculations. As it turns out, it is only necessary to have available the number of elements of **x** equal to the Toeplitz bandwidth plus one. All of the other elements may be 'thrown away'. That is, in example 10.1, when calculating x_5 it was only necessary to have elements x_4, x_3 and x_2 available. When calculating x_{31} it was only necessary to have elements x_{30}, x_{29} and x_{28} available, etc.

Algorithm 10.2

This means that Algorithm 10.2 can be slightly modified: each element of **x** is still evaluated using

$$x_{i+1} = (Tx_i - \sum_{j=1}^{i-1} |a_{-j}| x_{i-j}) / |a_1|$$

but now the vector **x** only has the number of elements equal to the Toeplitz bandwidth plus two. Therefore, at each iteration, the elements are reassigned. In a Matlab program, something such as the following might be used:

```
for j=1:k
    $x_j = x_{j+1}$
end
```

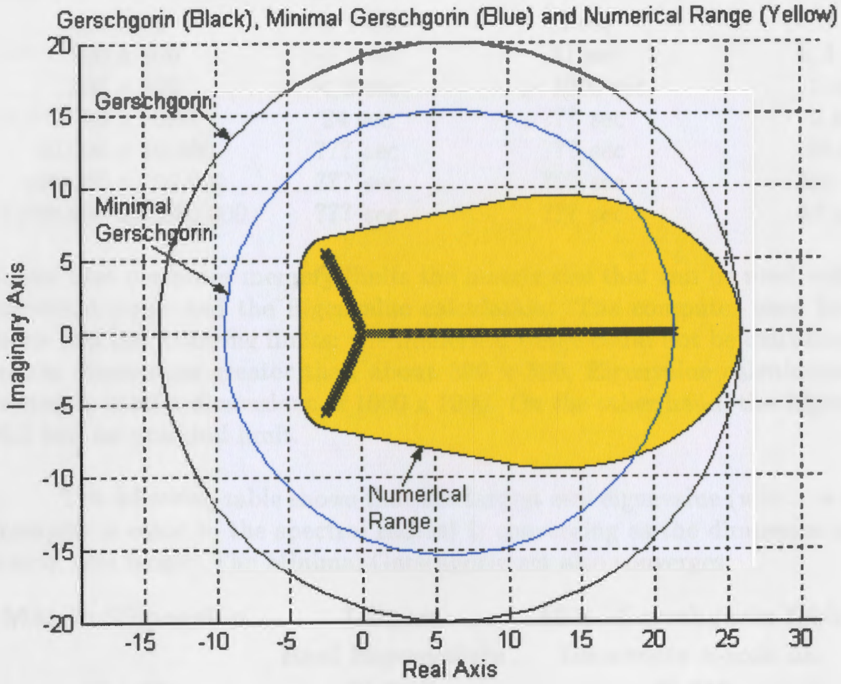
This means that when analyzing a Toeplitz matrix not only is it unnecessary to create the Toeplitz matrix, it is also unnecessary to create the matrix P or the entire vector **x** used in our similarity transformation. This makes it possible to analyze *huge* Toeplitz matrices by using only a handful of variables on computers with very limited memory.

In order to illustrate this, the Toeplitz matrix of example 10.1 will be used except with larger dimensions:

Example 10.6 Consider the 300 x 300 matrix:

$$\mathbf{A} = \begin{pmatrix} 6 & 5 & 0 & 0 & 0 & \dots \\ 7 & 6 & 5 & 0 & 0 & \dots \\ 8 & 7 & 6 & 5 & 0 & \dots \\ 0 & 8 & 7 & 6 & 5 & \dots \\ 0 & 0 & 8 & 7 & 6 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

The results are shown in figure 10.6. The Minimal Gerschgorin disk has a radius of 15.358 with center at (6,0). Notice how the bound on the Minimal Gerschgorin set coincides with the spectral radius: the Minimal Gerschgorin disk crosses the x-axis at 21.358. On the other hand, the largest real eigenvalue is 21.3561. With this size of matrix it is still possible to calculate the numerical range but, the execution time is very long.



Toeplitz 300 x 300 Matrix
Figure 10.6

The Matlab code to generate the Numerical Range in these figures was written by Cowen and Harel

Algorithm 10.2 was used on the matrix of example 10.6 using dimensions of 50 x 50 through 1,000,000 x 1,000,000. The results are summarized below. (Details and graphs may be found under examples 10.4 through 10.10 in the Appendix).

Calculation Time:

Matrix Dimension	Eigenvalues	Numerical Range	New Algorithm
50 x 50	< 1 sec	3 sec	< 1 sec
100 x 100	< 1 sec	21 sec	< 1 sec
300 x 300	< 2 sec	> 1000 sec	1 sec
1000 x 1000	24 sec	??? sec	2 sec
10,000 x 10,000	??? sec	??? sec	26 sec
100,000 x 100,000	??? sec	??? sec	268 sec
1,000,000 x 1,000,000	??? sec	??? sec	45 min

Notice that computer memory limits the matrix size that can be used with the numerical range and the Eigenvalue calculation. The computer used for this thesis had the following limits: the numerical range could not be calculated for matrix dimensions greater than about 300 x 300; Eigenvalue calculation was limited to matrix dimensions of 1000 x 1000. On the other hand, the Algorithm 10.2 had no practical limit.

The following table shows that the largest real eigenvalue (which in these examples is equal to the spectral radius) is converging as the dimension of the matrix gets larger. The Minimal Gerschgorin set also converges.

Matrix Dimension	Largest	Min. Gerschgorin Disk
	Real Eigenvalues	Intersects x-axis at:
50 x 50	21.3144	21.316
100 x 100	21.3464	21.348
300 x 300	21.3561	21.358
1000 x 1000	21.3573	21.358
10,000 x 10,000		21.358
100,000 x 100,000		21.358
1,000,000 x 1,000,000		21.358

The fact that the Minimal Gerschgorin set for this particular Toeplitz matrix converges *does not* make our algorithm any more stable or any less likely to be subject to roundoff errors. The algorithm must perform the same types of calculations and will be exposed to the same types of pitfalls whether the set converges or not. So, the fact that this particular example worked for a one million by one million matrix without error, while not *proving* the reliability of the algorithm, is very encouraging.

The times shown above will vary based on computer speed, memory, type of algorithm used, etc. but they show the relative speed of each of the methods.

The examples above utilized a relatively simple, real valued matrix with positive elements. Yet, Algorithm 10.2 will work for any square, complex Hessenberg matrix of practically any size as the next few examples will demonstrate.

Advanced Applications

The following examples utilize complex matrices with long Toeplitz bandwidths.

Example 10.11 Consider the 1000x1000 matrix:

$$\mathbf{A} = \begin{pmatrix} 6-3i & -6i & 0 & 0 & 0 & 0 & 0 & \dots \\ -4-7i & 6-3i & -6i & 0 & 0 & 0 & 0 & \dots \\ 3 & -4-7i & 6-3i & -6i & 0 & 0 & 0 & \dots \\ -5i & 3 & -4-7i & 6-3i & -6i & 0 & 0 & \dots \\ 2+2i & -5i & 3 & -4-7i & 6-3i & -6i & 0 & \dots \\ -8+3i & 2+2i & -5i & 3 & -4-7i & 6-3i & -6i & \dots \\ 0 & -8+3i & 2+2i & -5i & 3 & -4-7i & 6-3i & \dots \\ 0 & 0 & -8+3i & 2+2i & -5i & 3 & -4-7i & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

The results are shown in figure 10.11. The algorithm took 7 seconds while matlab required 66 seconds to calculate the actual eigenvalues. For the record, the radius of the minimal Gerschgorin disk is 17.7872 with the center at (6,-3). Notice that the Gerschgorin set is very sharp.

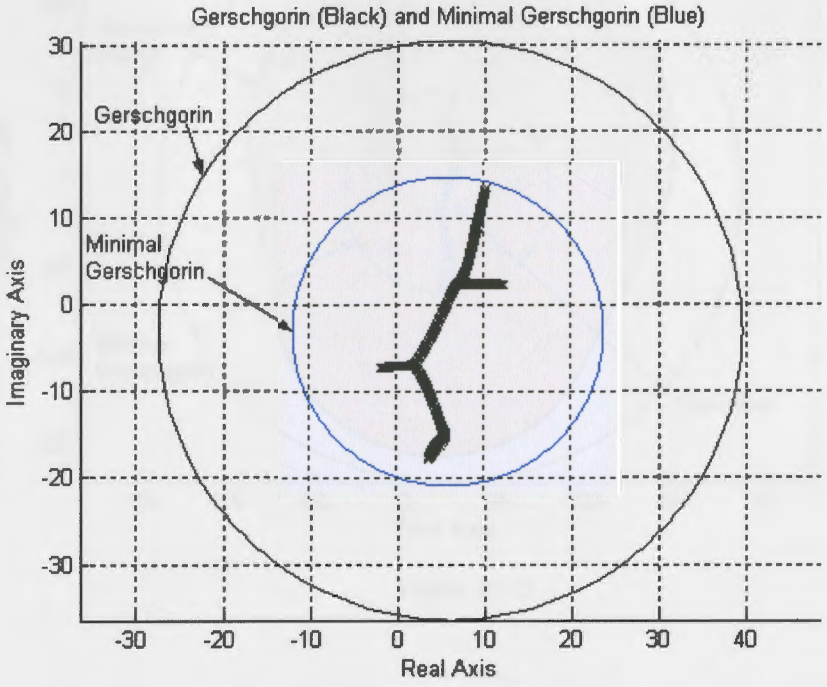


Figure 10.11

The next example shows that zeros in the bandwidth will not effect the Algorithm.

Example 10.12 Consider the 100x100 matrix:

$$\begin{pmatrix}
23-7i & 16i & 0 & 0 & 0 & 0 & \dots \\
47i & 23-7i & 16i & 0 & 0 & 0 & \dots \\
0 & 47i & 23-7i & 16i & 0 & 0 & \dots \\
0 & 0 & 47i & 23-7i & 16i & 0 & \dots \\
0 & 0 & 0 & 47i & 23-7i & 16i & \dots \\
53-20i & 0 & 0 & 0 & 47i & 23-7i & \dots \\
0 & 53-20i & 0 & 0 & 0 & 47i & \dots \\
0 & 0 & 53-20i & 0 & 0 & 0 & \dots \\
-42+13i & 0 & 0 & 53-20i & 0 & 0 & \dots \\
0 & -42+13i & 0 & 0 & 53-20i & 0 & \dots \\
0 & 0 & -42+13i & 0 & 0 & 53-20i & \dots \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots
\end{pmatrix}.$$

Figure 10.12 shows the Minimal Gerschgorin set calculated with Algorithm 10.2, Gerschgorin set, and the numerical range. Once again, the Minimal Gerschgorin set 'closes in' on the spectrum. On the other hand, the numerical range is not sharp at all.

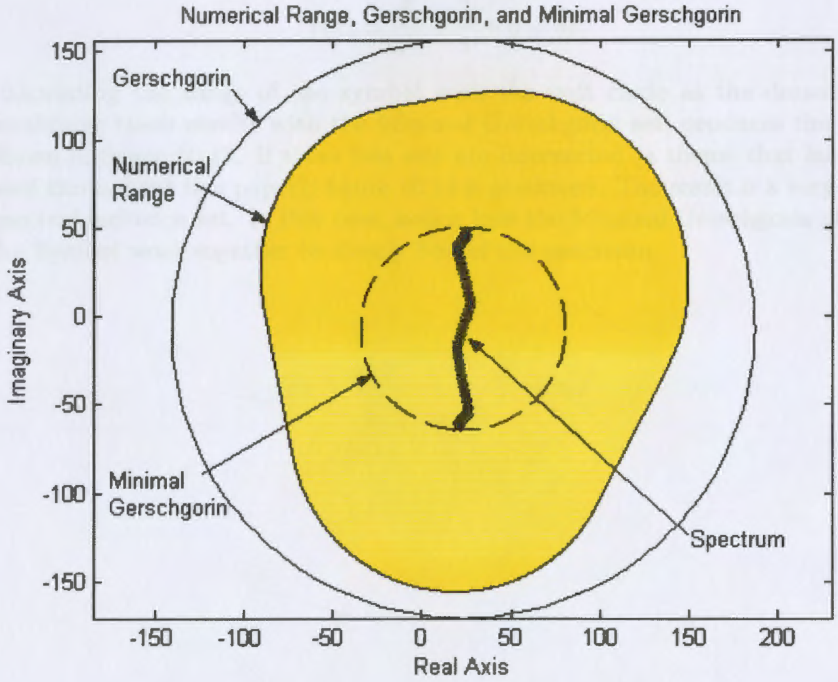


Figure 10.12

Bounding the inclusion set Further

The method described above produces a relatively small spectral inclusion set but the 'sharpness' of the set is limited in that the inclusion set is a circle. It would be desirable to reduce the size of this set even further. This will be attempted by using the symbol of the operator.

Recall from chapter six of this thesis that when one calculated the symbol of a matrix based on the unit circle, the result was the spectrum of the infinite dimensional operator. Furthermore, it was noted that the spectrum of the infinite dimensional operator can be used as a bound for the associated finite dimensional matrix. Therefore, these facts can be used here.

Revisiting example 10.6,
Example 10.13 Consider the 300 x 300 matrix:

$$\mathbf{A} = \begin{pmatrix} 6 & 5 & 0 & 0 & 0 & \dots \\ 7 & 6 & 5 & 0 & 0 & \dots \\ 8 & 7 & 6 & 5 & 0 & \dots \\ 0 & 8 & 7 & 6 & 5 & \dots \\ 0 & 0 & 8 & 7 & 6 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

The symbol of the associated infinite dimensional operator is

$$f(z) = \frac{8}{z^2} + \frac{7}{z} + 6 + 5z.$$

Calculating the range of the symbol with the unit circle as the domain and combining these results with the Minimal Gerschgorin set, produces the graph shown in figure 10.13. If these two sets are intersected (a theme that has been used throughout this paper), figure 10.14 is produced. The result is a very sharp spectral inclusion set. In this case, notice how the Minimal Gerschgorin set and the Symbol work together to closely bound the spectrum.

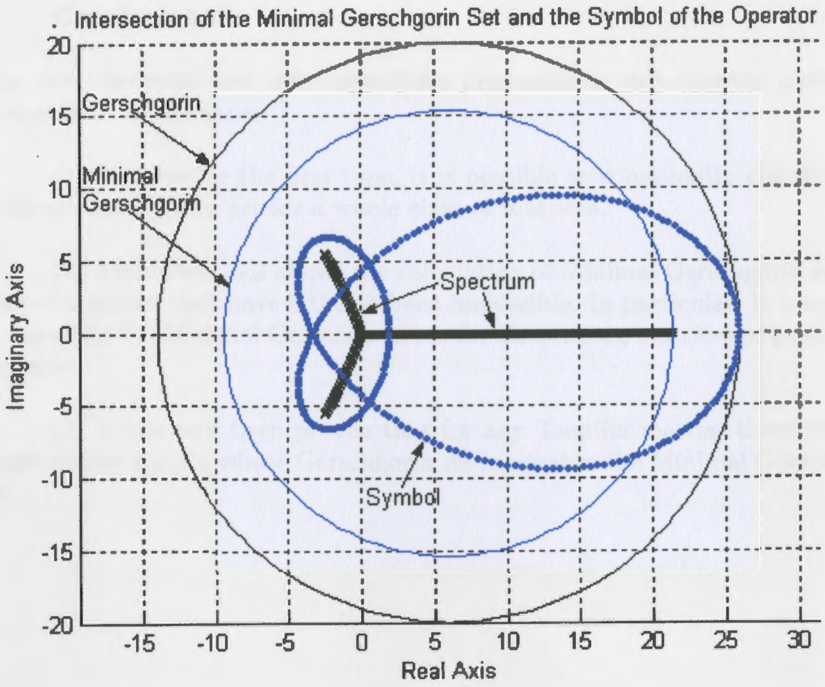


Figure 10.13

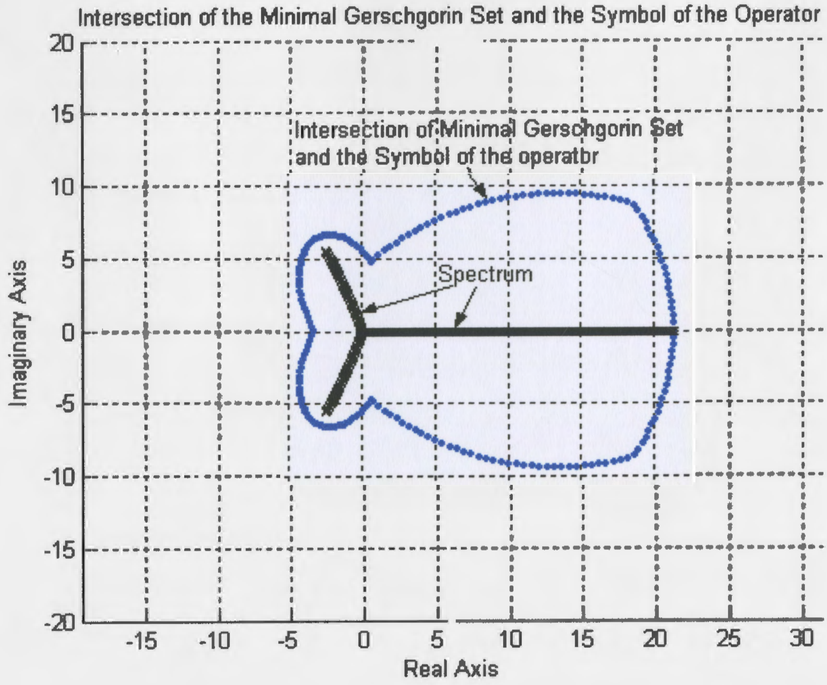


Figure 10.14
(This is the same as figure 10.13 except with the superfluous lines taken out)

Conclusions

The new theorems and new algorithms presented in this chapter contribute to a number of advances:

(1) Perhaps for the first time, it is possible to consistently calculate the Minimal Gerschgorin set for a whole class of matrices.

(2) Algorithm 10.2 allows the calculation of Minimal Gerschgorin sets for sizes of matrices that have hitherto been impossible. In particular, it is possible to calculate the Minimal Gerschgorin set for Hessenberg matrices of practically any size.

(3) It has now been proven that for any Toeplitz matrix, there exists a single similar matrix whose Gerschgorin set is equal to the Minimal Gerschgorin set.

References

[1] E.N.Barankin,*Bounds for the Characteristic Roots of a Matrix*, Bulletin of the American Mathematical Society. **51** (1945), pp. 767-770.

[2] A.Bottcher and B. Silbermann,*Introduction to Large Truncated Toeplitz Matrices* Springer, New York 1997.

[3] L.S.Boulton,*Non-Self-Adjoint Harmonic Oscillator Compact Semigroups and Pseudospectra*, J. of Operator Theory. **47** (2002), pp. 413-429.

[4] L.Brickman,*On the Field of Values of a Matrix*, Proceedings of the American Mathematical Society. **12** (1961), pp. 61-66.

[5] A.Brauer,*Characteristic Roots of a Matrix*, Duke Mathematical Journal. **13** (1946), pp. 387-395.

[6] A.Brauer,*Limits for the Characteristic Roots of a Matrix II*, Duke Mathematical Journal. **14** (Dec 1947), pp. 21-26.

[7] A.Brauer,*Limits for the Characteristic Roots of a Matrix III*, Duke Mathematical Journal. **15** (Dec 1948), pp. 871-877.

[8] A.Brauer,*Limits for the Characteristic Roots of a Matrix V*, Duke Mathematical Journal. **19 No. 4** (Dec 1952), pp. 553-562.

[9] A.Brauer,*Limits for the Characteristic Roots of a Matrix VII*, Duke Mathematical Journal. **25** (Dec 1958), pp. 583-590.

[10] A.Brauer and I.C.Gentry,*Bounds for the Greatest Characteristic Root of an Irreducible Nonnegative Matrix II*, Linear Algebra and Its Applications. **13** (1976), pp. 109-114.

[11] A.Brauer and H.T.LaBorde, *Limits for the Characteristic Roots of a Matrix. IV:*, Duke Mathematical Journal. **22 No. 2** (1955), pp. 253-261.

[12] A.Brauer and A.C.Mewborn, *The Greatest Distance Between Two Characteristic Roots of a Matrix.*, Duke Mathematical Journal. **26** (1959), pp. 653-661.

[13] E.T.Browne,*The Characteristic Roots of a Matrix* , Bulletin of the American Mathematical Society. **34** (1928), pp. 363-368.

[14] E.T.Browne,*The Characteristic Roots of a Matrix* , Bulletin of the American Mathematical Society. **36** (1930), pp. 705-710.

[15] E.T.Browne,*Limits to the Characteristic Roots of a Matrix* , American Mathematical Monthly. **46** (1939), pp. 252-265.

[16] R.A.Brualdi,*Matrices, Eigenvalues, and Directed Graphs*, Linear and Multilinear Algebra. **11** (1982), pp. 143-165.

[17] D.H.Carlson and R.S.Varga,*Minimal G-Functions. II*, Linear Algebra and Its Applications. **7** (1973), pp. 233-242.

[18] M.Chien,*On the Numerical Range of Tridiagonal Operators*, Linear Algebra and Its Applications. **246** (1996), pp. 203-214.

[19] T.Driscoll and L.N.Trefethen, *Pseudospectra for the Wave Equation with an Absorbing Boundary*, Journal of Computational and Applied Mathematics **69** (1996), pp. 125-142.

[20] T.E.Easterfield,*The Characteristic Roots of a Matrix:A Correction*, Duke Mathematical Journal. **23 No. 4** (1956), pp. 635-637.

[21] M.Embree and L.N.Trefethen,*Generalizing Eigenvalue Theorems to Pseudospectra Theorems.*, SIAM J. Sci. Comput. **23 No. 2** (2001), pp. 583-590.

[22] A.B. Farnell,*Limits for the Characteristic Roots of a Matrix*, Bulletin of the American Mathematical Society **vol. 50** (1944), pp. 789-794.

[23] M.Fiedler,*Some Estimates of the Proper Values of Matrices*, J. Soc. Inust. Appl. Math **13 No.1** (1965), pp. 1-5.

[24] S.Furtado and C.Johnson,*Spectral Variation under Congruence for a nonsingular matrix with 0 on the boundary of its field of values*, Linear Algebra and Its Applications. **359** (2003), pp. 67-78.

[25] G.Geldenhuys and C.Sippel,*Bounds for the Real Eigenvalues of a Cascade Matrix*, Linear Algebra and Its Applications. **93** (1987), pp.127-130.

[26] S.Gerschgorin,*Uber die Abgrenzung der Eigenwerte einer Matrix*, Isv. Akad. Nauk USSR **7** (1931), pp. 749-754.

[27] W.Givens,*Fields of Values of a Matrix*, Proceedings of the American Mathematical Society. **3** (1952), pp. 206-209.

[28] A.Greenbaum,*Generalizations of the Field of Values Useful in the Study of Polynomial Functions of a Matrix* , Linear Algebra and Its Applications. **347** (2002), pp. 233-249.

[29] K.Gustafson and D.Rao,*Numerical Range, the Field of Values of Linear Operators and Matrices*, Springer, New York 1997.

[30] C.A.Hall and T.A.Porsching, *Bounds for the Maximal Eigenvalue of a Nonnegative Irreducible Matrix.*, Duke Mathematical Journal. **36** (1969), pp. 159-164.

[31] P.Hartman and A.Wintner, *The Spectra of Toeplitz's Matrices*, American Journal of Mathematics. **76 No.4** (Oct. 1954), pp. 867-882.

[32] P.Henrici,*Bounds for Eigenvalues of Certain Tridiagonal Matrices*, J. Soc. Inust. Appl. Math **11 No.2** (1963), pp. 281-290.

[33] N.J.Higham and F.Tisseur,*More on Pseudospectra for Polynomial Eigenvalue Problems and Applications in Control Theory*, Linear Algebra and Its Applications. **351-352** (2002), pp. 435-453.

[34] D.Hinrichsen and A.J.Pritchard,*On Spectral Variations under Real Matrix Perturbations*, Numerische Mathematik **60** (1992), pp. 509-524.

[35] A.J.Hoffman and H.W.Wielandt,*The Variation of the Spectrum of a Normal Matrix*, Duke Mathematical Journal. **20 No.1** (1953), pp. 37-39.

[36] A.J.Hoffman and R.S.Varga,*Patterns of Dependence of Gerschgorin's Theorem*, SIAM J. Numerical Analysis. **7 No.4** (1970), pp. 571-574.

[37] C.Johnson,*Functional Characterizations of the Field of Values...*, Proceedings of the American Mathematical Society. **61 No. 2** (1977), pp. 201-204.

[38] C.Johnson,*A Gerschgorin Inclusion Set for the Field of Values of a Matrix*, Proceedings of the American Mathematical Society. **41 No. 1** (1973), pp. 57-60.

[39] R.L.Johnson,*Gerschgorin Theorems for Partitioned Matrices.*, Linear Algebra and Its Applications. **4** (1971), pp. 205-220.

[40] R.L.Johnson and D.D.Olesky,*Best Pseudo-Isolated Gerschgorin Disks for Eigenvalues.*, Linear Algebra and Its Applications. **4** (1971), pp. 205-220.

[41] A.Klawonn and G.Starke,*Block Triangular Preconditioners for nonsymmetric saddle point problems:field of value analysis.*, Numerische Mathematik. **81** (1999), pp. 577-594.

[42] Z.V.Kovarik,*Spectrum Localization in Banach Spaces I* , Linear Algebra and Its Applications. **8** (1974), pp. 225-236.

[43] Z.V.Kovarik,*Spectrum Localization in Banach Spaces II* , Linear Algebra and Its Applications. **12** (1975), pp. 223-229.

[44] Z.V.Kovarik,*Sharpness of Generalized Gerschgorin Disks*, Linear Algebra and Its Applications. **8** (1974), pp. 477-482.

[45] S.Leng,*Characteristic Roots of a Matrix*, Duke Mathematical Journal. **19** (1952), pp. 139-154.

[46] B.W.Levinger and R.S.Varga,*Minimal Gerschgorin Sets II*, Pacific Journal of Mathematics **17 No.2** (1966), pp. 199-210.

[47] B.W.Levinger,*Minimal Gerschgorin Sets III*, Linear Algebra and Its Applications. **2** (1969), pp. 13-19.

[48] C.Li,*A generalization of Spectral Radius, Numerical Radius, and Spectral Norm*, Linear Algebra and Its Applications. **90** (1987), pp.105-118.

[49] C.Li, B.Tam, and P.Wu,*The Numerical Range of a nonnegative Matrix*, Linear Algebra and Its Applications. **350** (2002), pp.1-23.

[50] C.Li and M.Tsatsomeros,*Doubly Diagonally Dominant Matrices*, Linear Algebra and Its Applications. **261** (1997), pp.221-235.

[51] C.Li, C.Sung, and N.Tsing,*c-Convex Matrices:Characterizations,Incl. Rel. and Normality*, Linear and Multilinear Algebra. **25** (1989), pp. 275-287.

[52] A.Lumsdaine and D.Wu,*Spectra and Pseudospectra of Block Toeplitz Matrices*, Linear Algebra and Its Applications. **272** (1998), pp.103-130.

[53] E.A.Maximenko,*Convolution Operators on Expanding Polyhedra: Limits of the Norms of Inverse Operators and Pseudospectra*, Siberian Mathematical Journal **44 No. 6** (2003), pp. 1027-1038.

[54] H.I.Medley,*A Note on G-Generating Families and Isolated Gerschgorin Disks*, Numerische Mathematik **21** (1973), pp. 93-95.

[55] H.I.Medley and R.S.Varga,*On the Smallest Isolated Gerschgorin Disks for Eigenvalues II*, Numerische Mathematik **11** (1968), pp. 320-323.

[56] H.I.Medley and R.S.Varga,*On the Smallest Isolated Gerschgorin Disks for Eigenvalues III*, Numerische Mathematik **11** (1968), pp. 361-369.

[57] G.W.Medlin,*Bounds for the Characteristic Roots of a Matrix with Real Elements*, Duke Mathematical Journal. **19 No. 4** (Dec 1952), pp. 563-565.

[58] C.Meyer,*Matrix Analysis and Applied Linear Algebra* SIAM, Philadelphia, 2000.

[59] C.Moler and C.Van Loan,*Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later*, SIAM Rview. **45** (2003), pp. 3-49.

[60] A.M.Ostrowski,*Note on a Theorem by A.Brauer*, Duke Mathematical Journal. **22 No. 3** (1955), pp. 469-470.

[61] A.M.Ostrowski and H.Schneider, *Bounds for the Maximal Characteristic Root of a Non-negative Irreducible Matrix.*, Duke Mathematical Journal. **27** (1960), pp. 547-553.

[62] W.V.Parker,*Characteristic Roots of a Matrix*, Duke Mathematical Journal. **3** (1937), pp. 484-487.

[63] W.V.Parker,*Characteristic Roots and the Field of Values of a Matrix*, Duke Mathematical Journal.

[64] W.V.Parker,*Sets of Complex Numbers Associated with a Matrix*, Duke Mathematical Journal. **15** (1948), pp. 711-715.

[65] J.M.Pena,*On an alternative to Gerschgorin circles and ovals of Cassini*, Numerische Mathematik **95** (2003), pp. 337-345.

[66] T.A.Porsching,*Diagonal Similarity Transformations which Isolate Gerschgorin Disks*, Numerische Mathematik **8** (1966), pp. 437-443.

[67] T.A.Porsching,*Analytic Eigenvalues and Eigenvectors*, Duke Mathematical Journal. (1968), pp. 363-367.

[68] D.L.Powers,*A Block Gerschgorin Theorem*, Linear Algebra and Its Applications. **13** (1976), pp. 45-52

[69] L.Reichel and L.Trefethen,*Eigenvalues and Pseudo-eigenvalues of Toeplitz Matrices*, Linear Algebra and Its Applications. **162-164** (1992), pp. 153-185.

[70] A.Rube,*On the Closeness of Eigenvalues and Singular Values for Almost Normal Matrices*, Linear Algebra and Its Applications. **11** (1975), pp. 87-94.

[71] P.N.Shivakumar and J.J.Williams,*Eigenvalues for Infinite Matrices*, Linear Algebra and Its Applications. **96** (1987), pp. 35-63

[72] D.S.Scott,*On the Accuracy of the Gerschgorin Circle Theorems for Bounding the Spread of a Real Symmetric Matrix.*, Linear Algebra and Its Applications. **65** (1985), pp. 147-155.

[73] G.Starke,*Field-of-Values Analysis of Preconditioned Iterative Methods for nonsymmetric Elliptic Problems.*, Numerische Mathematik **78** (1997), pp. 103-117.

[74] O.Taussky,*Bounds for Characteristic Roots of Matrices*, Duke Mathematical Journal. **15** (1948), pp. 1043-1044.

[75] O.Taussky,*A Recurring Theorem on Determinants*, American Mathematical Monthly. **56** (1949), pp. 672-676.

[76] O.Taussky,*Bounds for Characteristic Roots of Matrices II*, Journal of Research of the National Bureau of Standards **46 No. 2** (Feb 1951), pp. 124-125.

[77] F.Tisseur and N.J.Higham,*Structured Pseudospectra for Polynomial Eigenvalue Problems, with Applications.*, SIAM J. Matrix Anal. Appl. **23 No. 1** (2001), pp. 187-208.

[78] J. Todd,*On the Smallest Isolated Gerschgorin Disks for Eigenvalues*, Numerische Mathematik **7** (1965), pp. 171-175.

[79] L.N.Trefethen,*Pseudospectra of Linear Operators.*, SIAM **39 No. 3** (1997), pp. 383-406.

[80] L.N.Trefethen, M.Contedini, and M.Embree, *Spectra, Pseudospectra, and Localization for Random Bidiagonal Matrices.*, Communications on Pure and Applied Mathematics, **LIV** (2001), pp. 595-623.

[81] A.E.Trefethen, L.N.Trefethen, and P.J.Schmid, *Spectra and Pseudospectra for Pipe Poiseuille Flow*, Computer Methods in Applied Mechanics and Engineering **1926** (1999), pp. 413-420.

[82] W.Trench,*On the Eigenvalue Problem for Toeplitz Band Matrices*, Linear Algebra and Its Applications. **64** (1985), pp. 199-214.

[83] C.Tretter and M.Wagenhofer,*The Block Numerical Range of an n x n Block Operator Matrix.*, SIAM J. Matrix Anal. Appl. **24 No. 4** (2003), pp. 1003-1017.

[84] J.M.Varah,*A Lower Bound for the Smallest Singular Value of a Matrix*, Linear Algebra and Its Applications. **11** (1975), pp. 3-5.

[85] R.S.Varga,*Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, NJ 1962.

[86] R.S.Varga,*On the Smallest Isolated Gerschgorin Disks for Eigenvalues*, Numerische Mathematik **6** (1964), pp. 367-376.

[87] R.S.Varga,*Minimal Gerschgorin Sets*, Pacific Journal of Mathematics **15 No. 2** (1965), pp. 719-729.

[88] R.S.Varga,*Gerschgorin-Type Eigenvalue Inclusion Theorems and Their Sharpness*, Electronic Transactions on Numerical Analysis **12** (2001), pp. 113-133.

[89] R.S.Varga and A.Krautstengl,*Minimal Gerschgorin Sets for Partitioned Matrices III.*, Electronic Transactions on Numerical Analysis **3** (1995), pp. 83-95.

[90] R.S.Varga and A.Krautstengl,*On Gerschgorin-Type Problems and Ovals of Cassini.*, Electronic Transactions on Numerical Analysis **8** (1999), pp. 15-20.

[91] R.S.Varga and B.Levinger,*On Minimal Gerschgorin Sets for Families of Norms*, Numerische Mathematik **20** (1973), pp. 252-256.

[92] T.G.Wright and L.N.Trefethen,*Large Scale Computation of Pseudospectra using ARPAK and EIGS.*, SIAM J. Sci. Comput. **23 No. 2** (2001), pp. 591-605.

[93] T.G.Wright and L.N.Trefethen,*Pseudospectra of Rectangular Matrices.*, IMA Journal of Numerical Analysis **22** (2002), pp. 501-519.

A Detailed calculations of some of the examples

This appendix contains detailed calculations of some of the examples given in the text.

Example 1.2

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & -1 & 2 \\ 1 & 2 & 1 & -1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & .5 & -1 \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{-1.79, 2.17, .81 + 341i, .81 - 341i\}.$$

(Note that the eigenvalues in this thesis were computer with Matlab in double precision).

$$\mathbf{B} = \frac{A + A^*}{2} = \begin{pmatrix} 0 & .5 & -.5 & 1.5 \\ .5 & 2 & .5 & 0 \\ -.5 & .5 & 1 & .75 \\ 1.5 & 0 & .75 & -1 \end{pmatrix},$$

$$\mathbf{C} = \frac{A - A^*}{2i} = \begin{pmatrix} 0 & .5i & .5i & -.5i \\ -.5i & 0 & -.5i & i \\ -.5i & .5i & 0 & -.25i \\ .5i & -i & .25i & 0 \end{pmatrix},$$

$$R_{(A)1} = |0| + |0| + |-1| + |2| = 3,$$
$$R_{(B)1} = |0| + |.5| + |-.5| + |1.5| = 2.5,$$
$$R_{(C)1} = |0| + |.5i| + |.5i| + |-.5i| = 1.5,$$

$$T_{(A)1} = |0| + |1| + |0| + |1| = 2,$$

$$R_{(A)2} = |1| + |2| + |1| + |-1| = 5,$$
$$R_{(B)2} = |.5| + |2| + |.5| + |0| = 3,$$
$$R_{(C)2} = |-.5i| + |0| + |-.5i| + |i| = 2,$$

$$T_{(A)2} = |0| + |2| + |0| + |1| = 3,$$

$$R_{(A)3} = |0| + |0| + |1| + |1| = 2,$$

$$R_{(B)3} = |-.5| + |.5| + |1| + |.75| = 2.75,$$
$$R_{(C)3} = |-.5i| + |.5i| + |0| + |-.25i| = 1.25,$$

$$T_{(A)3} = |-1| + |1| + |1| + |.5| = 3.5,$$

$$R_{(A)4} = |1| + |1| + |.5| + |-1| = 3.5,$$
$$R_{(B)4} = |1.5| + |0| + |.75| + |-1| = 3.25,$$
$$R_{(C)4} = |5i| + |-i| + |.25i| + |0| = 1.75,$$

and

$$T_{(A)4} = |2| + |-1| + |1| + |-1| = 5.$$

From these, the maximums may be calculated,

$$R_A = max\{R_{(A)i}, ..., R_{(A)N}\} = max\{3, 5, 2, 3.5\} = 5,$$
$$R_B = max\{R_{(B)i}, ..., R_{(B)N}\} = max\{2.5, 3, 2.75, 3.25\} = 3.25,$$
$$R_C = max\{R_{(C)i}, ..., R_{(C)N}\} = max\{1.5, 2, 1.25, 1.75\} = 2,$$
$$T = max\{T_i, ..., T_N\} = max\{2, 3, 3.5, 5\} = 5.$$

and for any $\lambda \in \sigma(A)$,

$$|\lambda| \leq \frac{R_A + T}{2} = \frac{5+5}{2} = 5, \qquad |Re\,\lambda| \leq R_B = 3.25, \quad and \quad |Im\,\lambda| \leq R_C = 2.$$

This is graphed in figure 1.2 below. Note that the bounds on 'a' and 'b' (represented by the rectangle in figure 1.2) are contained completely within the bounds for λ (represented by the circle in figure 1.2). Therefore, in this example, the bound on λ (represented by the circle) is superfluous and may be ignored.

Example 1.7

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & -1 & 2 \\ 1 & 2 & 1 & -1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & .5 & -1 \end{pmatrix}.$$

The spectrum, $\sigma(A)$, was found to be

$$\{-1.79, 2.17, .81 + 341i, .81 - 341i\}.$$

$$\mathbf{B} = \frac{A + A^*}{2} = \begin{pmatrix} 0 & .5 & -.5 & 1.5 \\ .5 & 2 & .5 & 0 \\ -.5 & .5 & 1 & .75 \\ 1.5 & 0 & .75 & -1 \end{pmatrix},$$

$$\mathbf{C} = \frac{A - A^*}{2i} = \begin{pmatrix} 0 & .5i & .5i & -.5i \\ -.5i & 0 & -.5i & i \\ -.5i & .5i & 0 & -.25i \\ .5i & -i & .25i & 0 \end{pmatrix},$$

$$S_{(A)1} = (|0| + |0| + |-1| + |2| + |0| + |1| + |0| + |1|)/2 = 2.5,$$
$$S_{(B)1} = (|0| + |.5| + |-.5| + |1.5| + |0| + |.5| + |-.5| + |1.5|)/2 = 2.5,$$
$$S_{(C)1} = (|0| + |.5i| + |.5i| + |-.5i| + |0| + |-.5i| + |-.5i| + |.5i|)/2 = 1.5,$$

$$S_{(A)2} = (|1| + |2| + |1| + |-1| + |0| + |2| + |0| + |1|)/2 = 4,$$
$$S_{(B)2} = (|.5| + |2| + |.5| + |0| + |.5| + |2| + |.5| + |0|)/2 = 3,$$
$$S_{(C)2} = (|-.5i| + |0| + |-.5i| + |i| + |.5i| + |0| + |.5i| + |-i|)/2 = 2,$$

$$S_{(A)3} = (|0| + |0| + |1| + |1| + |-1| + |1| + |1| + |.5|)/2 = 2.75,$$
$$S_{(B)3} = (|-.5| + |.5| + |1| + |.75| + |-.5| + |.5| + |1| + |.75|)/2 = 2.75,$$
$$S_{(C)3} = (|-.5i| + |.5i| + |0| + |-.25i| + |.5i| + |-.5i| + |0| + |.25i|)/2 = 1.25,$$

$$S_{(A)4} = (|1| + |1| + |.5| + |-1| + |2| + |-1| + |1| + |-1|)/2 = 4.25,$$
$$S_{(B)4} = (|1.5| + |0| + |.75| + |-1| + |1.5| + |0| + |.75| + |-1|)/2 = 3.25,$$
$$S_{(C)4} = (|.5i| + |-i| + |.25i| + |0| + |-.5i| + |i| + |-.25i| + |0|)/2 = 1.75,$$

$$S_A = max\{2.5, 4, 2.75, 4.25\} = 4.25,$$
$$S_B = max\{2.5, 3, 2.75, 3.25\} = 3.25,$$
$$S_C = max\{1.5, 2, 1.25, 1.75\} = 2,$$

and

$$S_A = 4.25, \qquad S_B = 3.25, \qquad S_C = 2,$$
$$|\lambda| \le S_A = 4.25\,, \qquad |Re\ \lambda| \le S_B = 3.25\,, \quad \textit{and} \quad |Im\ \lambda| \le S_C = 2.$$

Note that the bound on $|\lambda|$ is 4.25 for Parker, compared with 5 for Browne. However, in this example, the bounds on 'a' and 'b' (represented by the rectangle in figure 1.7) are contained completely within the bounds for λ (represented by the rectangle in figure 1.7). Therefore, in this example, the bound on λ is superfluous so that, Browne and Parker, for practical purposes, produce the same results.

Example 1.28 Find the Gerschgorin Radii and Disks for the following matrix.

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & -1 & 2 \\ 1 & 2 & 1 & -1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & .5 & -1 \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is
$\{-1.79, 2.17, .81 + 341i, .81 - 341i\}$.

Solution The first Gerschgorin Circle is centered at $a_{11} = 0$ with radius:

$$r_1 = \sum_{\substack{j=1 \\ j \neq 1}}^{4} |a_{1j}| = |a_{12}| + |a_{13}| + |a_{14}| = |0| + |-1| + |2| = 3.$$

The second Gerschgorin Circle is centered at $a_{22} = 2$ with radius:

$$r_2 = \sum_{\substack{j=1 \\ j \neq 2}}^{4} |a_{2j}| = |a_{21}| + |a_{23}| + |a_{24}| = |1| + |1| + |-1| = 3.$$

The third Gerschgorin Circle is centered at $a_{33} = 1$ with radius:

$$r_3 = \sum_{\substack{j=1 \\ j \neq 3}}^{4} |a_{3j}| = |a_{31}| + |a_{32}| + |a_{34}| = |0| + |0| + |1| = 1.$$

The fourth Gerschgorin Circle is centered at $a_{44} = -1$ with radius:

$$r_4 = \sum_{\substack{j=1 \\ j \neq 4}}^{4} |a_{4j}| = |a_{41}| + |a_{42}| + |a_{43}| = |1| + |1| + |.5| = 2.5.$$

The Gerschgorin disks for this matrix are plotted in figure 1.28.
• • ••

Example 1.29 Find the Gerschgorin Radii and Disks for the following.

$$\mathbf{A} = \begin{pmatrix} 2+3i & i & 4 & i+1 \\ 4-4i & 2 & 1+i & 2+2i \\ 3i & 4 & -4i & 5i \\ -7 & 2-5i & 6 & -5+i \end{pmatrix}.$$

The spectrum, $\sigma(A)$, is

$$\{7.79 - .12i, .97 + 6.19i, -5.47 - 5.89i, -4.29 - .18i\}.$$

Solution The first Gerschgorin Circle is centered at $a_{11} = 2 + 3i$ with radius:

$$r_1 = \sum_{\substack{j=1 \\ j\neq 1}}^{4} |a_{1j}| = |a_{12}| + |a_{13}| + |a_{14}| = |i| + |4| + |i + 1| = 1 + 4 + \sqrt{2} = 6.4142.$$

The second Gerschgorin Circle is centered at $a_{22} = 2$ with radius:

$$r_2 = \sum_{\substack{j=1 \\ j\neq 2}}^{4} |a_{2j}| = |a_{21}|+|a_{23}|+|a_{24}| = |4-4i|+|1+i|+|2+2i| = \sqrt{32}+\sqrt{2}+\sqrt{8} = 9.9.$$

The third Gerschgorin Circle is centered at $a_{33} = -4i$ with radius:

$$r_3 = \sum_{\substack{j=1 \\ j\neq 3}}^{4} |a_{3j}| = |a_{31}| + |a_{32}| + |a_{34}| = |3i| + |4| + |5i| = 3 + 4 + 5 = 12.$$

The fourth Gerschgorin Circle is centered at $a_{44} = -5 + i$ with radius:

$$r_4 = \sum_{\substack{j=1 \\ j\neq 4}}^{4} |a_{4j}| = |a_{41}|+|a_{42}|+|a_{43}| = |-7|+|2-5i|+|6| = 7+\sqrt{29}+6 = 18.39.$$

The Gerschgorin disks for this matrix are plotted in figure 1.29.

• • • •

Example 4.10 Consider,

$$\mathbf{A} = \begin{pmatrix} 5 & 3 & 0 & 0 & 0 & -7 \\ 5 & 9 & 0 & 6 & 0 & -3 \\ 0 & 0 & 4 & 9 & 8 & 0 \\ 0 & 0 & -6 & 8 & 7 & 0 \\ 0 & 0 & 8 & 7 & 9 & 0 \\ -41 & 7 & 0 & 0 & 0 & 14 \end{pmatrix}$$

Construct the associated graph as follows:

$a_{12} \neq 0$ place a directed edge from 1 to 2.
$a_{13} = 0$ do not place a directed edge from 1 to 3.
$a_{14} = 0$ do not place a directed edge from 1 to 4.
$a_{15} = 0$ do not place a directed edge from 1 to 5.
$a_{16} \neq 0$ place a directed edge from 1 to 6.
$a_{21} \neq 0$ place a directed edge from 2 to 1.
$a_{23} = 0$ do not place a directed edge from 2 to 3.
$a_{24} \neq 0$ place a directed edge from 2 to 4.

$a_{25} = 0$ do not place a directed edge from 2 to 5.
$a_{26} \neq 0$ place a directed edge from 2 to 6.
$a_{31} = 0$ do not place a directed edge from 3 to 1.
$a_{32} = 0$ do not place a directed edge from 3 to 2.
$a_{34} \neq 0$ place a directed edge from 3 to 4.
$a_{35} \neq 0$ place a directed edge from 3 to 5.
$a_{36} = 0$ do not place a directed edge from 3 to 6.
$a_{41} = 0$ do not place a directed edge from 4 to 1.
$a_{42} = 0$ do not place a directed edge from 4 to 2.
$a_{43} \neq 0$ place a directed edge from 4 to 3.
$a_{45} \neq 0$ place a directed edge from 4 to 5.
$a_{46} = 0$ do not place a directed edge from 4 to 6.
$a_{51} = 0$ do not place a directed edge from 5 to 1.
$a_{52} = 0$ do not place a directed edge from 5 to 2.
$a_{53} \neq 0$ place a directed edge from 5 to 3.
$a_{54} \neq 0$ place a directed edge from 5 to 4.
$a_{56} = 0$ do not place a directed edge from 5 to 6.
$a_{61} \neq 0$ place a directed edge from 6 to 1.
$a_{62} \neq 0$ place a directed edge from 6 to 2.
$a_{63} = 0$ do not place a directed edge from 6 to 3.
$a_{64} = 0$ do not place a directed edge from 6 to 4.
$a_{65} \neq 0$ place a directed edge from 6 to 5.

The result is the directed graph shown in the figure 4.10. Notice that this graph ***is not strongly connected.*** For example, there does not exist a directed path from 5 to 2. Since the graph is not strongly connected, it is not irreducible. On the other hand, each vertex belongs to some cycle in the graph. That is, 1 belongs to C_{126}; 2 belongs to C_{126}; 3 belongs to C_{345};4 belongs to C_{345};5 belongs to C_{35}; and 6 belongs to C_{126}. Therefore, the matrix is weakly irreducible and the Brualdi Theorem may be applied to this matrix.

Example 10.4 Consider the 50 x 50 matrix:

$$\mathbf{A} = \begin{pmatrix} 6 & 5 & 0 & 0 & 0 & \dots \\ 7 & 6 & 5 & 0 & 0 & \dots \\ 8 & 7 & 6 & 5 & 0 & \dots \\ 0 & 8 & 7 & 6 & 5 & \dots \\ 0 & 0 & 8 & 7 & 6 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

The results are shown in figure 10.4. The Minimal Gerschgorin disk has a radius of 15.316 with center at (6,0). So, the Minimal Gerschgorin disk crosses the x-axis at 21.316. On the other hand, the largest real eigenvalue is 21.3144. That is very good.

Example 10.5

Same Toeplitz matrix as example 10.4 but with dimension 100 x 100.

The results are shown in figure 10.5A. This time there appears to be a serious problem - some of the eigenvalues fall *outside of* the Inclusion Set. This may lead one to conclude that there is some roundoff error or some other problem with the algorithm but this is not the case: Bottcher and Silbermann[] note in their book Introduction to Large Truncated Toeplitz Matrices (Example 3.16 page 72) that this kind of roundoff error often occurs when calculating the eigenvalues of large Toeplitz matrices. But is the phenomena that Bottcher and Silbermann refer to happening in *this* case? Apparently so. For when the eigenvalues of the transpose of the matrix are calculated, the picture changes (Figure 10.5B). Now things are back on track - the Spectrum is inside the Inclusion Set. The Minimal Gerschgorin disk has a radius of 15.348 with center at (6,0). So, the Minimal Gerschgorin disk crosses the x-axis at 21.348. On the other hand, the largest real eigenvalue is 21.3144. Again, that is very good.

It is not at all clear why Matlab was able to calculate the eigenvalues correctly for the transpose but not for the original matrix. Of course, that is typical of roundoff error - it is often sensitive to the order in which the calculations are performed. In any case, the next few examples will calculate eigenvalues based on the transpose of our Toeplitz matrix.

Example 10.6

Same Toeplitz matrix as 10.4 but with dimension 300 x 300.

The results are shown in figure 10.6. The Minimal Gerschgorin disk has a radius of 15.358 with center at (6,0). So, the Minimal Gerschgorin disk crosses the x-axis at 21.358. On the other hand, the largest real eigenvalue is 21.3561. With this size of matrix it is still possible to calculate the Numerical Range but, the execution time is very long.

Example 10.7

Same Toeplitz matrix as 10.4 but with dimension 1,000 x 1,000.

The results are shown in figure 10.7. The Minimal Gerschgorin disk has a radius of 15.358 with center at (6,0). So, the Minimal Gerschgorin disk crosses the x-axis at 21.358. On the other hand, the largest real eigenvalue is 21.3573. (The computer being used for this thesis is not powerful enough to do the Numerical Range algorithm for a 1000x1000 matrix.)

Example 10.8

Same Toeplitz matrix as example 10.4 but with dimension 10,000 x 10,000.

The results are shown in figure 10.8. The Minimal Gerschgorin disk has a radius of 15.358 with center at (6,0). So, the Minimal Gerschgorin disk crosses the x-axis at 21.358. It is not possible to calculate the eigenvalues for a 10,000 x 10,000 matrix on the computer being used for this thesis. (It is difficult to even *create* such a large matrix on the computer!)

Example 10.9

Same Toeplitz matrix as 10.4 but with dimension 100,000 x 100,000.

The results are the same as example 10.8. (See figure 10.8).

Example 10.10

Same Toeplitz matrix as 10.4 but with dimension 1,000,000 x 1,000,000.

The results are the same as example 10.8. (See figure 10.8).

B Methods Used in this Thesis

Simpler Methods

Browne's Theorem	Theorem 1.1
Parker's First	Theorem 1.6
Farnell's First	Theorem 1.11
Farnell's Second	Theorem 1.12
Brauer's First	Theorem 1.13
Gerschgorin's Theorem	Theorem 1.27
Gerschgorin's Column Theorem	Theorem 1.30
Parker's Second Theorem (1948)	Theorem 2.2

Involved Methods

Brauer Ovals of Cassini	Theorem 4.1
Cassini for Real Matrices	Theorem 4.7
Brualdi	Theorem 4.10
Minimal Gerschgorin	Theorem 4.14
Numerical Range	Theorem 5.1
Pseudospectra	Chapter 7

Composite BBFP is a composite of the first five 'Simple' methods listed above: Browne, Parker's First, Farnell's First, Farnell's Second, and Brauer's First.

Related Methods

Varga-Medley Methods	Chapter 3

C Symbols used in this thesis

$Br(A)$ - Brualdi Set
$K(A)$ - Cassini Set
$K'(A)$ - Cassini Set for real matrices
$G(A)$ - Gerschgorin set
$G^R(A)$ - Minimal Gerschgorin set
$G^T(A)$ - Gerschgorin *column* set. This is the set produced by applying Gerschgorin's theorem to the columns (rather than the rows of a matrix).
$W(A)$ - Numerical Range
λ - an eigenvalue
ρ - The spectral radius
$\sigma(A)$ - Spectrum of A
$\sigma_\varepsilon(A)$ - Pseudospectra of A
ω - The numerical radius. This is analogous to the spectral radius except that it is applied to the numerical range. That is, it is the largest absolute value of any point in the numerical range.

D Matlab Code for programs used in this thesis

A number of Matlab programs were written for this thesis. The Matlab code for some of the more important programs is presented in this appendix.

The Matlab code for these and other programs may also be found on the following websites:

www.baymite.com/TronzoThesis.htm

All of these programs must be run under Matlab. These programs all follow the same format: the user is prompted for the matrix dimension and then for the variable name that represents the matrix (the matrix must be in the Matlab workspace at the time the program is run).

These programs were not written for general distribution. Therefore, the code in these programs does not include 'error trapping' or other safeguards to ensure that the information entered is suitable for the programs. This means that erroneous results may produced without warning if information is not entered according the the instructions for the programs. For this reason, these programs should be run only by those who have some understanding of the methods that are used.

It should be noted that these programs were written over a long period of time. Therefore, the code and the graphics may appear to 'cleaner' in some programs than in others. In some cases the graphic produced by these programs looks much better than the graphics in the text. This is because some of the graphs in the text were produced by early versions of these programs.

Section D.1 Programs that intersect different spectral inclusion sets

CassiniBrowne This program produces inclusion sets by intersecting the following three sets: Brauer-Cassini (Theorem 4.1), Brauer-Cassini for the transpose, and Browne's box. This program produces a very sharp spectral inclusion set in a reasonable amount of time.

Note that the 'sharpness' of the inclusion set that is produced can be adjusted by changing the value of the variable 'TINC'. At the present time, this variable is set at 201. This appears to be optimal for most matrices. If this value is increased, the program will produce a sharper set but the calculation time will also be increased. If the number assigned to TINC is decreased, the set produced will be courser but the calculation time will be decreased. (The number assignment to TINC occurs in the first few lines of the program).

Example: Find the CassiniBrowne set for the following matrix:

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & -1 & 2 \\ 1 & 2 & 1 & -1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & .5 & -1 \end{pmatrix}.$$

First create the matrix in the Matlab workspace by entering the following:

$A = [0 \quad 0 \quad -1 \quad 2; 1 \quad 2 \quad 1 \quad -1; 0 \quad 0 \quad 1 \quad 1; 1 \quad 1 \quad .5 \quad -1]$

Start the program CassiniBrowne and answer the prompts as follows:

Enter the dimension of the matrix **4**

Enter the variable name that represents the matrix **A**

The computer will then calculate and plot the inclusion set.

IntersectAll1 This program produces inclusion sets by intersecting the following 'simply generated' sets: Browne's (Theorem 1.1), Brauer's First (Theorem 1.13), Farnell's First (Theorem 1.11), Farnell's Second (Theorem 1.12), Parker's First (Theorem 1.6), Parker's Second (1948) (Theorem 2.2), Gerschgorin (Theorem 1.27), and Gerschgorin's column (Theorem 1.3). This program produces a set that is slightly less sharp than the set produced by the 'CassiniBrowne' program (see above) but calculates the set somewhat more quickly than the 'CassiniBrowne' program.

Note that the 'sharpness' of the inclusion set that is produced can be adjusted by changing the value of the variable 'TINC'. At the present time, this variable is set at 201. This appears to be optimal for most matrices. If this value is increased, the program will produce a sharper set but the calculation time will also be increased. If the number assigned to TINC is decreased, the set produced will be courser but the calculation time will be decreased. (The number assignment to TINC occurs in the first few lines of the program).

Example: This program may be run in a similar way as CassiniBrowne. See under 'CassiniBrowne' above for an example.

CompositeBBFP (See Definition 9A) This program produces inclusion sets by intersecting the following 'Pre-Gerschgorin' sets: Browne's (Theorem 1.1), Brauer's First (Theorem 1.13), Farnell's First (Theorem 1.11), Farnell's Second (Theorem 1.12), and Parker's First (Theorem 1.6). The main purpose of this program is to exhibit the best of the 'Pre-Gerschgorin' methods. This program produces a set that is not, in general, nearly as sharp as the sets produced by the 'CassiniBrowne' or 'IntersectAll1' programs (see above) but the calculation time is very short.

Section D.2 Programs that produce Gerschgorin and Gerschgorin-type inclusion sets

GerschgorinBasic This program calculates and plots the Gerschgorin disks (Theorem 1.27) for a matrix.

Example: This program may be run in a similar way as CassiniBrowne. See under 'CassiniBrowne' above for an example.

GerschgorinAdvanced1 This program calculates and plots the Gerschgorin disks (Theorem 1.27) for matrices except that it *will not* plot disks that are completely contained within other disks. That is, this program will not plot superfluous Gerschgorin disks.

Example: This program may be run in a similar way as CassiniBrowne. See under 'CassiniBrowne' above for an example.

GerschgorinExtra1 This program calculates the Gerschgorin disks (Theorem 1.27) for a matrix but plots only the *boundary* of the resulting Gerschgorin inclusion set.

Example: This program may be run in a similar way as CassiniBrowne. See under 'CassiniBrowne' above for an example.

CassiniNewOpt1 This program produces inclusion sets based on the Brauer-Cassini Theorem (Theorem 4.1).

Example: This program may be run in a similar way as CassiniBrowne. See under 'CassiniBrowne' above for an example.

CassiniRealNew1 This program produces inclusion sets based on the Brauer-Medlin-Cassini Theorem (Theorem 4.7) but this program works for *real matrices only.*

Example: Find the Cassini set for the following real matrix:

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & -1 & 2 \\ 1 & 2 & 1 & -1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & .5 & -1 \end{pmatrix}.$$

First create the matrix in the Matlab workspace by entering the following:

$A = [0 \quad 0 \quad -1 \quad 2; 1 \quad 2 \quad 1 \quad -1; 0 \quad 0 \quad 1 \quad 1; 1 \quad 1 \quad .5 \quad -1]$

Start the program CassiniRealNew1 and answer the prompts as follows:

Enter the dimension of the REAL matrix **4**

Enter the variable name that represents the matrix **A**

The computer will then calculate and plot the inclusion set.

BrualdiNew1 This is very experimental program that works only for 3 x 3 weakly irreducible matrices. This program calculates the Brualdi Lemniscate set (Theorem 4.10). This program will calculate all 3 x 3 cycles plus it will calculate only those 2 x 2 cycles that are on a circuit. This program produces inclusion sets that are not quite as sharp as a true Brualdi calculation.

Example: This program may be run in a similar way as CassiniBrowne. See under 'CassiniBrowne' above for an example.

BrualdiSpecial2 This is very experimental program that works with weakly irreducible matrices as large as 4 x 4. This program checks to see that the matrix is weakly irreducible. This program calculates the Brualdi Lemniscate set (Theorem 4.10).

Example: This program may be run in a similar way as CassiniBrowne. See under 'CassiniBrowne' above for an example.

GerschMinCombined This is an experimental program. This program calculates the minimal Gerschgorin set (Theorem 4.14). **This program is extremely slow.** Therefore, it will be wise to try this program with a 3x3 or a 4x4 matrix in order to get an indication of the calculation required *before* attempting to run this program with a large matrix.

Note that the 'sharpness' of the inclusion set that is produced can be adjusted by changing the value of the variables 'MTRS' and 'VXINCR'. At the present time, MTRS is set to 51 and VXINCR is set to .025. One may want to experiment with different values for these variables. If the value of MTRS is increased AND/OR the value of VINCR is decreased, the program will produce a sharper set but the calculation time be increased. If the value of MTRS is decreased AND/OR the value of VINCR is increased, the set produced will be courser but the calculation time will be decreased.

Example: This program may be run in a similar way as CassiniBrowne. See under 'CassiniBrowne' above for an example.

Section D.3 Programs that produce Pre-Gerschgorin inclusion sets

Brown1 This program produces spectral inclusion sets based on Browne's theorem (Theorem 1.1).

BrauerPower This program produces spectral inclusion sets based on Brauer's Power Method (Theorem 1.14). The Brauer Power Method creates a spectral inclusion set utilizing the original matrix raised to a power. That power is always in the form of 2^r where 'r' is a natural number. Therefore, when r=1, the matrix is raised to the 2nd power; when r=2, the matrix is raised to the 4th power; when r=3, the matrix is raised to the 8th power, etc. So, when this program is run, the user will be prompted to enter 'r'. As noted in the text, the Brauer Power Method is considered unreliable. Therefore, **it is possible that the set created by the Brauer Power Method may not include the spectrum!**

FarnellOriginal1 This program produces spectral inclusion sets based on Farnell's first theorem (Theorem 1.11).

FarnellSecond This program produces spectral inclusion sets based on Farnell's second theorem (Theorem 1.12).

Parker1 This program produces spectral inclusion sets based on Parker's first theorem (Theorem 1.6).

ParkerSecondNew This program produces spectral inclusion sets based on Parker's second (1948) theorem (Theorem 2.2).

Section D.4 Programs that utilize Varga-Medley methods

IsolateGersch3 This program calculates exact eigenvalues that are contained inside isolated Gerschgorin disks. This program utilizes theorems developed by Olga Taussky (Theorem 3.1 and Corollary 3.3) and algorithms developed by Helen Medley and Richard Varga (see chapter three for the full discussion of this subject).

Section D.5 Programs that utilize new methods

GerschgorinHessenberg This program produces a minimal Gerschgrin spectral inclusion set for Hessenberg matrices. The method used in this program was developed by Mark Tronzo for this thesis (See theorems 10.1 and 10.2 and Algorithms 10.1 and 10.2).

(Unlike the other programs, the matrix need not be created in the workspace).

Example: Find the minimal Gerschgorin set for the following Hessenberg matrix:

$$\mathbf{A} = \begin{pmatrix} 5 & 18-3i & 0 & 0 & 0 & 0 & 0 & \dots \\ 4i & 5 & 18-3i & 0 & 0 & 0 & 0 & \dots \\ 0 & 4i & 5 & 18-3i & 0 & 0 & 0 & \dots \\ -9 & 0 & 4i & 5 & 18-3i & 0 & 0 & \dots \\ 0 & -9 & 0 & 4i & 5 & 18-3i & 0 & \dots \\ 0 & 0 & -9 & 0 & 4i & 5 & 18-3i & \dots \\ 6+8i & 0 & 0 & -9 & 0 & 4i & 5 & \dots \\ 0 & 6+8i & 0 & 0 & -9 & 0 & 4i & \dots \\ 0 & 0 & 6+8i & 0 & 0 & -9 & 0 & \dots \\ 0 & 0 & 0 & 6+8i & 0 & 0 & -9 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

Start the program GerschgorinHessenberg and answer the prompts as follows:

Enter the dimension of the matrix **200**

How many bands are there below the center diagonal **6**

Enter an element from the leftmost band **6+8i**

Enter an element from the next band **0**

Enter an element from the next band **0**

Enter an element from the next band **-9**

Enter an element from the next band **0**

Enter an element from the next band **4i**

Enter an element from the center diagonal **5**

Enter an element from the band above the center diagonal **18-3i**

Section D.6 Programs written by others

NumericalRangeAlg1 This program calculates the numerical range for a matrix this program was written by Carl C. Cowen (Purdue University) and Elad Harel.

eigtool This program calculates the pseudospectra for a matrix. The Matlab code for this program was written by Tom Wright. The Matlab code for this program is not given in this appendix but may be found at the website http://web.comlab.ox.ac.uk/projects/pseudospectra.