Gershgorin's Circle Theorem for Estimating the Eigenvalues of a Matrix with Known Error Bounds

Author:
David Marquis

Advisors:
Professor Hans De Moor
Dr. Kathryn Porter

Reader:
Dr. Michael Nathanson

May 15, 2016

Figure 1: Semyon Aranovich Gershgorin

Abstract

Square matrices are characterized by their eigenvalues, so we would like to be able to produce the eigenvalues for any square matrix. Cofactor expansion is an algorithm used to produce what is known as the characteristic polynomial of a matrix A, the roots of which are the eigenvalues of A. However, the algorithm for cofactor expansion is computationally intensive and as the output is an n^{th} degree polynomial, we do not have any guaranteed method to find the roots for large matrices.
Semyon Aranovich Gershgorin (1901-1933), was a mathematician who worked mainly in the study of differential equations. Among his accomplishments were a degree from the Petrograd Technological Institute, a professorship at the Leningrad Institute of Mechanical Engineering, and he became head of the Division of Mechanics at the Turbine Institute. In 1931, he authored a paper titled *Über die Abgrenzung der Eigenwerte einer Matrix* in which he gave the proof of what has come to be called the Gershgorin Circle Theorem with provided a simple method to estimate the eigenvalues of a complex matrix with known error bounding.

Introduction

In linear algebra, linear systems of equations in the form of matrices are studied.

Definition: A <u>matrix</u> is an $n \times m$ array of elements from a vector space. In this essay, we will only consider matrices with entries from the complex numbers.

An example of a matrix is

$$\begin{bmatrix} i & 0 & 87 \\ -12 & -12 & 3+2i \\ 8i & 1 & 0.11 \\ 54 & -45 & 3-i \end{bmatrix}$$

Matrices can easily represent a system of n equations in m variables. A matrix that has the same number of rows and columns, is called a square matrix. One way we can study such a matrix is to find its eigenvalues and eigenvectors.

Definition: An eigenvector is a column vector which when multiplied on the left by a matrix results in a scalar multiple of the original vector or $A\mathbf{x} = \lambda\mathbf{x}$. Although the equation $A\vec{0} = \lambda\vec{0}$ is true for all A and λ, we define $\vec{0}$ to not be an eigenvector for any matrix.

Definition: The value λ is what is known as the eigenvalue associated with the eigenvector $\mathbf{x}$.
The matrix

$$A = \begin{bmatrix} 3 & -2 \\ -1 & 2 \end{bmatrix}$$

has eigenvalues $\lambda_1 = 4$ and $\lambda_2 = 1$. The corresponding eigenvectors are $\mathbf{x}_1 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and all scalar multiples of either. This is because

$$\begin{bmatrix} 3 & -2 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} -2 \\ 1 \end{bmatrix} = \begin{bmatrix} -8 \\ 4 \end{bmatrix} = 4\begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

and

$$\begin{bmatrix} 3 & -2 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Thus $A\mathbf{x}_1 = \lambda_1\mathbf{x}_1$ and $A\mathbf{x}_2 = \lambda_2\mathbf{x}_2$.

Eigenvalues and eigenvectors characterize a matrix. Another related value associated with a matrix is its determinant. Although for this paper, we won't need to examine the determinants of any matrices, we will need to know the algorithm for finding determinants as it leads us to an algorithm for finding eigenvalues.

The general method for finding the determinant of a matrix is called cofactor expansion. For a matrix $R = [r_{i,j}]$, the minor $M_{k,l}$ of the entry $r_{k,l}$ is the determinant of the matrix that is left after removing the k^{th} row and l^{th} column from R. The cofactor $C_{k,l}$ of $r_{k,l}$ is then $(-1)^{k+l} M_{k,l}$.

If the cofactor produced then requires one to evaluate the determinant of a matrix one dimension smaller than the previous matrix, we then perform cofactor expansion of that minor. We continue this process until all of the determinants are of 2×2 matrices. The formula to evaluate the determinant of a 2×2 matrix is

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$

By selecting a row or column of R and for that row or column, summing the product of each entry with its cofactor, we find the determinant of R.

The equation $A\mathbf{x} = \lambda\mathbf{x}$ is equivalent to the equation $A\mathbf{x} = \lambda I\mathbf{x}$ where I, the identity matrix, is the matrix with the same dimension as A whose entries are 1's along the main diagonal and 0 elsewhere. Then, we can subtract $A\mathbf{x}$ from both sides and factor $\mathbf{x}$ from both terms to get

$$(\lambda I - A)\mathbf{x} = 0$$

Because $\mathbf{x} \neq \vec{0}$, $(\lambda I - A)$ must be singular, so

$$\det(\lambda I - A) = 0$$

Solving for which values of λ satisfy the above equation produces the eigenvalues of A.

The eigenvalues in certain matrices are very easy to determine.
Definition: An upper triangular matrix is a square matrix with all entries below the main diagonal equal to zero.

To illustrate, if you replace each ? with any complex number, then the following matrix will be upper triangular

$$\begin{bmatrix} ? & ? & ? & ? & ? \\ 0 & ? & ? & ? & ? \\ 0 & 0 & ? & ? & ? \\ 0 & 0 & 0 & ? & ? \\ 0 & 0 & 0 & 0 & ? \end{bmatrix}$$

Theorem: The eigenvalues of a triangular matrix (upper, lower or diagonal) are exactly the entries on the main diagonal.

This is because, for a matrix $[a_{i,j}]$ if you imagine performing cofactor expansion on the first column of $(\lambda I - A)$, there will only be one non-zero term to sum and it will be $(\lambda - a_{1,1})$ times the determinant of the minor. The determinant of that minor is then $(\lambda - a_{2,2})$ times the determinant of its minor, et cetera until we get the factored characteristic polynomial $(\lambda - a_{1,1})(\lambda - a_{2,2})...(\lambda - a_{n,n}) = 0$. Thus the eigenvalues of a triangular matrix are the values along its main diagonal.

Example 1: Determine the eigenvalues of the matrix

$$\begin{bmatrix} 2 & 3+i & 8 \\ 0 & 0 & 0.5 \\ 0 & 0 & 9 \end{bmatrix}$$

Because this is an upper triangular matrix, we can easily tell that the eigenvalues are 2, 0, and 9.

Example 2: Determine the eigenvalues of the matrix

$$\begin{bmatrix} 10 & 2 & 3 \\ 0 & 11 & 1 \\ 0 & -1 & 13 \end{bmatrix}$$

It is incorrect to claim that the eigenvalues are 10, 11, & 13 because this matrix has non-zero entries both above and below the main diagonal, that claim would be false. We have to find the solutions to the characteristic equation $\det(\lambda I - A) = 0$ or

$$\begin{vmatrix} \lambda - 10 & -2 & -3 \\ 0 & \lambda - 11 & -1 \\ 0 & 1 & \lambda - 13 \end{vmatrix} = 0$$

Performing cofactor expansion on the first column we obtain

$$(\lambda - 10)((\lambda - 11)(\lambda - 13) - ((1)(-1))) = 0$$

$$(\lambda - 10)(\lambda^2 - 24\lambda + 144) = 0$$

which when factored is

$$(\lambda - 10)(\lambda - 12)(\lambda - 12) = 0$$

and finally tells us that the eigenvalues of this matrix are 10 and 12.

Calculating eigenvalues in this way is computationally intensive even for a small matrix, so as the size of the matrix becomes very large, a significant amount of processing power is necessary to find the eigenvalues.

Notice also that for the matrix

$$\begin{bmatrix} 10 & 2 & 3 \\ 0 & 11 & 1 \\ 0 & 0 & 13 \end{bmatrix}$$

(notice that $a_{3,2}$ was changed from a 1 to a 0) we can see that the eigenvalues are 10, 11, and 13. So it would seem that small changes to the values of the entries may only result in small changes to the eigenvalues. From this observation, back in 1931, Semyon Aronovich Gershgorin devised a method to estimate the eigenvalues of a square matrix with a known error bound.

Gershgorin's Circle Theorem

The concept of the Gershgorin Circle Theorem is that one can take the diagonal entries of an $n \times n$ matrix as the coordinates in the complex plane. These points then act as the centers of n discs which have radii of the sum of the magnitudes of the $n-1$ other entries from the same row. Then, all of the eigenvalues of this matrix will lie within the union of these discs. We now state the theorem formally preceded by a definition.

Definition: If A is an $n \times n$ matrix with complex entries $a_{i,j}$ then $r_i(A) = \sum_{i \neq j} |a_{i,j}|$ is defined as the sum of the magnitudes of the non-diagonal entries of the i^{th} row. Then a Gershgorin disc is the disc $D(a_{i,i}, r_i(A))$ centered at a_{ii} on the complex plane with radius $r_i(A)$.

Theorem: Every eigenvalue of a matrix lies within at least one Gershgorin disc.

Proof (Varga 2010). Let λ be an eigenvalue of a matrix A and let $\mathbf{x} = (x_j)$ be its corresponding non-zero eigenvector. Since $\mathbf{x}$ is an eigenvector of A, we can rewrite $A\mathbf{x} = \lambda\mathbf{x}$ using sigma notation as inner product of each row of A with $\mathbf{x}$:

$$\sum_j a_{ij} x_j = \lambda x_i$$

Subtracting $a_{ii}\mathbf{x}_i$ from both sides splits the sum:

$$\sum_{i \neq j} a_{ij}\mathbf{x}_j = \lambda \mathbf{x}_i - a_{ii}\mathbf{x}_i$$

Due to the fact that $\mathbf{x} \neq 0$, we know there is some $k \in \mathbb{N}$ such that

$$0 < |x_k| = \max\{|x_i| : i \in \mathbb{N}\}$$

With this k, we see

$$\sum_{i \in \mathbb{N}} a_{ki}x_i = \lambda x_k$$

which is the same as

$$\sum_{i \neq j} a_{ki}x_i = (\lambda - a_{k,k})x_k$$

Using the triangle inequality,

$$|\lambda - a_{k,k}||x_k| = \sum_{i \neq j} |a_{ai}||x_i| \leq \sum_{i \neq j} |a_{ki}||x_k| = |x_k|r_k(A)$$

where $\sum_{i \neq j} |a_{ki}||x_k| = |x_k|r_k(A)$ comes from the definition of a Gershgorin Disc. Finally, since $|x_k|$ is greater than 0 we can divide by $|x_k|$, to get

$$|\lambda - a_{kk}| \leq r_k(A)$$

□

As an example, consider the graph of the Gershgorin discs for our sample matrix

$$A = \begin{bmatrix} 10 & 2 & 3 \\ 0 & 11 & 1 \\ 0 & -1 & 13 \end{bmatrix}$$

Since the entries on the main diagonal are $10+0i$, $11+0i$, and $13+0i$, those will be the centers of the three discs in the complex plane. Those three discs will have radii of $|2| + |3| = 5$, $|0| + |1| = 1$, and $|0| + |1| = 1$ respectively see *Figure 2.*

Notice in *Figure 2*, that the two smaller discs are contained within the larger disc. This leads us to an important corollary to Gershgorin's Circle Theorem.

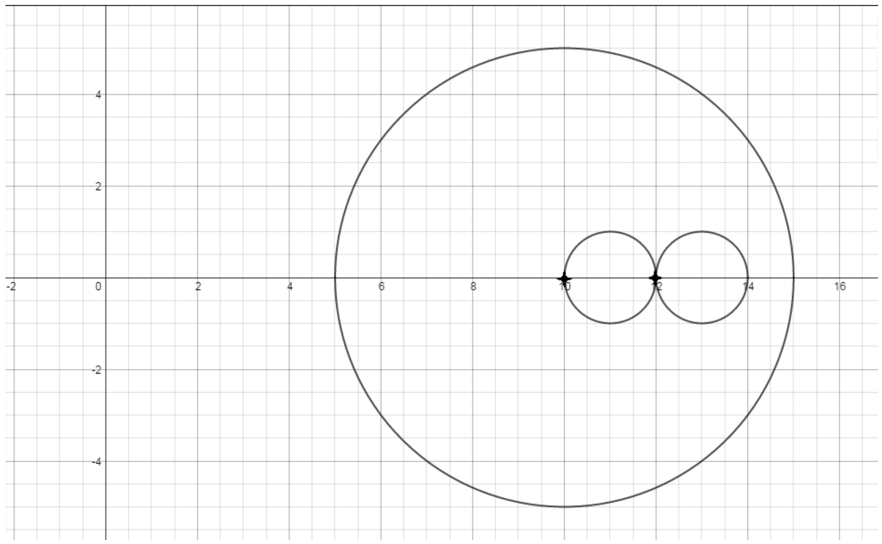

Figure 2: The Gershgorin discs for matrix A

Corollary 1: If $\hat{S}$ is a union of m discs R_i such that $\hat{S}$ is disjoint from all other discs, then $\hat{S}$ contains precisely m eigenvalues (counting multiplicities) of A

Proof of Corollary 1: (Ortega 1990). Let A be an $n \times n$ matrix. Then, we can create a family of matrices $A(t) = tB + D$ where D is the same as A with all the off-diagonal entries reduced to zero and B is the same as A with all the diagonal entries reduced to zero all along the interval $0 \leq t \leq 1$. Then, each Gershgorin disc of $A(t)$ will have the same center of a disc from A, but with its radius scaled by t. Now if we let $R_i(t)$ be the disc with center $a_{i,i}$ and radius $t \sum_{j \neq i} |a_{i,j}|$ and define $\hat{S} = \bigcup_{i=1}^{m} R_i$, we can then set $\hat{S}(t) = \bigcup_{i=1}^{m} R_i(t)$ and $\tilde{S}(t) = \bigcup_{i=m+1}^{n} R_i$. From our assumptions, the sets $\hat{S}$ and $\tilde{S}$ are disjoint and therefore $\hat{S}(t)$ and $\tilde{S}(t)$ are disjoint throughout the interval of t. We can see that $\hat{S}(0)$ contains m eigenvalues of A_0. Since Gershgorin's Circle Theorem tells us that all eigenvalues of A_t are contained in $\hat{S} \cup \tilde{S}$. However, since $\hat{S} \cap \tilde{S} = \emptyset$ and an eigenvalue cannot 'jump' from one set to the other, $\hat{S}(0)$ containing exactly m eigenvalues implies that $\hat{S}(t)$ contains exactly m eigenvalues for the entire interval of t. $\square$

To demonstrate this corollary, see *figure 3* which depicts a subset of the Gershgorin discs for some matrix (with the actual locations of the eigenvalues marked with x's).

Notice that of the three discs, the one that is disjoint contains exactly one eigenvalue, yet of the two continuous discs, one contains both of their eigenvalues and the other has none.

Refinement of Gershgorin's Circle Theorem

Even though we already know a method to precisely calculate the eigenvalues of a matrix, namely cofactor expansion, we may still want to be able to estimate them. Cofactor expansion becomes exponentially more computationally expensive as n increases. This can become cumbersome even when automated, so application of Gershgorin's Theorem may become more useful for many real-world applications.

One thing we can do to make our estimations better is to realize that taking the transpose of a square matrix does not change its eigenvalues. Therefore, we can compare the results of using Gershgorin's Circle Theorem on a matrix A with those from A^T and use the more favorable results.

Example 3: Use Gershgorin's Circle Theorem to find a bound for A^T where

$$A = \begin{bmatrix} 10 & 2 & 3 \\ 0 & 11 & 1 \\ 0 & -1 & 13 \end{bmatrix}$$

The transpose of A is

$$A^T = \begin{bmatrix} 10 & 0 & 0 \\ 2 & 11 & -1 \\ 3 & 1 & 13 \end{bmatrix}$$

See *Figure 4*.

Another, less commonly applicable, corollary of Gershgorin's Circle Theorem which relates back to the first corollary is thus:

Corollary 2. Any Gershgorin disc with radius 0 is an eigenvalue.

Proof of Corollary 2. (Marquis 2016). Suppose $A = [a_{i,j}]$ is a matrix such that at least one of its Gershgorin discs has radius 0. That means that A has at least one row with non-diagonal entries all equal to zero. Without loss of generality, call this the k^{th} row. If we

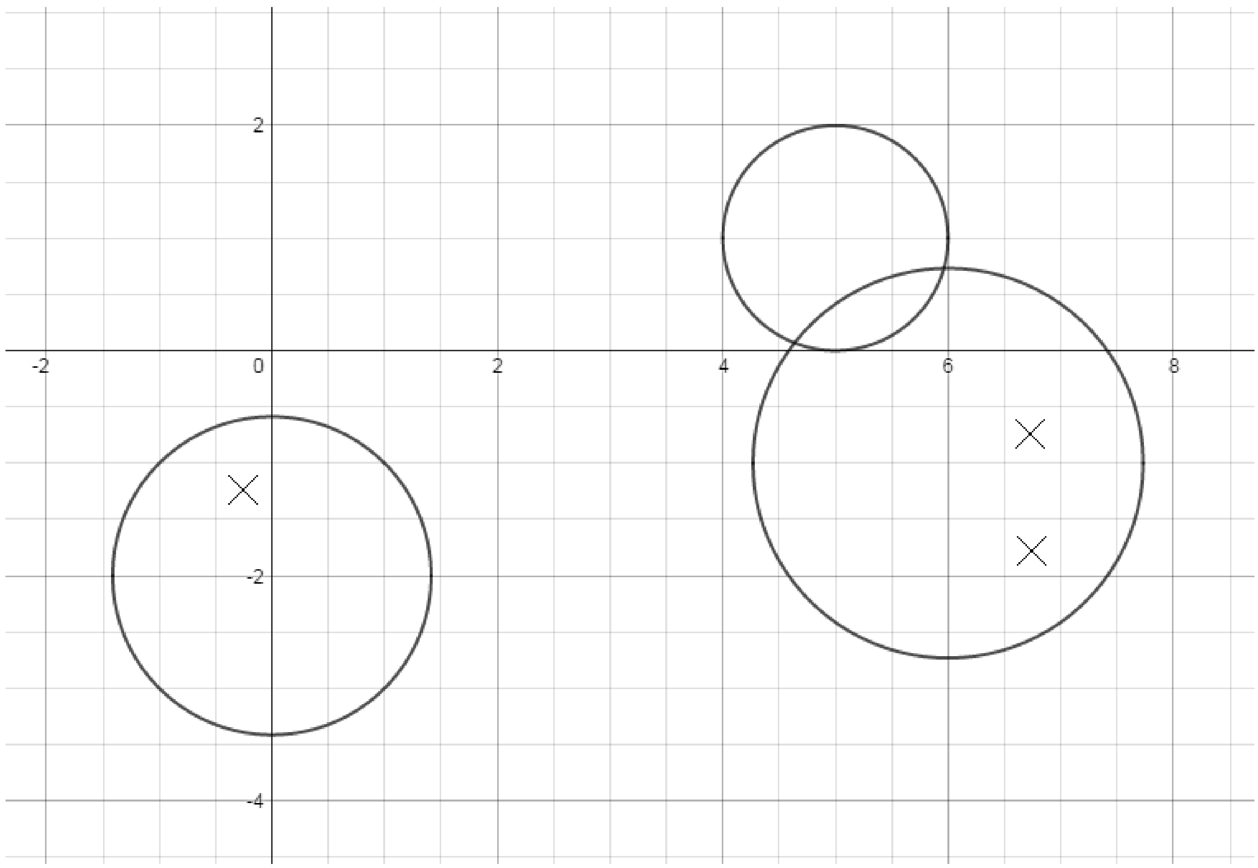

Figure 3: Demonstration of the application of corollary 1

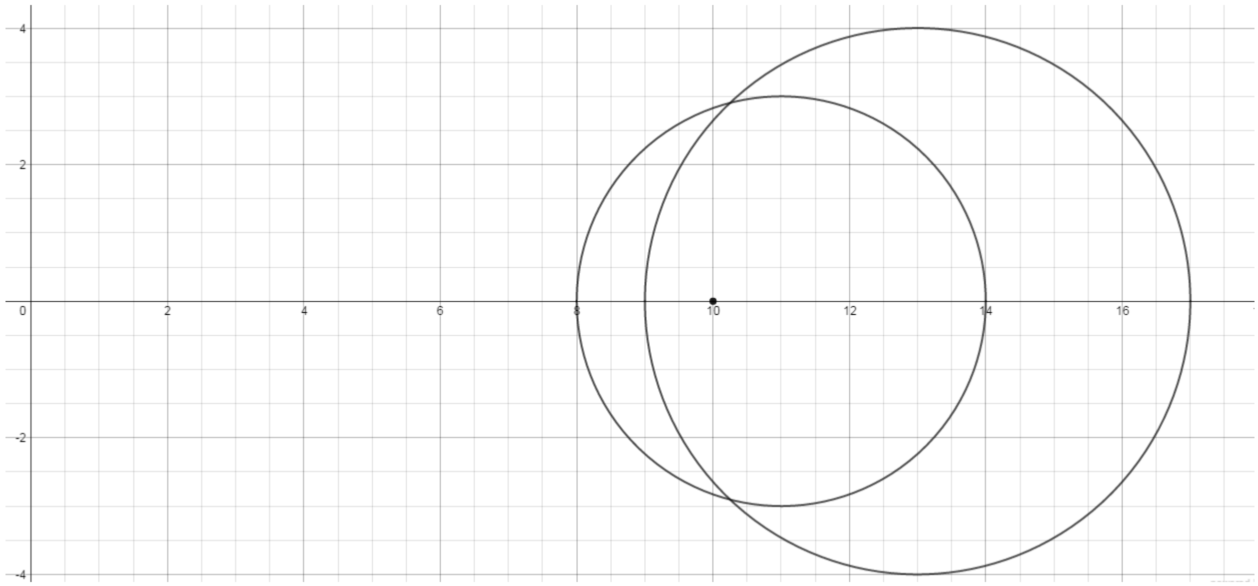

Figure 4: The Gershgorin Discs for A^T

form the matrix $(\lambda I - A)$ and perform cofactor expansion along the k^{th} row, we will get $(\lambda - a_{k,k})(-1)^{k+k}M_{k,k}$. Thus, $(\lambda - a_{k,k})$ is a factor of the characteristic polynomial, so $a_{k,k}$ is a root, so $a_{k,k}$ is an eigenvalue of A. □

Corollary 2 is obvious for a disc of radius 0, or point-disc, by itself, so corollary 2 is only really useful when a point-disc lies within another disc. In that case, corollary 1 would tell us that the two eigenvalues could be anywhere within the two discs, but corollary 2 guarantees that every disc with radius 0 is an eigenvalue.

Notice also that when comparing the Gershgorin Discs for A and A^T, there are different bounds on the possible eigenvalues because the discs are not the same unless A is a symmetric matrix. In our examples along the real axis, the discs of A reach from 5 to 15 whereas the discs of A^T reach from 8 to 17. Similarly, the upper and lower bounds of the complex components at any value along the real axis vary between the two sets of discs. This leads to the realization that we can bound the eigenvalues by the intersection of the unions of the Gershgorin discs for A and A^T i.e. for $A = [a_{i,j}]$ and $A^T = [b_{k,l}]$,

$$\bigcup D(a_{i,i}, R_i) \cap \bigcup D(b_{k,k}, R_k)$$

So, for our example matrix, our bound for the eigenvalues of A becomes the shaded area in *Figure 5.*

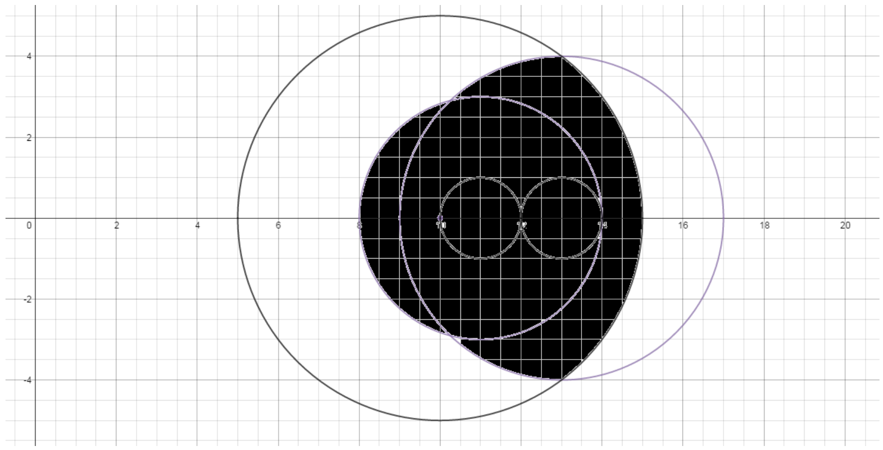

Figure 5: Improved bound using the intersection of the unions of the discs.

Final Thoughts

Whether we are interested in the eigenvalues of a matrix to learn how a linear system acts, as a stepping stone for finding the eigenvectors, or to study matrices in an abstract sense, we need to have an algorithmic method for finding these eigenvalues. Although cofactor expansion works well on small matrices, pending advances in establishing roots of n^{th} degree polynomials, alternative methods must be used.

The power method and inverse power method are popular choices for estimating the eigenvalue with greatest magnitude, but still require iterative processes and, theoretically, could fail if iterated on a vector orthogonal to the dominant eigenvector.

The real power of Gershgorin's Circle Theorem lies in how easily it allows approximation of all of the eigenvalues. Summing $n-1$ terms n times is about as simple of a calculation as can be. In addition, finding the intersection of the union of a matrix $A's$ gershgorin discs with the union of the discs from A^T can be easily implemented to improve estimates of any non-symmetric matrix.

In practice, though Gershgorin's Circle Theorem is powerful, using other techniques to augment Gershgorin's Theorem can make ever more potent estimations. For instance, one can use Gershgorin's Theorem to estimate the dominant eigenvalue, use that to estimate the corresponding eigenvector, and then use the power method on that to improve the estimates of both. One can also precondition a matrix before applying Gershgorin's Theorem in an attempt to decrease the area of the union of the Gershgorin discs.

References

Anton, Howard. Elementary Linear Algebra. Print.

O'Connor, J. J., and E. F. Robertson. "Semyon Aranovich Gershgorin." Gershgorin Biography. Jan. 2012. Web. 12 Apr. 2016.

Ortega, James M. Numerical Analysis: A Second Course. Philadelphia: Society for Industrial and Applied Mathematics, 1990. Print.

On the Accuracy of the Gershgorin Circle Theorem for Bounding the Spread of a Real Symmetric Matrix by David S. Scott https://stmarys-ca.illiad.oclc.org/illiad/illiad.dll?Action%3D10&Form%3I

Varga, Richard S. Gerssgorin and His Circles. Heidelberg: Springer-Verlag, 2010. Print.