5-1996

Geršgorin and Beyond•••

Jason Knight Belnap
Utah State University

Geršgorin and Beyond ···

by

Jason Knight Belnap

Thesis submitted in partial fulfillment
of the requirements for the degree

of

DEPARTMENT HONORS

in

MATHEMATICS

UTAH STATE UNIVERSITY
Logan, UT

1996

Geršgorin's Disk Theorem and Eigenvalue Location

Eigenvalues are useful in various areas of mathematics, such as in testing the critical values of a multivariable function to see if it is a local extrema. One of the more common ways to define eigenvalues is:

Definition (1): Given that A is an n by n matrix, λ is an **eigenvalue** of A if and only if $\det(A - \lambda I_n) = 0$. Any nonzero vector in $Null(A - \lambda I)$ is called an **eigenvector** associated with λ.

It turns out that whenever λ is an eigenvalue of A, there is also a nonzero vector y satisfying $y^T A = \lambda y^T$; such a y is called a **left eigenvector** of A associated with λ. From here on, unless otherwise stated, we shall assume that all matrices are n by n, with n an arbitrary natural number. For notational convenience, we shall denote the entry in the ith row and jth column of A by a_{ij} and the ith row by $A_{(i)}$. For a given vector x, we shall denote its ith entry by x_i. Finally, we introduce the following definition:

Definition (2): The **spectrum** of a matrix A is the set of all eigenvalues of A. We shall denote this by $\sigma(A)$.

Definition 1 provides us with a means of finding eigenvalues; one needs only find the roots of an nth degree polynomial given by the determinant of $A - \lambda I$. An easy task, right? Well, maybe if you're using *Mathematica*©. In general, these roots can be extremely difficult to locate, expecially in the complex plane. Often an approximation is the best that we can do. Before computers and current software, finding eigenvalues or even approximating them was extremely tedious and difficult. Due to that difficulty, many methods for approximating eigenvalues have evolved.

One such method was a theorem by Geršgorin, often called the Geršgorin disc theorem. In order to discuss this theorem and those that followed from it, we must first introduce some notation that we will use throughout this presentation.

$$\text{Given } A,\ R_i(A) \equiv \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|,\ 1 \leq i \leq n$$

$$G_{R,i}(A) \equiv \{z \text{ in } \mathbb{C} \text{ such that } |z - a_{ii}| \leq R_i(A)\}$$

$$G_R(A) \equiv \bigcup_{i=1}^{n} G_{R,i}(A).$$

Where it is not ambiguous, the "(A)" will be left off. We shall refer to G_R as the **Geršgorin Row Region**, the $G_{R,i}$ as the **ith Geršgorin row disk**, and the R_i as the **ith row radius**. The theorem follows:

Theorem (1): Given A, $\sigma(A)$ is a subset of $G_R(A)$. Furthermore, if the union of k of the Geršgorin row disks forms a connected region that is disjoint from the other $n - k$ row disks, then that union must contain exactly k eigenvalues.[1]

Proof: Let λ be an eigenvalue of A and x be an eigenvector associated with λ, such that $Ax = \lambda x$. For this x, there exists an integer p $(1 \leq p \leq n)$ such that $|x_i| \leq |x_p|$ for all i, and $x_p \neq 0$. By computation we get the following:

$$\lambda x_p = (\lambda x)_p = (Ax)_p = A_{(p)}x = \sum_{i=1}^{n} a_{pi}x_i\,,$$

which by subracting $a_{pp}x_p$, gives us

$$x_p(\lambda - a_{pp}) = \lambda x_p - a_{pp}x_p = \sum_{\substack{i=1\\i\neq p}}^{n} a_{pi}x_i$$

By applying the Euclidian length, or absolute value, to both sides we get:

$$|x_p|{\cdot}|\lambda - a_{pp}| = |x_p(\lambda - a_{pp})| = \left|\sum_{\substack{i=1\\i\neq p}}^{n} a_{pi}x_i\right| \leq \sum_{\substack{i=1\\i\neq p}}^{n} |a_{pi}x_i| = \sum_{\substack{i=1\\i\neq p}}^{n} |a_{pi}|{\cdot}|x_i|$$

$$\leq \sum_{\substack{i=1\\i\neq p}}^{n} |a_{pi}|{\cdot}|x_p| = |x_p|\sum_{\substack{i=1\\i\neq p}}^{n} |a_{pi}| = |x_p|R_p(A)$$

We immediately see that $|\lambda - a_{pp}| \leq R_p(A)$, which implies that λ is in $G_{R,p}(A)$, and hence is in $G_R(A)$.

To show the second part of the theorem, suppose that $n_1, \cdots, n_k$ is an increasing sequence of integers from $\{1, 2, \cdots, n\}$, chosen such that $\bigcup_{i=1}^{k} G_{R,n_i}(A)$ is connected and disjoint from the other $n - k$ disks. A can be written as $D + B$ where $D \equiv diag(a_{11}, a_{22}, \cdots, a_{nn})$. Furthermore, let $A_\epsilon \equiv D + \epsilon B$. (Notice that $A_0 = D$ and $A_1 = D + B = A$. Also note that A_ϵ is entrywise continuous on [0,1].) We immediately notice the following:

For any i, $(0 \leq \epsilon \leq 1)$, $R_i(A_\epsilon) = \sum_{\substack{j=1\\j\neq i}}^{n} |a_{ij}{\cdot}\epsilon| = \sum_{\substack{j=1\\j\neq i}}^{n} \epsilon|a_{ij}| = \epsilon R_i(A) \leq R_i(A)$

and therefore, $G_{R,i}(A_\epsilon) = \{z \text{ in } \mathbb{C} : |z - a_{ii}| \leq R_i(A_\epsilon)\} = \{z \text{ in } \mathbb{C} : |z - a_{ii}| \leq \epsilon R_i(A)\}$
$\subseteq \{z \text{ in } \mathbb{C} : |z - a_{ii}| \leq R_i(A)\} = G_{R,i}(A)$

This means that $\bigcup_{i=1}^{k} G_{R,n_i}(A_\epsilon)$ is disjoint from the other $n - k$ disks for A_ϵ. According to Matrix Analysis, the eigenvalues of a matrix are continuous functions of its entries.[2] Consequently, the eigenvalues of A_ϵ are continuous with respect to ϵ; hence $\bigcup_{i=1}^{k} G_{R,n_i}(A_\epsilon)$ will contain the same number of eigenvalues, for any ϵ in [0,1]. Well, for $\epsilon = 0$, it contains the eigenvalues $\{a_{n_1n_1}, a_{n_2n_2}, \cdots, a_{n_kn_k}\}$, and thus it must contain exactly k eigenvalues for $\epsilon = 1$.□

This theorem has an immediate corollary. If we let $C_i(A)$ denote the ith column radius given by $C_i(A) = \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ji}|$, $G_{C,i}(A)$ denote the Geršgorin column disk with radius $C_i(A)$ and center a_{ii}, and $G_C(A)$ represent the Geršgorin column region obtained by taking the union over all column disks, then we get the following corollaries:

Corollary (1): Given A, $\sigma(A)$ is contained in $G_C(A) = \bigcup_{i=1}^{n} G_{C,i}(A)$.[3]
Corollary (2): Given A, $\sigma(A)$ is located in $G_R(A) \cap G_C(A)$.

The first is a result of the facts $G_C(A) = G_R(A^T)$ and $\sigma(A) = \sigma(A^T)$. The second corollary comes from combining Theorem 1 with Corollary 1. These corollaries allow us to limit the eigenvalues of a matrix to an even smaller region. Another corollary that is useful for this purpose isCorollary 3.
Corollary (3): For any matrix A, $\sigma(A)$ is contained in $\bigcap_{S \text{ in } D} \left(G_R(SAS^{-1}) \cap G_C(SAS^{-1})\right)$, where D is any nonempty subset of the set of all invertible n by n matrices.
Proof: Given A, let D be a nonempty set of invertible n by n matrices. By application of Geršgorin's theorem to the matrix SAS^{-1}, the eigenvalues of SAS^{-1} are found in $G_R(SAS^{-1}) \cap G_C(SAS^{-1})$. For any S in D, we know that SAS^{-1} is similar to A. Since similar matrices have the same eigenvalues,[4] we know that the eigenvalues of A must also be found in $G_R(SAS^{-1}) \cap G_C(SAS^{-1})$. Since S was an arbitrary element of D, the eigenvalues of A must be contained in $\bigcap_{S \text{ in } D} \left(G_R(SAS^{-1}) \cap G_C(SAS^{-1})\right)$.□

This corollary provides another means for "cutting down" the region, to get a better estimate for the location of eigenvalues. To see how the previous theorem and corollaries apply, consider the following matrix:

Sample 1

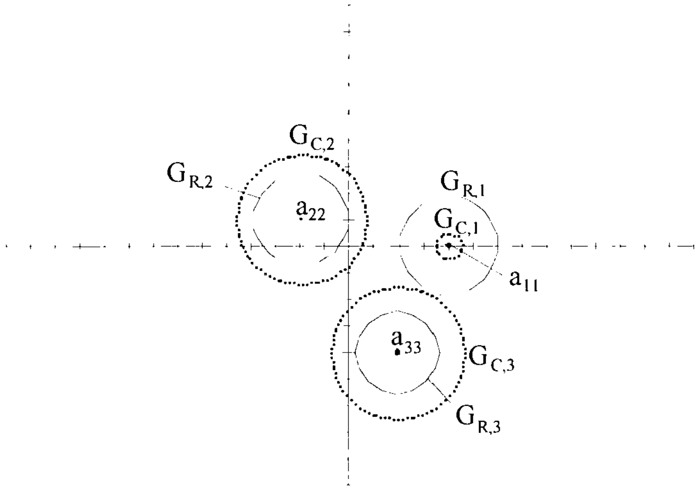
$$A = \begin{bmatrix} 4 & 1 & i \\ -\frac{1}{2} & i-2 & 1-i \\ 0 & \sqrt{3} & 2-4i \end{bmatrix}$$

This gives us the following radii:

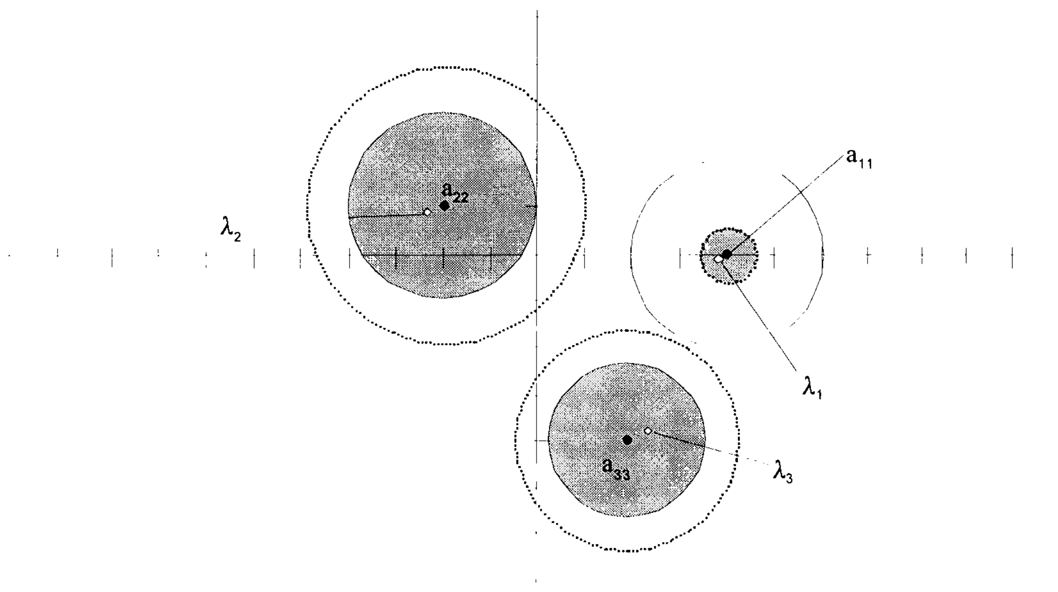
$$R_1(A) = 2, \quad R_2(A) = \frac{1+2\sqrt{2}}{2}, \quad R_3(A) = \sqrt{3}$$

$$C_1(A) = \frac{1}{2}, \quad C_2(A) = 1+\sqrt{3}, \quad C_3(A) = 1+\sqrt{2}$$

Graphically, here are the disks:



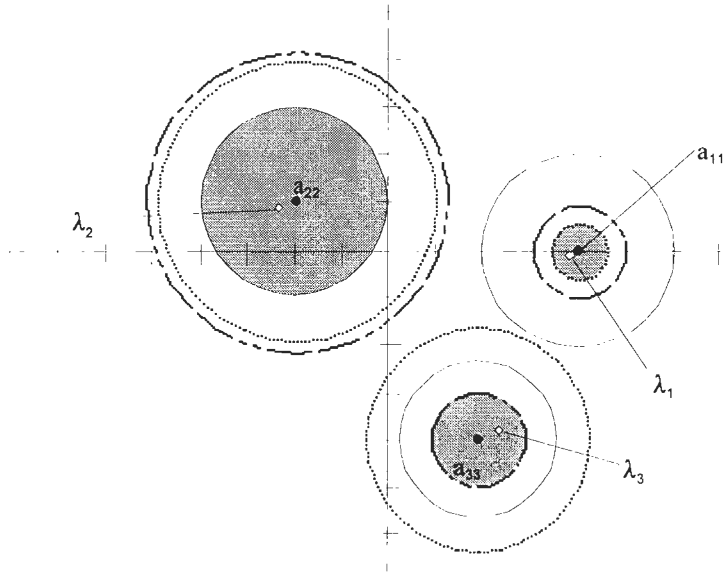
The eigenvalues must be contained in the following regions (eigenvalues marked, λ_i):



Now, if we consider the matrices:

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \; D^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}$$

the region can be changed as follows, using corollary 3:



Notice, that the intersection of the new disks with the old region left the first and second disks unchanged, but reduced the third region. By using other choices for D we could get an even better approximation region, this will be seen later.

Geršgorin's Theorem has other applications. To see some of them we look at a class of matrices, called diagonally dominant matrices.

Definition (3): A matrix A is **strictly diagonally dominant** if and only if for every i in $\{1, 2, \cdots, n\}$, $|a_{ii}| > R_i$. [5]

Theorem (2): If A is strictly diagonally dominant, then the following hold:

a) A is invertible,

b) if all a_{ii} are positive reals, then $Re(\lambda) > 0$, for every λ in $\sigma(A)$,

c) and if A is Hermitian, with the main diagonal positive, then $\sigma(A)$ is contained in the set of positive reals.[6]

Proof: Let A be strictly diagonally dominant, then for any i, $R_i < |a_{ii}| = |a_{ii} - 0|$.

a) This means that 0 is not in $G_{R,i} = \{z \text{ in } \mathbb{C} \text{ such that } |z - a_{ii}| \leq R_i\}$. This means that 0 is not in $\sigma(A)$; therefore, A is invertible.

b) Next, if all a_{ii} is positive and real, then $G_{R,i}$ will be a circle in the complex plane centered on the positive real axis, with the origin as an exterior point (from part **a**). Thus each row disk will be found to the right of the complex axis; which means $Re(\lambda) > 0$.

c) Finally, if A is Hermitian, then the eigenvalues must all be real.[7] By part **b**, since all a_{ii} are positive, this means that for any λ in $\sigma(A)$, $\lambda = Re(\lambda) > 0$. Thus $\sigma(A)$ is contained in the positive reals.□

This theorem, in conjunction with Schur's Unitary Triangularization Theorem, gives us a condition equivalent to the invertibility of a matrix. To see this, we start with a lemma.

Lemma (1): Given an upper triangular matrix A and positive real q, there exists a positive real number r in $[1, \infty)$ such that $R_i(D_r^{-1}AD_r) < q$ for every i. Where $D_r = diag\{r^n, r^{n-1}, \cdots, r\}$.

Proof: By simple matrix multiplication, for any positive real s in $[1, \infty)$, we get:

$$D_s^{-1}AD_s = \begin{bmatrix} a_{11} & \frac{a_{12}}{s} & \cdots & \frac{a_{1n}}{s^{n-1}} \\ 0 & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{a_{(n-1)n}}{s} \\ 0 & \cdots & 0 & a_{nn} \end{bmatrix},$$

giving us $R_i(D_s^{-1}AD_s) = \sum_{j=1}^{n-i} \left| \frac{a_{1(i+j)}}{s^j} \right| = \sum_{j=1}^{n-i} \frac{|a_{1(i+j)}|}{s^j} \leq \frac{1}{s} \sum_{j=1}^{n-i} |a_{1(i+j)}| = \frac{1}{s} R_i(A)$.

Now we know that $t \equiv max\left\{ \max_i \left(\frac{R_i(A)}{q} \right), 1 \right\}$ is finite, so pick $r > t$ and it follows that for every i, $r > \frac{R_i(A)}{q}$ and $R_i(D_r^{-1}AD_r) \leq \frac{1}{r} R_i(A) < q$ with r in $[1, \infty)$□

Theorem (3): A given matrix A is invertible if and only if there exists an invertible matrix Q such that QAQ^{-1} is strictly diagonally dominant.

Proof: ($\Rightarrow$): Let A be an invertible matrix, then by Schur's Theorem,[8] there exists a unitary matrix U such that $T \equiv UAU^*$ is upper triangular, with $t_{ii} \neq 0$ for every i because A is invertible. Now, let $q = \min_i (|(UAU^*)_{i,i}|)$, which is positive; by Lemma 1 there exists a positive real r, such that $R_i(D_r^{-1}UAU^*D_r) < q \leq |t_{ii}|$ for every i. Thus $QAQ^{-1} = (D_r^{-1}U)A(U^*D_r)$ is strictly diagonally dominant where $Q = D_r^{-1}U$.

($\Leftarrow$): Suppose that there exists such a Q, then by Theorem 2a, QAQ^{-1} is invertible. Since Q and Q^{-1} are invertible, by multiplication we know that $Q^{-1}(QAQ^{-1})Q$ is invertible and $Q^{-1}(QAQ^{-1})Q = IAI = A$. Therefore, A is invertible.

□

Not only is $\sigma(A) \subseteq \bigcap_{S \text{ in } D} \left(G_R(SAS^{-1}) \cap G_C(SAS^{-1}) \right)$ for a nonempty set D of invertible matrices, but we can now show a stronger result:

Theorem (4): $\sigma(A) = \bigcap_{S \text{ in } D} \left(G_R(SAS^{-1}) \right)$, where D is the set of all invertible n by n matrices.

Proof: Given A, there exists a unitary U such that $T \equiv UAU^*$ is upper triangular, with A's eigenvalues on the main diagonal. Let $\{q_k\}_{k=1}^{\infty}$ be a strictly decreasing sequence of positive real numbers, such that $\lim_{k \to \infty} (q_k) = 0$. By lemma 1, for every k, there exists a positive real number r_k such that $R_i(D_{r_k}^{-1}TD_{r_k}) < q_k$, for every i. Thus, for every i, $\lim_{k \to \infty} \left(R_i(D_{r_k}^{-1}TD_{r_k}) \right) = 0$.

Therefore, since $(U^*D_{r_k})^{-1} = D_{r_k}^{-1}U$ we get:

$$\bigcap_{S \text{ in } D}\left(G_R(SAS^{-1})\right) \subseteq \lim_{k\to\infty}\left(G_R(D_{r_k}^{-1}UAU^*D_{r_k})\right)$$
$$= \lim_{k\to\infty}\left(\bigcup_{i=1}^{n} G_{R,i}(D_{r_k}^{-1}TD_{r_k})\right)$$
$$= \lim_{k\to\infty}\left(\bigcup_{i=1}^{n}\left\{z \text{ in } \mathbb{C} \text{ such that } |t_{ii} - z| \le R_i(D_{r_k}^{-1}TD_{r_k})\right\}\right)$$
$$= \{t_{11}, t_{22}, \cdots, t_{nn}\} = \sigma(A).$$

Finally by corollary 3, $\sigma(A) \subseteq \bigcap_{S \text{ in } D}\left(G_R(SAS^{-1})\right) \subseteq \sigma(A)$.□

One might now ask if anything useful results can come from partitioning matrices. The answer turns out to be yes, assuming we partition them in a useful manner. For our purposes, we will partition an arbitrary n-square matrix A as follows:

Form 1

$$A = \begin{bmatrix} A_{1,1} & \cdots & A_{1,N} \\ \vdots & \ddots & \vdots \\ A_{N,1} & \cdots & A_{N,N} \end{bmatrix},$$

where $A_{i,i}$ represents an n_i-square matrix with $1 \le i \le N$ and n_1 through n_N nonzero. If $N = 1$, then $A = [A_{1,1}]$. All subsequent partitioned matrices will be assumed to fit this form, unless otherwise stated. The first result comes rather quickly.

Theorem (5): If a matrix A is form 1 and block upper (or lower) triangular ($A_{i,j} = 0$ for $i > j$), then $\sigma(A) \subseteq \bigcup_{i=1}^{n}(F(A_{i,i}) \cap G_R(A_{i,i})) \subseteq \bigcup_{i=1}^{n}(G_R(A_{i,i})) \subseteq G_R(A)$, where $F(A_{i,i})$ is $G_R(SA_{i,i}S^{-1})$, $G_C(QA_{i,i}Q^{-1})$, or the intersection of a collection of these for any number of invertible S and Q.

Proof: Let A be defined as above and F be chosen to meet the described conditions. Since A is block upper triangular, $\sigma(A) = \bigcup_{i=1}^{n}\sigma(A_{i,i})$, but for any i, $\sigma(A_{i,i}) \subseteq F(A_{i,i}) \cap G_R(A_{i,i}) \subseteq G_R(A_{i,i})$ by corollaries 1, 2, and 3. Thus, $\sigma(A) \subseteq \bigcup_{i=1}^{n}(F(A_{i,i}) \cap G_R(A_{i,i})) \subseteq \bigcup_{i=1}^{n} G_R(A_{i,i})$. Now the jth row of $A_{i,i}$ corresponds to $A_{(n_1+n_2+\ldots+n_{i-1}+j)}$ and so for all integers $1 \le i \le N$ and $1 \le j \le n_i$ the following holds:

$G_{R,j}(A_{i,i}) \subseteq G_{R,(n_1+n_2+\ldots+n_{i-1}+j)}(A)$, which implies that

$\bigcup_{j=1}^{n_i} G_{R,j}(A_{i,i}) \subseteq \bigcup_{j=1}^{n_i} G_{R,(n_1+n_2+\ldots+n_{i-1}+j)}(A)$, giving us

$$\begin{aligned}\bigcup_{i=1}^{N} G_R(A_{i,i}) &= \bigcup_{i=1}^{N}\left(\bigcup_{j=1}^{n_i} G_{R,j}(A_{i,i})\right)\\ &\subseteq \bigcup_{i=1}^{N}\left(\bigcup_{j=1}^{n_i} G_{R,(n_1+n_2+\ldots+n_{i-1}+j)}(A)\right)\\ &= \bigcup_{i=1}^{n} G_{R,i}(A) = G_R(A)\end{aligned}$$

□

The righthand side of the previous containment inequality tells us that for block upper triangular matrices, we get the same, if not a better approximation of the eigenvalues of A by working only with the diagonal blocks. However, this theorem does apply to a wider range of matrices. Recall that a permutation matrix (P) is a matrix obtained by permuting rows and/or columns of an identity matrix. Permutation matrices are unitary and real, so $P^{-1} = P^T$. This wider range of matrices are known as reducible matrices:

Definition (4): A matrix A is **reducible** if and only if there exists a permutation matrix P such that:

Form 2

$$PAP^T = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix},$$

where B and D are square and 0 is a zero block.

Notice that block upper triangular matrices are reducible, with $P = I$. Now, since a reducible matrix A is similar to a form 2 matrix, we can get better approximating regions for the eigenvalues of A by looking only at the diagonal blocks of the resultant form 2 matrix.

The only problem with these results is that there are many matrices that are irreducible; even the matrix in sample 1 was irreducible. Because of this, some theorems deal with form 1 matrices in general. Definition 5, Corollary 4, and Theorems 6 and 7 can be found in the Pacific Journal of Mathematics.[10]

Partitioning a matrix A to fit form 1, we can consider $A_{i,i}$ to be a linear transformation from $\mathbb{C}^{n_i}$ to $\mathbb{C}^{n_i}$. For each i, we will let $\|\circ\|_{\alpha_i}$ be any vector norm associated with $\mathbb{C}^{n_i}$ (possibly differing with each i). Then for any i, j $(1 \leq i, j \leq N)$, $A_{i,j}$ can be reguarded as a linear transformation from $\mathbb{C}^{n_j}$ to $\mathbb{C}^{n_i}$. We shall define the norm

$$\|A_{i,j}\| \equiv \sup_{x \epsilon \mathbb{C}^{n_j},\, x \neq 0} \frac{\|A_{i,j}x\|_{\alpha_i}}{\|x\|_{\alpha_j}}.$$

Unless ambiguity will result, we shall leave off the subscripts.

Definition (5): Let A be a form 1 matrix, with the norms defined for each of A's submatrices. A is block diagonally dominant provided for all $1 \leq i \leq N$, $A_{i,i}$ is invertible and

$$\sum_{\substack{k=1 \\ k \neq i}}^{N} \|A_{i,k}\| \leq \left(\left\|A_{i,i}^{-1}\right\|\right)^{-1};$$

A is block strictly diagonally dominant if the inequality is strict.

Here, Feingold and Varga note that we can define

$$\left(\left\|A_{i,i}^{-1}\right\|\right)^{-1} \equiv \inf_{x \in \mathbb{C}^{n_i}, x \neq 0} \left(\frac{\|A_{i,i}\, x_i\|_{\alpha_i}}{\|x_i\|_{\alpha_i}}\right), \text{ for } A_{i,i} \text{ nonsingular.}$$

This allows $\left(\left\|A_{i,i}^{-1}\right\|\right)^{-1}$ to be defined for singular $A_{i,i}$.

Theorem (6): If a matrix A (of form 1) is block strictly diagonally dominant, then A is nonsingular.

Proof: By way of contradiction, assume that A is singular, then there exists a nonzero vector y such that $Ay = 0$. Let x be a normalized version of y partitioned as follows:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}, \text{ where } x_i \text{ has the same number of rows as } A_{i,i}.$$

Additionally, it is normalized such that for any i, $\|x_i\|_{\alpha_i} \leq 1$ and $\|x_r\|_{\alpha_r} = 1$ for some r (this is possible since for some r, $x_r \neq 0$). Before proceeding, let us note that for nonzero x_i we get:

$$\|A_{r,i}\, x_i\|_{\alpha_r} = \frac{\|A_{r,i}\, x_i\|_{\alpha_r} \cdot \|x_i\|_{\alpha_i}}{\|x_i\|_{\alpha_i}} \leq \|A_{r,i}\| \cdot \|x_i\|_{\alpha_i}$$

and that for any $x_i = 0$,

$$\|A_{r,i}\, x_i\|_{\alpha_r} = 0 = \|A_{r,i}\| \cdot \|x_i\|_{\alpha_i}, \text{ since } \|x_i\|_{\alpha_i} = 0.$$

Since $Ax = 0$, $\sum_{i=1}^{N} A_{r,i}\, x_i = 0$ and $\sum_{\substack{i=1 \\ i \neq r}}^{N} A_{r,i}\, x_i = -A_{r,r}\, x_r$. Thus,

$$\|A_{r,r}\, x_r\|_{\alpha_r} = \left\| \sum_{\substack{i=1 \\ i \neq r}}^{N} A_{r,i}\, x_i \right\|_{\alpha_r} \leq \sum_{\substack{i=1 \\ i \neq r}}^{N} \|A_{r,i}\, x_i\|_{\alpha_r}$$

$$\leq \sum_{\substack{i=1 \\ i \neq r}}^{N} \|A_{r,i}\| \cdot \|x_i\|_{\alpha_i} \leq \sum_{\substack{i=1 \\ i \neq r}}^{N} \|A_{r,i}\|$$

But since $A_{r,r}$ was assumed to be invertible, we get the following:

$$\left(\left\|A_{r,r}^{-1}\right\|\right)^{-1} \leq \left(\frac{\left\|A_{r,r}^{-1}(A_{r,r}\, x_r)\right\|_{\alpha_r}}{\|A_{r,r}\, x_r\|_{\alpha_r}} \right)^{-1} = \frac{\|A_{r,r}\, x_r\|_{\alpha_r}}{\|x_r\|_{\alpha_r}} = \|A_{r,r}\, x_r\|_{\alpha_r}.$$

This is our contradiction, since this gives us $\left(\left\|A_{r,r}^{-1}\right\|\right)^{-1} \leq \sum_{\substack{i=1 \\ i \neq r}}^{N} \|A_{r,i}\|$, contrary to the fact that A is block strictly diagonally dominant. Therefore A must be nonsingular.☐

Although this theorem deals with the nonsingularity of certain matrices, it also sets us up perfectly for another eigenvalue perterbation theorem and corollary. The next theorem and corollary, for partitioned matrices, are analogous to Geršgorin's Theorem.

Theorem (7): Let A be a partitioned matrix, then for any λ in $\sigma(A)$,

$$\left(\left\|(A_{j,j} - \lambda I)^{-1}\right\|\right)^{-1} \leq \sum_{\substack{i=1 \\ i \neq j}}^{N} \|A_{j,i}\| \text{ , for some } j.$$

Proof: Let A be of form 1. If λ is an eigenvalue of A, then $A - \lambda I$ is singular and thus cannot be block strictly diagonally dominant, which means by the contrapositive of our last proof that there exists some j such that the inequality of our hypothesis holds.☐

Corollary (4): If A is of form 1 and we define the following block regions,

$$B_i(A) = \left\{ z \text{ in } \mathbb{C} \text{ such that } \left(\left\|(A_{j,j} - zI)^{-1}\right\|\right)^{-1} \leq \sum_{\substack{i=1 \\ i \neq j}}^{N} \|A_{j,i}\| \right\}, \text{ then}$$

$\sigma(A) \subseteq \bigcup_{i=1}^{N} (B_i(A))$.

We have looked at many theorems that have evolved from Geršgorin's Theorem. All of the ones we have looked at basically use or have the appearance of his original theorem. The last theorem we shall present is strongly based upon Geršgorin's Theorem, but makes a unique improvement on it; the theorem is known as the Ovals of Cassini.

For Theorem 8, we introduce two notational definitions:

$$P_k(A) \equiv \sum_{\substack{i=1 \\ i \neq k}}^{n} |a_{ki}| \quad \text{and} \quad Q_k(A) \equiv \sum_{\substack{i=1 \\ i \neq k}}^{n} |a_{ik}|$$

Theorem (8): Given A, each eigenvalue can be found in the complex plane, both on at least one of the ovals of Cassini given by

$$|z - a_{kk}| \cdot |z - a_{mm}| \leq P_k(A) \cdot P_m(A), \quad \text{for } 1 \leq k < m \leq n$$

and on at least one of the ovals given by

$$|z - a_{kk}| \cdot |z - a_{mm}| \leq Q_k(A) \cdot Q_m(A), \quad \text{for } 1 \leq k < m \leq n.$$ [11]

Proof: Given A, suppose λ is an eigenvalue of A with x as a corresponding eigenvector. Now, there exist two distinct integers, $1 \leq k, m \leq n$ such that $|x_i| \leq |x_m| \leq |x_k|$ for every $i \neq k, m$. As in Geršgorin's theorem, we immediately get the following results:

$$\lambda x_k = \sum_{i=1}^{n} a_{ki} x_i \quad \text{which implies } (\lambda - a_{kk}) x_k = \sum_{\substack{i=1 \\ i \neq k}}^{n} a_{ki} x_i \quad \text{and}$$

$$\lambda x_m = \sum_{i=1}^{n} a_{mi} x_i \quad \text{which implies } (\lambda - a_{mm}) x_m = \sum_{\substack{i=1 \\ i \neq m}}^{n} a_{mi} x_i.$$

Now, by applying the absolute value (or euclidean norm) we obtain the following results:

Eq. 1 $$|\lambda - a_{kk}| \cdot |x_k| \leq \sum_{\substack{i=1 \\ i \neq k}}^{n} |a_{ki}| \cdot |x_i| \leq |x_m| \cdot \sum_{\substack{i=1 \\ i \neq k}}^{n} |a_{ki}| = |x_m| \cdot P_k(A) \quad \text{and}$$

$$|\lambda - a_{mm}| \cdot |x_m| \leq \sum_{\substack{i=1 \\ i \neq m}}^{n} |a_{mi}| \cdot |x_i| \leq |x_k| \cdot \sum_{\substack{i=1 \\ i \neq m}}^{n} |a_{mi}| = |x_k| \cdot P_m(A).$$

Finally by multiplication we get

Eq. 2 $$|\lambda - a_{kk}| \cdot |\lambda - a_{mm}| \cdot |x_k| \cdot |x_m| \leq |x_m| \cdot |x_k| \cdot P_k(A) \cdot P_m(A)$$

In all cases, we know that $|x_k| > 0$.

Case 1: Suppose $|x_m| = 0$. By eq.1, $|\lambda - a_{kk}| = 0$ and so $|\lambda - a_{kk}| \cdot |\lambda - a_{mm}| = 0 \leq P_k(A) \cdot P_m(A)$, concluding this case.

Case 2: Suppose $|x_m| \neq 0$. Then by dividing eq.2 by $|x_k| \cdot |x_m|$, we get $|\lambda - a_{kk}| \cdot |\lambda - a_{mm}| \leq P_k(A) \cdot P_m(A)$.

For the Q ovals, the exact procedure can be repeated by using a left eigenvector for x, switching the row/column indexes, and switching P's and Q's.□

In conclusion, let us look at how some of these theorems can improve upon the original Geršgorin regions. Consider the matrix:

Sample 2

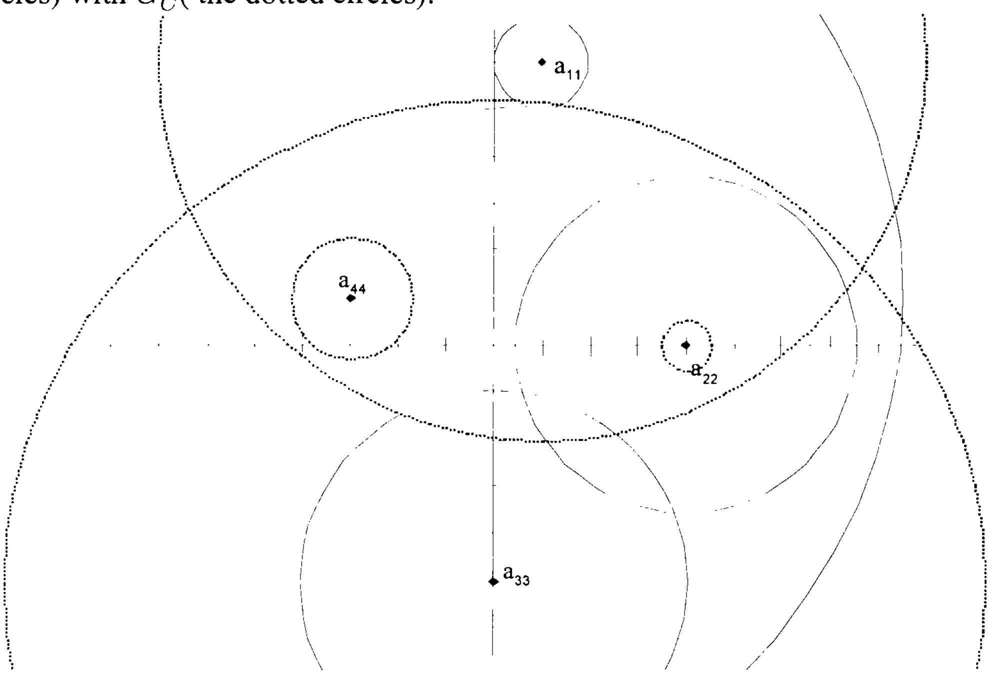
$$A = \begin{bmatrix} 1+6i & 0 & 1 & 0 \\ 2i & 4 & -1 & i+1 \\ 4 & 0 & -5i & 0 \\ 3 & \frac{1}{2} & 8+i & i-3 \end{bmatrix}.$$

This matrix gives the following row and column radii:

$$R_1 = 1 \quad R_2 = 3 + \sqrt{2} \quad R_3 = 4 \quad R_4 = \frac{7}{2} + \sqrt{65}$$

$$C_1 = 9 \quad C_2 = \frac{1}{2} \quad C_3 = 2 + \sqrt{65} \quad C_4 = \sqrt{2}$$

This gives us the following approximation region, obtained by intersecting G_R (the solid circles) with G_C(the dotted circles):



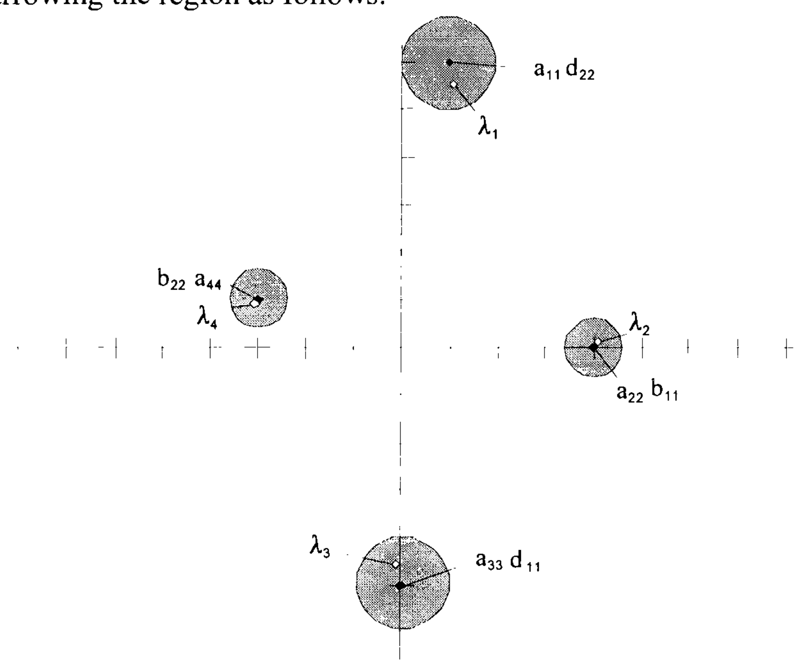
It is clear that this is a terrifying approximation. However, by using the permutation matrix

$$P = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix},$$

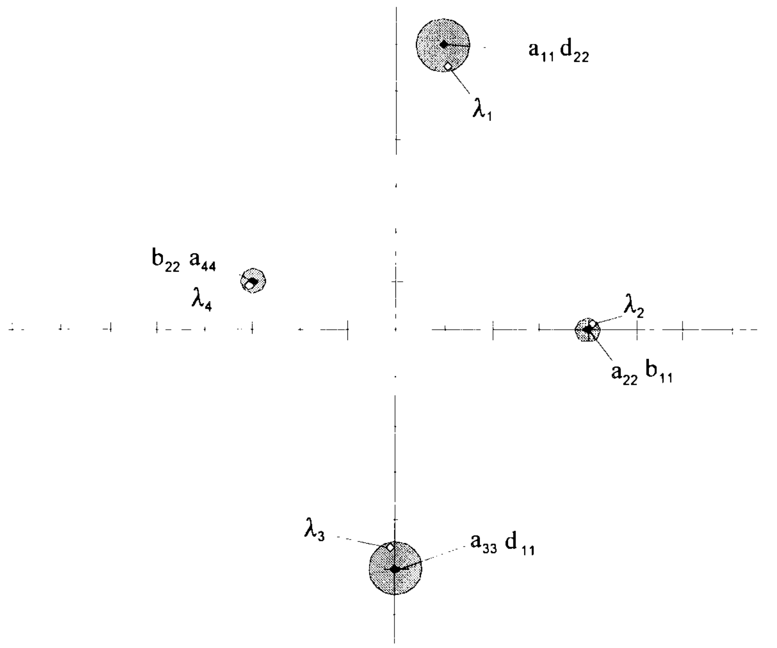
we get

$$P^T AP = \begin{bmatrix} 4 & i+1 & -1 & 2i \\ \frac{1}{2} & i-3 & 8+i & 3 \\ 0 & 0 & -5i & 4 \\ 0 & 0 & 1 & 1+6i \end{bmatrix}$$

which fits Form 2, with $B = \begin{bmatrix} 4 & i+1 \\ \frac{1}{2} & i-3 \end{bmatrix}$ and $D = \begin{bmatrix} -5i & 4 \\ 1 & 1+6i \end{bmatrix}$. Now, applying Theorem 5, we can replace $G_R(A)$ with $(G_R(B) \cap G_C(B)) \cup (G_R(D) \cap G_C(D))$, greatly narrowing the region as follows:



We could also improve on this region by replacing $G_R(A)$ with $\left(G_R(S^{-1}BS) \cap G_C(S^{-1}BS)\right) \cup \left(G_R(Q^{-1}DQ) \cap G_C(Q^{-1}DQ)\right)$ where $S = \begin{bmatrix} 1 & 0 \\ 0 & 2\sqrt{2} \end{bmatrix}$ and $Q = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ which gives an even better region as shown in the next figure (with actual eigenvalues λ_i noted).



These eigenvalue location games could go on and on. As noted, with good choices, these theorems can give good approximations for the eigenvalues of a matrix; however, the matrix operations needed, the row/column summing, and finding good matrices to perturb the regions are often very tedious and time consuming (I can testify to that). Never the less, we must note these theorems for their true worth. Computer software may have progressed far enough to taken their practical use for finding eigenvalues; but as we discussed, the benefits of these theorems go beyond numbers.

Works Cited

[1] Roger A. Horn and Charles R. Johnson, Matrix Analysis (New York : Cambridge University Press 1985) 344.

[2] Horn and Johnson, 539-40.

[3] Horn and Johnson, 346.

[4] Horn and Johnson, 45.

[5] Horn and Johnson, 349.

[6] Horn and Johnson, 349-50.

[7] Horn and Johnson, 170-1.

[8] Stephen Friedburg, Arnold Insel, and Lawrence Spence, Linear Algebra, Second Edition (New Jersey : Prentice Hall 1989) 338.

[9] Horn and Johnson, 351.

[10] David G. Feingold and Richard S. Varga, "Block Diagonally Dominant Matrices and Generalizations of the Gerschgorin Circle Theorem." Pacific Journal of Mathematics 12 (1962) : 1241-9.

[11] Alfred Brauer, "Limits for the Characteristic Roots of a Matrix. II." Duke Math. Journal 14 (1947) : 21-3.