Numerische Mathematik 6, 366—376 (1964)

On Smallest Isolated Gerschgorin Disks for Eigenvalues*

By
RICHARD S. VARGA**

* To ALSTON S. HOUSEHOLDER on his sixtieth birthday.

** The results of this paper were announced at the SIAM Matrix Symposium at Gatlinburg, Tenn., in April, 1964.

§ 1. Introduction

Let $A=(a_{i,j})$ be a fixed $n\times n$ complex matrix, and let $X(\boldsymbol{x})=\operatorname{diag}(x_1, x_2, \ldots, x_n)$, where $x_i>0$, $1\leqq i\leqq n$. Upon forming the matrix $X^{-1}(\boldsymbol{x})\,A\,X(\boldsymbol{x})$, the radii $\Lambda_i(\boldsymbol{x})$ of the Gerschgorin disks $|z-a_{i,i}|\leqq\Lambda_i(\boldsymbol{x})$, determined from row sums of $X^{-1}(\boldsymbol{x})\,A\,X(\boldsymbol{x})$, are given by

$$\Lambda_i(\boldsymbol{x})=\left(\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{i,j}|\,x_j\right)\Big/x_i, \qquad 1\leqq i\leqq n. \tag{1.1}$$

Further, let

$$d_{k,j}=|a_{k,k}-a_{j,j}|, \qquad 1\leqq j,\quad k\leqq n. \tag{1.2}$$

In analogy to [4], we make the following definition.

Definition 1. Let P_k be the set of all vectors $\boldsymbol{x}>\mathbf{0}$ such that

$$d_{k,j}-\Lambda_j(\boldsymbol{x})-\Lambda_k(\boldsymbol{x})\geqq 0 \quad \text{for all} \quad j\neq k. \tag{1.3}$$

If P_k is non-null, then the matrix A admits, under diagonal similarity transformations, an *isolated k-th Gerschgorin disk*.

Assuming there exists an $\boldsymbol{x}_0\in P_k$ for which strict inequality is valid in (1.3) for all $j\neq k$, then it follows from GERSCHGORIN's original paper [3] that there is *exactly* one eigenvalue of A in the disk $|z-a_{k,k}|\leqq\Lambda_k(\boldsymbol{x}_0)$. In order to find improved eigenvalue bounds for this isolated eigenvalue, it is quite natural to think in terms of varying the vector $\boldsymbol{x}_0\in P_k$ so as to decrease $\Lambda_k(\boldsymbol{x}_0)$. Indeed, this approach was already suggested by GERSCHGORIN [3], and has subsequently been employed numerically with success by TAUSSKY [8] and WILKINSON [12]. Yet, in terms of finding the *smallest* possible radius μ for this disk, obviously given by

$$\mu=\inf_{\boldsymbol{x}\in P_k}\Lambda_k(\boldsymbol{x}), \tag{1.4}$$

only recently did HENRICI [4] give a convergent algorithm for finding μ in a special tridiagonal case. The first object of this paper is to extend (Theorem 1) this convergent algorithm to the *general* case. We do this by using the theory of M-matrices [5].[1] Next, we also give a convergent non-linear Gauss-Seidel

[1] The next paper, by Professor JOHN TODD, gives an alternate matrix proof of this extension. The author wishes to express his thanks to Professor TODD for helpful discussions.

type algorithm (Theorem 2) which, for large order matrices, may be more useful in practical computations.

The results for the specific matrix A actually are valid for a *set* of matrices $\mathring{\Omega}_A$. Continuing the investigations of [*10*], we show (Theorem 3) that every point of an annulus $\tau \leqq |z| \leqq \mu$ is an isolated eigenvalue of some $n \times n$ matrix $B \in \mathring{\Omega}_A$. Finally, a numerical example is included.

§ 2. Some basic Lemmas

Since our goal is to obtain bounds for certain eigenvalues of the matrix A, we may first assume without loss of generality that A is *irreducible*[2]. Next, if some set P_k is non-null, we may further assume, again without loss of generality, that $k=1$ so that the *first* Gerschgorin disk $|z-a_{1,1}| \leqq \Lambda_1(\boldsymbol{x})$ can be isolated. This reduction to the case $k=1$ can obviously be accomplished by means of a similarity transformation by a permutation matrix applied to A. We henceforth assume that the set P_1 is non-null, and thus from Definition 1,

$$d_{1,j} - \Lambda_j(\boldsymbol{x}) - \Lambda_1(\boldsymbol{x}) \geqq 0, \qquad 2 \leqq j \leqq n, \tag{2.1}$$

for all $\boldsymbol{x} \in P_1$. Since $\Lambda_j(\beta \boldsymbol{x}) = \Lambda_j(\boldsymbol{x})$ for any $\beta > 0$ and any $1 \leqq j \leqq n$, we may assume, as a final normalization, that $x_1 = 1$ for any $\boldsymbol{x} \in P_1$. Note from (1.1) that $\Lambda_1(\boldsymbol{x})$ is then linear in the variables $x_2, x_3, \ldots, x_n$.

Let $Q = (q_{i,j})$ now be a fixed real $n \times n$ matrix, whose entries are defined by

$$\begin{cases} q_{i,i} = d_{1,i} = |a_{1,1} - a_{i,i}|, & 1 \leqq i \leqq n, \\ q_{1,j} = |a_{1,j}|, \quad 2 \leqq j \leqq n; & q_{i,j} = -|a_{i,j}|, \quad i \neq j, \quad i \neq 1. \end{cases} \tag{2.2}$$

The matrix Q is, by construction, irreducible. For notational simplicity in what is to follow, we introduce the following partitioning of Q:

$$Q = \left[\begin{array}{c|c} 0 & \hat{\boldsymbol{\alpha}}^T \\ \hline -\hat{\boldsymbol{a}} & \widetilde{Q} \end{array}\right], \tag{2.3}$$

where $\widetilde{Q}$ is an $(n-1) \times (n-1)$ principal submatrix of Q, $\hat{\boldsymbol{\alpha}}^T = (|a_{1,2}|, |a_{1,3}|, \ldots, |a_{1,n}|)$ and $\hat{\boldsymbol{a}}^T = (|a_{2,1}|, |a_{3,1}|, \ldots, |a_{n,1}|)$. Let I_{n-1} denote the $(n-1) \times (n-1)$ identity matrix, and let $\hat{\boldsymbol{y}}$ be the column vector with $n-1$ components obtained from $\boldsymbol{y}$ such that $\hat{\boldsymbol{y}}^T = (y_2, \ldots, y_n)$. Conversely, given $\hat{\boldsymbol{y}}$, let $\boldsymbol{y}$ denote the unique column vector with n components determined by $\hat{\boldsymbol{y}}$ and $y_1 = 1$. With this notation, we have

Lemma 1. $\boldsymbol{x} \in P_1$ if and only if $\hat{\boldsymbol{x}} > \hat{\boldsymbol{0}}$ and

$$(\widetilde{Q} - \Lambda_1(\boldsymbol{x}) I_{n-1}) \hat{\boldsymbol{x}} \geqq \hat{\boldsymbol{a}}. \tag{2.4}$$

Moreover, $\boldsymbol{x} \in P_1$ implies

$$Q\boldsymbol{x} \geqq \Lambda_1(\boldsymbol{x}) \boldsymbol{x}. \tag{2.5}$$

[2] An $n \times n$ matrix A is *irreducible* if there exists no $n \times n$ permutation matrix P such that $PAP^T = \begin{bmatrix} A_{1,1} & A_{1,2} \\ 0 & A_{2,2} \end{bmatrix}$, where $A_{1,1}$ and $A_{2,2}$ are square nonvoid submatrices. Otherwise, A is *reducible*.

Proof. Both parts of this lemma are a direct consequence of Definition 1 and our normalization in P_1.

Based on a simple argument using the irreducibility of the matrix A and the inequalities of (2.1), we can show that there exist constants r_i, s_i such that

$$0<r_i\leqq x_i\leqq s_i, \qquad 2\leqq i\leqq n, \tag{2.6}$$

for any $\boldsymbol{x}\in P_1$. Thus, as $\Lambda_1(\boldsymbol{x})$ is linear in the components $x_2, \ldots, x_n$, we see from (2.6) and (2.4) that the set of vectors P_1 is compact. Hence, there necessarily exist vectors $\boldsymbol{z}$ and $\boldsymbol{w}$ in P_1 such that

$$\mu\equiv \inf_{\boldsymbol{x}\in P_1} \Lambda_1(\boldsymbol{x})=\Lambda_1(\boldsymbol{z}), \quad \text{and} \quad \sigma\equiv \sup_{\boldsymbol{x}\in P_1} \Lambda_1(\boldsymbol{x})=\Lambda_1(\boldsymbol{w}). \tag{2.7}$$

From (2.5) of Lemma 1, we necessarily have that $Q\boldsymbol{z}\geqq\mu\boldsymbol{z}$ and $Q\boldsymbol{w}\geqq\sigma\boldsymbol{w}$. We shall actually show by means of the next lemmas that μ and σ are eigenvalues of Q. Moreover, we shall show that the eigenvectors $\boldsymbol{z}$ and $\boldsymbol{w}$ corresponding respectively to μ and σ are uniquely determined in P_1. Finally, the method of proof we employ gives rise to an algorithm for determining $\mu=\Lambda_1(\boldsymbol{z})$ and $\boldsymbol{z}$ which, in the tridiagonal case, reduces to the algorithm of [4].

The next lemma establishes a connection between our problem and the theory of M-matrices[3] [5].

Lemma 2. For any real number $t\leqq\sigma$, $(\widetilde{Q}-tI_{n-1})$ is an M-matrix, and $(\widetilde{Q}-tI_{n-1})^{-1}\,\widehat{\boldsymbol{a}}>\widehat{\boldsymbol{0}}$.

Proof. We consider first the special case $t=\sigma$. With $\sigma=\Lambda_1(\boldsymbol{w})$, $\boldsymbol{w}\in P_1$, let $\widetilde{X}=\operatorname{diag}(w_2, w_3, \ldots, w_n)$, and let $\widetilde{E}\equiv\widetilde{X}^{-1}(\widetilde{Q}-\sigma I_{n-1})\,\widetilde{X}\equiv(e_{i,j})$, $1\leqq i, j\leqq n-1$. It follows that the diagonal entries of $\widetilde{E}$, given by $e_{i,i}=d_{1,i+1}-\sigma$, are positive real numbers from (2.1) since $\boldsymbol{w}\in P_1$, and the off-diagonal entries of $\widetilde{E}$, given by $e_{i,j}=-|a_{i+1,j+1}|\,w_{j+1}/w_{i+1}$, are non-positive real numbers satisfying

$$\sum_{\substack{j=1\\ j\neq i}}^{n-1} |e_{i,j}| =\Lambda_{i+1}(\boldsymbol{w})-\left(|a_{i+1,1}|/w_{i+1}\right), \qquad 1\leqq i\leqq n-1. \tag{2.8}$$

Thus, since $\boldsymbol{w}\in P_1$, we have from (2.1) and (2.8) that

$$e_{i,i}-\sum_{\substack{j=1\\ j\neq i}}^{n-1} |e_{i,j}| \geqq |a_{i+1,1}|/w_{i+1}\geqq 0, \qquad 1\leqq i\leqq n-1, \tag{2.9}$$

which proves that $\widetilde{E}$ is *diagonally dominant* [9, p. 23]. Since A is irreducible, not all of the components of $\widehat{\boldsymbol{a}}$ can vanish, so that for some l, $1\leqq l\leqq n-1$,

$$e_{l,l}>\sum_{j\neq l} |e_{l,j}|. \tag{2.10}$$

If $\widetilde{Q}$ is irreducible, so is $\widetilde{E}$, and thus, with (2.10), $\widetilde{E}$ is *irreducibly diagonally dominant* [9, p. 23]. Because of the sign pattern of the entries of $\widetilde{E}$, it follows

[3] A real $n\times n$ matrix $B=(b_{i,j})$ with $b_{i,j}\leqq 0$ for all $i\neq j$ is defined to be an M-Matrix if B is non-singular, and $B^{-1}\geqq 0$.

[9, p. 85] that $\widetilde{E}$ is an irreducible M-matrix, and as such $\widetilde{E}^{-1}>0$. This of course implies that $\widetilde{Q}-\sigma I_{n-1}$ is an irreducible M-matrix with $(\widetilde{Q}-\sigma I_{n-1})^{-1}>0$. Again, since $\hat{\boldsymbol{a}}$ has at least one component positive, this positivity insures that $(\widetilde{Q}-\sigma I_{n-1})^{-1}\hat{\boldsymbol{a}}>\hat{\mathbf{0}}$. But the above argument is actually valid for $t\leqq\sigma$, and we have thus established the desired result in the case that $\widetilde{Q}$ is irreducible. If $\widetilde{Q}$ is reducible, the desired conclusions are established in a similar (but more tedious) way, the proofs now making use of the reduced normal form [2, p. 74; 9, p. 46] of a reducible matrix. For brevity, we have omitted the proof in the reducible case.

Corollary 1. If λ is any eigenvalue of $\widetilde{Q}$, then $\operatorname{Re}\lambda>\sigma$.

Proof. If B is an M-matrix, it is known [5] that any eigenvalue α of B satisfies $\operatorname{Re}\alpha>0$. The result then follows from Lemma 2.

For any $t\leqq\sigma$, we define the vector $\boldsymbol{y}_t$ from $(\widetilde{Q}-tI_{n-1})\,\hat{\boldsymbol{y}}_t\equiv\hat{\boldsymbol{a}}$, and it follows from Lemma 2 that $\hat{\boldsymbol{y}}_t>\hat{\mathbf{0}}$. Thus, as $\Lambda_1(\boldsymbol{y}_t)>0$, we see that the function $g(t)$ defined by $g(t)\equiv\Lambda_1(\hat{\boldsymbol{y}}_t)$ is positive for all $t\leqq\sigma$. We now show that $g(t)$ is *strictly increasing* and *strictly convex upward.*[4]

Lemma 3. For any $t<\sigma$ and for $\varepsilon>0$ sufficiently small,

$$g(t+\varepsilon)=g(t)+\sum_{k=1}^{\infty}c_k\varepsilon^k,\quad\text{where}\quad c_k>0\ \text{for all}\ k\geqq1. \tag{2.11}$$

Proof. By definition, $(\widetilde{Q}-(t+\varepsilon)I_{n-1})\,\hat{\boldsymbol{y}}_{t+\varepsilon}=\hat{\boldsymbol{a}}$, and $(\widetilde{Q}-tI_{n-1})\,\hat{\boldsymbol{y}}_t=\hat{\boldsymbol{a}}$. If we let $B\equiv(\widetilde{Q}-tI_{n-1})^{-1}$, then it follows that

$$\hat{\boldsymbol{y}}_{t+\varepsilon}=\{I-\varepsilon B\}^{-1}B\hat{\boldsymbol{a}}=\{I-\varepsilon B\}^{-1}\hat{\boldsymbol{y}}_t. \tag{2.12}$$

Thus, for $\varepsilon>0$ sufficiently small,

$$\hat{\boldsymbol{y}}_{t+\varepsilon}=\hat{\boldsymbol{y}}_t+\sum_{k=1}^{\infty}\varepsilon^k B^{k+1}\hat{\boldsymbol{a}}. \tag{2.13}$$

From Lemma 2, we know that $B\geqq0$ and $B\hat{\boldsymbol{a}}>\hat{\mathbf{0}}$, from which it follows inductively that $B^{k+1}\hat{\boldsymbol{a}}>\hat{\mathbf{0}}$. Recalling that $\Lambda_1(\boldsymbol{x})$ is linear and that $\Lambda_1(\boldsymbol{x})=\hat{\boldsymbol{\alpha}}^T\hat{\boldsymbol{x}}$, then the series expansion for $g(t+\varepsilon)=\Lambda_1(\boldsymbol{y}_{t+\varepsilon})$ in (2.11) directly follows from (2.13) with

$$c_k\equiv\hat{\boldsymbol{\alpha}}^T B^{k+1}\hat{\boldsymbol{a}},\qquad k\geqq1. \tag{2.14}$$

Thus, the coefficients c_k are all positive, completing the proof.

Lemma 4. $\boldsymbol{y}_t\in P_1$ if and only if $t\geqq g(t)$.

Proof. Since $\hat{\boldsymbol{y}}_t>\hat{\mathbf{0}}$ for all $t\leqq\sigma$, it follows from Lemma 1 that $\boldsymbol{y}\in P_1$ if and only if $(\widetilde{Q}-\Lambda_1(\boldsymbol{y}_t)I_{n-1})\,\hat{\boldsymbol{y}}_t\geqq\hat{\boldsymbol{a}}$, or equivalently, $(\widetilde{Q}-g(t)I_{n-1})\,\hat{\boldsymbol{y}}_t\geqq\hat{\boldsymbol{a}}$. Now,

$$(\widetilde{Q}-g(t)I_{n-1})\,\hat{\boldsymbol{y}}_t=\hat{\boldsymbol{a}}+(t-g(t))\,\hat{\boldsymbol{y}}_t, \tag{2.15}$$

and it therefore follows that $(\widetilde{Q}-g(t)I_{n-1})\,\hat{\boldsymbol{y}}_t\geqq\hat{\boldsymbol{a}}$ if and only if $t\geqq g(t)$, which completes the proof.

We now show that there exist values of t such that $t\geqq g(t)$.

[4] More precisely, $g(t)$ is strictly absolutely monotonic [11].

Lemma 5. For any $\boldsymbol{x} \in P_1$, $t \geqq g(t)$ where $\Lambda_1(\boldsymbol{x}) \equiv t$. Moreover, $(\widetilde{Q} - tI_{n-1})\hat{\boldsymbol{x}} \geqq \hat{\boldsymbol{a}}$ with strict inequality for some component if and only if $t > g(t)$.

Proof. Since $\boldsymbol{x} \in P_1$ and $\Lambda_1(\boldsymbol{x}) \equiv t$, then from Lemma 1 we know that $(\widetilde{Q} - tI_{n-1})\hat{\boldsymbol{x}} \geqq \hat{\boldsymbol{a}}$, and hence $\hat{\boldsymbol{x}} \geqq \hat{\boldsymbol{y}}_t > \hat{\boldsymbol{0}}$, using Lemma 2. Thus, $t = \Lambda_1(\boldsymbol{x}) \geqq \Lambda_1(\boldsymbol{y}_t) = g(t)$, which, by Lemma 4, assures us that $\boldsymbol{y}_t \in P_1$. Moreover, if equality is valid in all components, i.e., $(\widetilde{Q} - tI_{n-1})\hat{\boldsymbol{x}} = \hat{\boldsymbol{a}}$, then surely $\boldsymbol{x} = \boldsymbol{y}_t$ and thus $t = \Lambda_1(\boldsymbol{x}) = \Lambda_1(\boldsymbol{y}_t) = g(t)$. To prove the remaining part of this Lemma, we assume for simplicity that $\widetilde{Q}$ is irreducible. From the proof of Lemma 2, we then have that $(\widetilde{Q} - tI_{n-1})^{-1} > 0$. If $(\widetilde{Q} - tI_{n-1})\hat{\boldsymbol{x}} \geqq \hat{\boldsymbol{a}}$ with *strict* inequality for some component, the positivity of $(\widetilde{Q} - tI_{n-1})^{-1}$ allows us to deduce that $\hat{\boldsymbol{x}} > \hat{\boldsymbol{y}}_t$, i.e., strict inequality in *all* components, and thus $t = \Lambda_1(\boldsymbol{x}) > \Lambda_1(\boldsymbol{y}_t) = g(t)$. The extension of these results to the case when $\widetilde{Q}$ is reducible is again based on the reduced normal form of $\widetilde{Q}$, and is omitted.

Corollary 2. There exist two positive real numbers μ_1, σ_1 with $\mu_1 = g(\mu_1)$ and $\sigma_1 = g(\sigma_1)$ such that $t \geqq g(t)$ if and only if $0 < \mu_1 \leqq t \leqq \sigma_1$.

Proof. Since, by Lemma 5, there exist positive values of t with $t \geqq g(t)$, this result then follows directly from the strictly increasing and strictly convex property of $g(t)$, established in Lemma 3. We remark that if $\mu_1 < \sigma_1$, then $t > g(t)$ if and only if $\mu_1 < t < \sigma_1$.

Lemma 6. $\mu_1 = \mu = \inf_{\boldsymbol{x} \in P_1} \Lambda_1(\boldsymbol{x})$, and $\sigma_1 = \sigma = \sup_{\boldsymbol{x} \in P_1} \Lambda_1(\boldsymbol{x})$. Moreover, μ and σ are eigenvalues of Q, whose eigenvectors $\boldsymbol{y}_\mu$ and $\boldsymbol{y}_\sigma$ are uniquely determined in P_1.

Proof. Since $\boldsymbol{y}_{\mu_1} \in P_1$ by Lemma 4 and Corollary 2, it follows that $\mu_1 \geqq \inf_{\boldsymbol{x} \in P_1} \Lambda_1(\boldsymbol{x}) \equiv \mu$. On the other hand, there is a vector $\boldsymbol{z} \in P_1$ such that $\Lambda_1(\boldsymbol{z}) = \mu$. Thus, from Lemma 5 and Corollary 2, $\mu \geqq \mu_1$, and we thus have $\mu = \mu_1$. Next, since $(\widetilde{Q} - \mu I_{n-1})\hat{\boldsymbol{y}}_\mu = \hat{\boldsymbol{a}}$ by definition and $g(\mu) = \mu$, it is readily verified from (2.3) that $Q\boldsymbol{y}_\mu = \mu \boldsymbol{y}_\mu$, and as $\boldsymbol{y}_\mu > \boldsymbol{0}$, then $\boldsymbol{y}_\mu$ is an eigenvector of Q corresponding to the eigenvalue μ. To prove a somewhat stronger form of the final result, suppose that $Q\boldsymbol{x} = \mu \boldsymbol{x}$ where $\boldsymbol{x}$ is any complex vector with $\boldsymbol{x} \neq \boldsymbol{0}$. If x_1 were zero, then μ would be an eigenvalue of $\widetilde{Q}$, which contradicts the result of Corollary 1. Thus, $x_1 \neq 0$. Forming $(\hat{\boldsymbol{x}}/x_1)$, it follows from $Q\boldsymbol{x} = \mu \boldsymbol{x}$ that $(\widetilde{Q} - \mu I_{n-1})(\hat{\boldsymbol{x}}/x_1) = \hat{\boldsymbol{a}}$, but as $(\widetilde{Q} - \mu I_{n-1})$ is non-singular, then necessarily $(\hat{\boldsymbol{x}}/x_1) = \hat{\boldsymbol{y}}_\mu > \hat{\boldsymbol{0}}$. Thus, $\boldsymbol{x} = x_1 \boldsymbol{y}_\mu$. Similar arguments prove analogous results for $\sigma_1 = \sigma$.

To complete this section, we point out that the Gerschgorin circles for the irreducible matrix $X^{-1}(\boldsymbol{y}_\mu) Q X(\boldsymbol{y}_\mu)$ *all* pass through μ, i.e.,

$$|q_{i,i} - \mu| = \Lambda_i(\boldsymbol{y}_\mu) \quad \text{for all} \quad 1 \leqq i \leqq n, \tag{2.16}$$

which is related to a result of TAUSSKY [7]. From (2.16), it follows [*10*, Theorem 3] that μ is a boundary point of the minimal Gerschgorin set for Q. The same is true for σ.

§ 3. First Convergence Theorem

With the lemmas of the previous section, we now prove

Theorem 1. Let A be an irreducible $n \times n$ matrix which admits a first isolated Gerschgorin disk. Then, the smallest radius μ, under all diagonal similarity

transformations, for this isolated disk is an eigenvalue of the $n \times n$ matrix Q, and its corresponding eigenvector $\boldsymbol{y}_\mu$ is uniquely determined in P_1. If $\boldsymbol{x}_0 \in P_1$, and $Q\boldsymbol{x}_0 \geqq \Lambda_1(\boldsymbol{x}_0)\boldsymbol{x}_0$, with strict inequality for at least one component, then the sequence of vectors $\{\boldsymbol{x}_i\}_{i=0}^{\infty}$ defined by

$$(\widetilde{Q} - \Lambda_1(\boldsymbol{x}_i) I_{n-1})\, \hat{\boldsymbol{x}}_{i+1} = \hat{\boldsymbol{a}}, \qquad i \geqq 0, \tag{3.1}$$

are all elements of P_1 with $\lim_{i\to\infty} \boldsymbol{x}_i = \boldsymbol{y}_\mu$, and the sequence $\{\Lambda_1(\boldsymbol{x}_i)\}_{i=0}^{\infty}$ is strictly decreasing with $\lim_{i\to\infty} \Lambda_1(\boldsymbol{x}_i) = \mu$.

Proof. The first part of this theorem is a restatement of Lemma 6. Next, for any $\boldsymbol{r} \in P_1$, it is readily verified that $Q\boldsymbol{r} \geqq \Lambda_1(\boldsymbol{r})\boldsymbol{r}$ with strict inequality for at least one component if and only if $(\widetilde{Q} - \Lambda_1(\boldsymbol{r}) I_{n-1})\, \hat{\boldsymbol{r}} \geqq \hat{\boldsymbol{a}}$, with strict inequality for at least one component. Thus, with Lemma 5, the hypothesis tells us that $t_0 > g(t_0)$ where $t_0 \equiv \Lambda_1(\boldsymbol{x}_0)$. Next, from (3.1), it follows that $\boldsymbol{x}_1 \equiv \boldsymbol{y}_{t_0}$, and we now ask if $Q\boldsymbol{x}_1 \geqq \Lambda_1(\boldsymbol{x}_1)\boldsymbol{x}_1$, with strict inequality for some component. Since $\Lambda_1(\boldsymbol{x}_1) = g(t_0)$, it follows from (2.15) that

$$\begin{aligned} &(\widetilde{Q} - \Lambda_1(\boldsymbol{x}_1) I_{n-1})\, \hat{\boldsymbol{x}}_1 \\ &\quad = (\widetilde{Q} - g(t_0) I_{n-1})\, \hat{\boldsymbol{y}}_{t_0} \\ &\quad = \hat{\boldsymbol{a}} + (t_0 - g(t_0))\, \hat{\boldsymbol{y}}_{t_0} > \hat{\boldsymbol{a}}, \end{aligned} \tag{3.2}$$

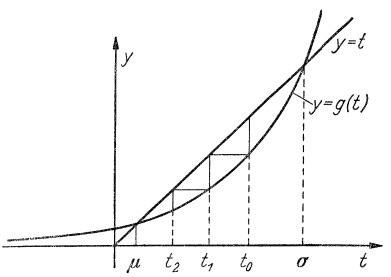

Fig. 1

so that inequality now holds for *all* $n-1$ components. This means that the above argument applied to $\boldsymbol{x}_0$ can be applied to $\boldsymbol{x}_1$, and thus $t_1 = \Lambda_1(x_1) > g(t_1)$. Inductively, with $t_i \equiv \Lambda_1(\boldsymbol{x}_i)$ and $\boldsymbol{x}_{i+1} = \boldsymbol{y}_{t_i}$, it follows that

$$t_0 > g(t_0) = t_1 > g(t_1) = t_2 > \cdots \tag{3.3}$$

so that the sequence $\{\Lambda_1(\boldsymbol{x}_i)\}_{i=0}^{\infty}$ is strictly decreasing. Similarly, the components of the vector sequence $\{\boldsymbol{x}_i\}_{i=0}^{\infty}$ are nonincreasing. If $\lim_{i\to\infty} t_i = \delta$, then clearly $\delta \geqq \mu$, but since $t_i > g(t_i) = t_{i+1} > t_{i+2}$, it necessarily follows that $\delta = g(\delta)$, which from Corollary 2 implies that $\delta = \mu$, completing the proof.

The iterative procedure of (3.1) has a rather simple geometrical interpretation. Starting with $\boldsymbol{x}_0 \in P_1$ such that $\mu < \Lambda_1(\boldsymbol{x}_0) < \sigma$, it follows that the Gerschgorin disks $|q_{k,k} - z| \leqq \Lambda_k(\boldsymbol{x}_0)$, $k > 1$, do *not* all touch the first (isolated) disk $|z| \leqq \Lambda_1(\boldsymbol{x}_0)$. The essence of (3.1) is that we now determine a positive vector $\boldsymbol{x}_1$ such that the *new* Gerschgorin disks $|q_{k,k} - z| \leqq \Lambda_k(\boldsymbol{x}_1)$, $k > 1$, are now all *tangent* to the *old* first disk $|z| \leqq \Lambda_1(\boldsymbol{x}_0)$. This is done by decreasing the components x_j of $\boldsymbol{x}_0$, $j > 1$, in such a way that the radii $\Lambda_j(\boldsymbol{x})$ all increase for $j > 1$. But then, the first disk has smaller radius, viz. $\Lambda_1(\boldsymbol{x}_1)$, and the process can be repeated.

The iteration of (3.1) to find μ, the smallest radius of the first isolated Gerschgorin disk, is actually the *method of successive substitutions*, which is represented by the staircase curve lying between $y = t$ and $y = g(t)$ in Fig. 1. We also remark that monotone convergence of the $\Lambda_1(\boldsymbol{x}_i)$ to μ is had for *any* real initial value $\Lambda_1(\boldsymbol{x}_0)$, with $\Lambda_1(\boldsymbol{x}_0) < \sigma$.

The iterative procedure of (3.1) requires at each step the solution of a system of $(n-1)$ linear equations in $(n-1)$ unknowns. In [4], HENRICI considered the special case where $\widetilde{Q}$ is essentially tridiagonal, since such matrices arose naturally in the problems of finding the zeros of a polynomial. For this case, the algorithm of [4] for solving the matrix equation (3.1) simply reduces to Gaussian elimination which is known to be efficient for such matrix problems.

It is natural to ask about the *rate* of convergence of the $\Lambda_1(\boldsymbol{x}_i)$ to μ. First, it is quite clear that the rate of convergence for the iterative method of (3.1) must be *linear*. In fact, writing $\boldsymbol{x}_i=\boldsymbol{y}_\mu+\boldsymbol{\epsilon}_i$, then $\Lambda_1(\boldsymbol{x}_i)=\mu+\Lambda_1(\boldsymbol{\epsilon}_i)$, $i\geqq 0$, and since $(\widetilde{Q}-\Lambda_1(\boldsymbol{x}_i)I_{n-1})\hat{\boldsymbol{x}}_{i+1}=\hat{\boldsymbol{a}}=(\widetilde{Q}-\mu I_{n-1})\boldsymbol{y}_\mu$, it follows that

$$\frac{\Lambda_1(\boldsymbol{\epsilon}_{i+1})}{\Lambda_1(\boldsymbol{\epsilon}_i)}=\Lambda_1\{(\widetilde{Q}-\Lambda_1(\boldsymbol{x}_i)I_{n-1})^{-1}\hat{\boldsymbol{y}}_\mu\}\sim\Lambda_1\{(\widetilde{Q}-\mu I_{n-1})^{-2}\hat{\boldsymbol{a}}\},\qquad i\to\infty. \tag{3.4}$$

It is also natural to apply other iterative methods to this problem of finding the smallest real zero, μ, of $f(t)\equiv t-g(t)$. As an example, consider *Newton's method*. Given any real number $t<\sigma$, then previous definitions give us

$$(\widetilde{Q}-tI_{n-1})\hat{\boldsymbol{y}}_t=\hat{\boldsymbol{a}}\quad\text{and}\quad\Lambda_1(\boldsymbol{y}_t)=g(t). \tag{3.5}$$

Define now the vector $\boldsymbol{w}_t$ and the scalar $h(t)$ by

$$(\widetilde{Q}-tI_{n-1})\hat{\boldsymbol{w}}_t=\hat{\boldsymbol{y}}_t\quad\text{and}\quad\Lambda_1(\boldsymbol{w}_t)=h(t). \tag{3.6}$$

Clearly, $\hat{\boldsymbol{w}}_t=(\widetilde{Q}-tI_{n-1})^{-2}\hat{\boldsymbol{a}}$, and thus from (2.13), we see that $h(t)=g'(t)$. Hence, Newton's method is then defined by

$$t_{i+1}\equiv t_i-\frac{t_i-g(t_i)}{1-h(t_i)},\qquad\text{when } h(t_i)\neq 1. \tag{3.7}$$

The point here is that the derivative $g'(t)$, which is necessary in the formulation of Newton's method, can be directly calculated by solving an additional system (3.6) of $(n-1)$ equations in $(n-1)$ unknowns. Though the rate of convergence of this method is quadratic, convergence of these iterates t_i to μ cannot be guaranteed for all initial values $t_0<\sigma$, as is clear from Fig. 1. On the other hand, we remark that the iterates t_i of Newton's method (3.7) converge monotonically for the initial value $t_0=0$, but the associated vectors $\boldsymbol{y}_i$ are no longer elements of P_1.

§ 4. Second Convergence Theorem

The particular algorithm given in Theorem 1 required the solution of a system of $(n-1)$ linear equations in $(n-1)$ unknowns at each iteration. For large n, this algorithm may not be practical unless the matrix $\widetilde{Q}$ has some special structure (e.g. tridiagonal). To derive another convergent algorithm, we recall from Lemma 1 that

$$\widetilde{Q}\,\hat{\boldsymbol{x}}\geqq\hat{\boldsymbol{a}}+\Lambda_1(\boldsymbol{x})\,\hat{\boldsymbol{x}}\quad\text{for any}\quad\boldsymbol{x}\in P_1. \tag{4.1}$$

As in Theorem 1, we assume that we are given an initial $\boldsymbol{x}^{(0)}\in P_1$ such that strict inequality is valid for at least one component in (4.1). We may, without loss of generality, assume that strict inequality in (4.1) occurs in the second vector

component. This implies, upon rewriting, that

$$|a_{1,2}|\,(x_2^{(0)})^2+\Big\{\sum_{l>2}|a_{1,l}|\,x_l^{(0)}-q_{2,2}\Big\}\,(x_2^{(0)})+\sum_{l\neq 2}|a_{2,l}|\,x_l^{(0)}<0, \tag{4.2}$$

which is a (possibly degenerate) *quadratic* inequality in $x_2^{(0)}$. Since A is irreducible and $\boldsymbol{x}^{(0)}>\mathbf{0}$, the last term of (4.2) cannot vanish and is thus positive. Hence, if we define $x_2^{(1)}$ as the least positive root of

$$|a_{1,2}|\,(x_2^{(1)})^2+\Big\{\sum_{l>2}|a_{1,l}|\,x_l^{(0)}-q_{2,2}\Big\}\,(x_2^{(1)})+\sum_{l\neq 2}|a_{2,l}|\,x_l^{(0)}=0, \tag{4.3}$$

it is clear that $0<x_2^{(1)}<x_2^{(0)}$. We now show that this new vector $\boldsymbol{x}^{(1)}$, obtained by reducing the second component $x_2^{(0)}$ to $x_2^{(1)}$ is still an element of the set P_1. The remaining inequalities of (4.1) can be written for $j>2$ as

$$\begin{aligned}\{|a_{1,2}|\,x_j^{(0)}+|a_{j,2}|\}\,x_2^{(0)}+\Big\{(x_j^{(0)})\Big(\sum_{l>2}|a_{1,l}|\,x_l^{(0)}\Big)+\\ +\sum_{\substack{l\neq j\\ l\neq 2}}|a_{j,l}|\,x_l^{(0)}-q_{j,j}\,x_j^{(0)}\Big\}\leqq 0,\end{aligned} \tag{4.4}$$

which are *linear* inequalities in $x_2^{(0)}$ for all $j>2$. Because the coefficients of $x_2^{(0)}$ in (4.4) are nonnegative for $j>2$, it is clear that replacing $x_2^{(0)}$ by a smaller quantity leaves all these inequalities unchanged. Thus, we have that $\boldsymbol{x}^{(1)}\in P_1$. Moreover, since $\boldsymbol{x}^{(1)}\leqq\boldsymbol{x}^{(0)}$, then $\Lambda_1(\boldsymbol{x}^{(1)})\leqq\Lambda_1(\boldsymbol{x}^{(0)})$.

We now point out that since A is irreducible, at least one coefficient of $x_2^{(0)}$ in (4.4) is positive for some $j>2$. Thus, while the new vector $\boldsymbol{x}^{(1)}$ by definition gives equality in (4.3), there is some $j>2$ for which strict inequality is now valid in (4.4) for this vector. This means that the above process can be *continued*. In fact, there are several natural ways in which the iteration can be continued. One can *cyclically* improve the components x_j in succession, $2\leqq j\leqq n$, by solving at each step a quadratic equation similar to (4.3). This corresponds to a *nonlinear Gauss-Seidel iterative method* [cf. *1*]. One can also improve the components x_j by a *free-steering method* [cf. *6*], where one makes the added assumption that each component x_j, $2\leqq j\leqq n$, is improved infinitely often. For simplicity, we consider only the cyclic non-linear Gauss-Seidel iterative method, although the basic conclusions are valid also for free-steering methods. Starting with strict inequality in (4.2) for the second component x_2, let $\boldsymbol{x}^{(1)}$ now denote the vector obtained after having improved the components x_j in succession, $2\leqq j\leqq n$. Using the irreducibility of A, it follows that $\Lambda_1(\boldsymbol{x}^{(1)})<\Lambda_1(\boldsymbol{x}^{(0)})$, and in general $\Lambda_1(\boldsymbol{x}^{(n+1)})<\Lambda_1(\boldsymbol{x}^{(n)})$. Moreover, since each $\boldsymbol{x}^{(n)}\in P_1$, and $\boldsymbol{x}^{(n+1)}\leqq\boldsymbol{x}^{(n)}$, it is not difficult to verify that $\lim_{n\to\infty}\boldsymbol{x}^{(n)}=\boldsymbol{y}_\mu$ and $\lim_{n\to\infty}\Lambda_1(\boldsymbol{x}^{(n)})=\mu$, which gives us

Theorem 2. Let A be an irreducible $n\times n$ matrix which admits a first isolated Gerschgorin disk, and let $\{\boldsymbol{x}^{(n)}\}_{n=0}^{\infty}$ be the sequence of vector iterates of the cyclic non-linear Gauss-Seidel iterative method, where $Q\boldsymbol{x}^{(0)}\geqq\Lambda_1(\boldsymbol{x}^{(0)})\boldsymbol{x}^{(0)}$ with strict inequality for at least one component. Then, all the vectors $\boldsymbol{x}^{(n)}$ are elements of P_1 with $\lim_{n\to\infty}\boldsymbol{x}^{(n)}=\boldsymbol{y}_\mu$, and $\lim_{n\to\infty}\Lambda_1(\boldsymbol{x}^{(n)})=\mu$.

This nonlinear iterative method of Theorem 2 can also be given a simple geometric interpretation. Starting with $\boldsymbol{x}_0\in P_1$ such that $\mu<\Lambda_1(\boldsymbol{x}_0)<\sigma$, we now

cyclically treat each component x_j separately, $2 \leqq j \leqq n$. Evidently, decreasing x_j, $2 \leqq j \leqq n$, decreases $\Lambda_1(\boldsymbol{x})$, but increases $\Lambda_j(\boldsymbol{x})$. Thus, the iterative method of Theorem 2 in essence selects a new component x_j such that the disks $|z| \leqq \Lambda_1(\boldsymbol{x})$ and $|q_{k,k} - z| \leqq \Lambda_j(\boldsymbol{x})$ are now tangent. Since the radii $\Lambda_k(\boldsymbol{x})$ of the remaining disks $k \neq j$, are made smaller in the process, we can then apply this procedure to other components.

As a final comment in this section, we point out that the procedure of WILKINSON [12] in essence is the first step of the above Gauss-Seidel procedure for very diagonally dominant matrices Q. He shows in these cases that an excellent approximation to the solution $x_2^{(1)}$ of (4.3) is obtained simply by dropping the first term of (5.3) and solving the resultant linear equation for $x_2^{(1)}$. For such very diagonally dominant matrices, it is also clear that standard linear methods, such as the Gauss-Seidel method, should be rapidly convergent when used in conjunction with the iterative procedure of (3.1). In fact, just one iteration is often sufficient to give good estimates of μ in such cases [12]. In any event, since the vector iterates of these convergent algorithms are elements of the set P_1, rigorous bounds for the first isolated eigenvalue can be obtained at any step in the iteration.

§ 5. The Set of Matrices $\mathring{\Omega}_A$

In this section, we again assume that the $n \times n$ matrix $A = (a_{i,j})$ is an irreducible matrix which admits a first isolated Gerschgorin disk. We further assume, without loss of generality, that the entry $a_{1,1}$ of A is zero. Otherwise, we could consider the new matrix $A - a_{1,1} I$. In analogy with [10], let $\mathring{\Omega}_A$ be the set of all $n \times n$ matrices $B = (b_{i,j})$ such that $|b_{i,j}| = |a_{i,j}|$ for all $1 \leqq i, j \leqq n$. Then, the $n \times n$ matrix Q of (2.2), derived from A, is the same for all $B \in \mathring{\Omega}_A$. Hence, each matrix $B \in \mathring{\Omega}_A$ has an eigenvalue in the disk $0 \leqq |z| \leqq \mu$, where $\mu = \inf_{\boldsymbol{x} \in P_1} \Lambda_1(\boldsymbol{x})$. Let $s(\mathring{\Omega}_A)$ be the set of all complex numbers z with $0 \leqq |z| \leqq \mu$ such that z is an eigenvalue of some $B \in \mathring{\Omega}_A$. In other words, $s(\mathring{\Omega}_A)$ is the spectrum of the set $\mathring{\Omega}_A$ restricted to the disk $0 \leqq |z| \leqq \mu$. Our object now is to determine precisely $s(\mathring{\Omega}_A)$.

Lemma 7. $s(\mathring{\Omega}_A)$ is a closed, bounded, and connected set. Moreover, if $z \in s(\mathring{\Omega}_A)$, then $z \exp i\varphi \in s(\mathring{\Omega}_A)$ for any real φ.

Proof. That $s(\mathring{\Omega}_A)$ is closed and bounded is obvious from the definitions above. To show that $s(\mathring{\Omega}_A)$ is connected, let $\lambda_0, \lambda_1 \in s(\mathring{\Omega}_A)$, where $B_l \boldsymbol{z}_l = \lambda_l \boldsymbol{z}_l$, $B_l \in \mathring{\Omega}_A$, $l = 0, 1$. Writing $B_l = (|a_{j,k}| \exp i\vartheta_{j,k}^{(l)})$, then the $n \times n$ matrix T_ν, defined by

$$T_\nu = \left(|a_{j,k}| \exp i\{(1-\nu)\vartheta_{j,k}^{(0)} + \nu\vartheta_{j,k}^{(1)}\}\right), \qquad 0 \leqq \nu \leqq 1, \tag{5.1}$$

is surely in the set $\mathring{\Omega}_A$, and $T_0 = B_0$, $T_1 = B_1$. Since the entries of T_ν vary continuously with ν, then T_ν possesses an eigenvalue $t(\nu)$ with $|t(\nu)| \leqq \mu$ which also varies continuously with ν. Since $t(0) = \lambda_0$ and $t(1) = \lambda_1$, then $s(\mathring{\Omega}_A)$ is connected.

Next, suppose $z \in s(\mathring{\Omega}_A)$. Then, there is a $B \in \mathring{\Omega}_A$ such that $B\boldsymbol{x} = z\boldsymbol{x}$, $\boldsymbol{x} \neq \boldsymbol{0}$. But clearly, $\exp(i\varphi)B$ is also in the set $\mathring{\Omega}_A$, and thus $z \exp i\varphi \in s(\mathring{\Omega}_A)$, which completes the proof.

With this lemma, we now prove

Theorem 3. There exists a nonnegative real number τ with $\tau < \mu$ such that $s(\mathring{\Omega}_A) = \{z \mid \tau \leqq |z| \leqq \mu\}$.

Proof. First, we note that the $n \times n$ matrix Q of (2.2) is an element of the set $\mathring{\Omega}_A$. Thus, from Lemmas 6 and 7, $\mu \exp i\varphi \in s(\mathring{\Omega}_A)$ for any real φ. It is now clear that $s(\mathring{\Omega}_A) = \{z \mid \tau \leqq |z| \leqq \mu\}$ for some $\tau \geqq 0$, and it remains to show that $\tau < \mu$. Equivalently, we shall show that there is some $z \in s(\mathring{\Omega}_A)$ with $|z| < \mu$. Starting with the $n \times n$ matrix Q of (2.2), change the sign of the last diagonal entry of Q, thereby forming the new $n \times n$ matrix Q_1. Clearly, $Q_1 \in \mathring{\Omega}_A$. One can then show that the minimal Gerschgorin set (defined in [10]) for Q_1, is disconnected, with one connected component lying in the open disk $|z| < \mu$. From this, it follows [10] that Q_1 has one eigenvalue which satisfies $|z| < \mu$, which completes the proof.

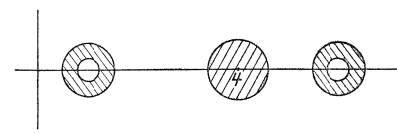

Fig. 2

We remark that the nonnegative quantity τ of Theorem 3 can be positive in some cases and thus $s(\mathring{\Omega}_A)$ is a proper annulus. This is illustrated in § 6.

§ 6. Examples

To illustrate the preceding results for a particularly simple example, consider the following matrix

$$A = \begin{bmatrix} 1 & i/2 & i/2 \\ 1/2 & 4 & i/2 \\ 1/2 & 1/2 & 6 \end{bmatrix}. \tag{6.1}$$

It is readily verified from Definition 1 that the vector $\boldsymbol{\xi} = (1, 1, 1)^T$ is simultaneously in the sets P_1, P_2, and P_3, so that all the eigenvalues of A can be isolated by positive diagonal similarity transformations. Using the results of the previous sections, it can be shown that there is exactly one eigenvalue of A in each of the following annuli:

$$\begin{cases} 0.01584 \leqq |z-1| \leqq 0.1608, \\ 0 \leqq |z-4| \leqq 0.3139, \\ 0.04593 \leqq |z-6| \leqq 0.2301, \end{cases} \tag{6.2}$$

which is schematically indicated in Fig. 2.

The actual eigenvalues of the matrix A of (6.1) are:

$$\lambda_1 = 0.9897 - 0.1243i; \quad \lambda_2 = 4.0121 - 0.06423i, \quad \text{and} \quad \lambda_3 = 5.9982 + 0.1885i.$$

We remark that the iterative method of Theorem 1 is a rapidly convergent method for finding the exterior radii of the annuli of (6.2), for the initial iterate

unity. For the annulus with center $z=1$, the first four iterates $\Lambda_1(\boldsymbol{x}_i)$ of Theorem 1 are 1, 0.2258, 0.1645, and 0.1610. For the cyclic non-linear Gauss-Seidel iterative method of Theorem 2, the corresponding first four iterates $\Lambda_1(\boldsymbol{x}_i)$ are 1, 0.2957, 0.1644, and 0.1609.

Finally, we briefly mention that there is a 6×6 example given[5] in [3], which isolates each of two complex eigenvalues in a disk of radus 0.000287. Because of the increased generality of Theorem 1 of § 3, it is possible to *decrease* this upper bound to 0.0001242. We omit the numerical details.

References

[1] BERS, L.: On mildly nonlinear partial difference equations of elliptic type. J. Research Nat. Bureau of Standards **51**, 229–236 (1953).

[2] GANTMACHER, F. R.: The Theory of Matrices, Vol. Two. New York: Chelsea Publishing Co. 1959.

[3] GERSCHGORIN, S.: Über die Abgrenzung der Eigenwerte einer Matrix. Izv. Akad. Nauk SSSR Ser. Mat. **7**, 749–754 (1931).

[4] HENRICI, P.: Bounds for eigenvalues of certain tridiagonal matrices. J. Soc. Indust. Appl. Math. **11**, 281–290 (1963).

[5] OSTROWSKI, A. M.: Über die Determinanten mit überwiegender Hauptdiagonale. Comment. Math. Helv. **10**, 69–96 (1937).

[6] SCHECHTER, S.: Iteration methods for nonlinear problems. Trans. Amer. Math. Soc. **104**, 179–189 (1963).

[7] TAUSSKY, O.: Bounds for characteristic roots of matrices. Duke Math. J. **15**, 1043–1044 (1948).

[8] — A method for obtaining bounds for characteristic roots of matrices with application to flutter calculations. Aeronautical Research Council of Great Britain, Report **10**, 508 (1947).

[9] VARGA, R. S.: Matrix Iterative Analysis. Englewood Cliffs, New Jersey: Prentice-Hall Inc. 1962.

[10] — Minimal Gerschgorin Sets (to appear in the Pacific J. of Math.).

[11] WIDDER, D.: The Laplace Transform. Princeton: Princeton University Press 1946.

[12] WILKINSON, J. H.: Rigorous error bounds for computed eigensystems. The Computer J. **4**, 230–241 (1961).

Computing Center, Case Institute of Technology
10900 Euclid Avenue, Cleveland, Ohio 44106, USA

(Received May 15, 1964)

[5] The numerical results of [3] are corrected in the J. Soc. Indust. Appl. Math. **12** (1964); pp. 497.