GERSCHGORIN THEORY FOR THE
GENERALIZED EIGENVALUE PROBLEM
$Ax = \lambda Bx$

G. W. Stewart[*]

Department of Mathematics
Department of Computer Science
Carnegie-Mellon University
Pittsburgh, Pa.

October, 1973

ABSTRACT

A generalization of Gerschgorin's theorem is developed for the eigen-
value problem $Ax = \lambda Bx$ and is applied to obtain perturbation bounds for mul-
tiple eigenvalues. The results are interpreted in terms of the chordal metric
on the Riemann sphere, which is especially convenient for treating infinite
eigenvalues.

1. Introduction

The object of this paper is to develop a perturbation theory for the generalized eigenvalue problem $Ax = \lambda Bx$ that parallels the perturbation theory developed by Wilkinson [4, Ch. 2] for the ordinary eigenvalue problem. The theory rests on a generalization of the Gerschgorin theorem, whose results are interpreted in terms of the chordal metric in the Riemann sphere. This approach has the advantage that it deals readily with multiple and infinite eigenvalues.

Throughout this paper we shall identify the (possibly infinite) eigenvalue $\lambda = \alpha/\beta$ with the point in the projective complex line defined by

$$[\alpha,\beta] = \{(\alpha,\beta) \neq (0,0) : \alpha/\beta = \lambda\}.$$

We shall equip the projective complex line with the metric χ defined by

$$\chi([\alpha,\beta],[\alpha',\beta']) = \frac{|\alpha\beta'-\alpha'\beta|}{\sqrt{|\alpha|^2+|\beta|^2}\sqrt{|\alpha'|^2+|\beta'|^2}}$$

The number $\chi([\alpha,\beta],[\alpha',\beta'])$ is the chordal distance between the two points $\lambda = \alpha/\beta$ and $\lambda' = \alpha'/\beta'$ when they are projected in the usual way onto the Riemann sphere. By abuse of notation we shall let $\chi(\lambda,\lambda')$ denote this distance, so that in this context χ is the chordal metric for the set of complex numbers [1, p. 81].

The justification of the use of the chordal metric lies in the simplicity of the final results, in particular in the uniform treatment of large and small eigenvalues. To illustrate this, we shall begin with an application to first order perturbation theory for a simple eigenvalue.

Let A, B, E, and F be square n×n matrices with complex elements and let λ be a simple eigenvalue of the problem $Ax = \lambda Bx$ with eigenvector x normalized so that $||x||_2 = 1$. Let y be the left eigenvector corresponding to λ, again normalized so that $||y||_2 = 0$. Let

$$\alpha = y^H Ax, \quad \beta = y^H Bx,$$

so that the point $[\alpha, \beta]$ is identified with λ. It has been shown by the author [3] that for sufficiently small E, and F, there is an eigenvalue λ' satisfying $(A+E)x' = \lambda'(B+F)x$ that can be identified with

$$[\alpha + y^H Ex, \quad \beta + y^H Fx],$$

except for terms of order $||E||^2$ and $||F||^2$. Thus in our approach the sensitivity of λ to perturbations in A and B will be measured by

$$\chi(\lambda, \lambda') \stackrel{\sim}{=} \chi([\alpha, \beta], \ [\alpha + y^H Ex, \ \beta + y^H Fx])$$

$$= \frac{|\alpha y^H Fx - \beta y^H Ex|}{\sqrt{|\alpha|^2 + |\beta|^2} \ \sqrt{|\alpha + y^H Ex|^2 + |\beta + y^H Fx|^2}}$$

To obtain meaningful approximate perturbation bounds, let

$$\theta = \tan^{-1}|\lambda|,$$
$$\nu = \sqrt{|\alpha|^2 + |\beta|^2} \ ,$$

and

$$\epsilon = \sqrt{\left|y^H Ex\right|^2 + \left|y^H Fx\right|^2}$$

Then $\alpha/\nu = \sin\theta$ and $\beta/\nu = \cos\theta$. Hence we have the bound

(1.1)
$$\chi(\lambda, \lambda') \stackrel{<}{\sim} \frac{\cos\theta|y^H Ex| + \sin\theta|y^H Fx|}{\nu - \epsilon},$$

which is accurate up to terms of order ϵ^2.

In the terminology of numerical analysis, the bound (1.1) says that ν^{-1} is a condition number for the eigenvalue λ, in the sense that it measures how perturbations in A and B will affect λ. If ν is small compared with E and F, then one can expect large changes in λ. It should be noted that ill-conditioned eigenvalues need not be large, and conversely a large eigenvalue need not be ill-conditioned. For example, the bound (1.1) can be quite small even when $\beta = 0$ and hence $\lambda = \infty$, which illustrates the utility of the chordal metric in dealing with this problem.

The factors $\cos\theta$ and $\sin\theta$ appearing in (1.1) are a little unusual, but they make sense. For example when $\theta = 0$, the eigenvalue λ is zero and the matrix A is singular. The disappearance of the term $\sin\theta|y^H Fx|$ in (1.1) then says that perturbations in B cannot affect the singularity of A.

Although the above bound is quite satisfactory for practical work, it is interesting to relate ν to the condition number for the ordinary eigenvalue problem. This may be done as follows. Suppose $Bx \neq 0$ (if $Bx = 0$, use Ax in what follows). Define

$$\kappa = \frac{||Bx||_2}{|\beta|} \equiv \sec \angle(y, Bx)$$

and

$$\rho = \sqrt{||Ax||_2^2 + ||Bx||_2^2} .$$

Then

$$\nu^{-1} = \frac{\kappa}{\rho} .$$

When $B = I$, the number κ is the secant of the angle between the left and right eigenvectors corresponding to λ, which is the condition number for the ordinary

eigenvalue problem [4, Ch. 2]. The number ρ measures how nearly x is an approximate null vector of both A and B (when $Ax = Bx = 0$, $\det(A-\lambda B) \equiv 0$ and any number λ is an eigenvalue). When $B = I$ and $|\lambda| \leq 1$, $\rho \leq \sqrt{2}$ and the bound becomes essentially the bound for the ordinary eigenvalue problem. If $|\lambda| \geq 1$, then the bound (1.1) deviates from the usual bound because of the distorting effects of the chordal metric outside the unit circle.

The above theory has two drawbacks. First, although the bound is asymptotically accurate, the theory does not provide a bound on the remainder, nor does it specify where it is applicable. Second, the theory is not applicable to multiple eigenvalues. It is the object of the next two sections to remedy these defects. In Section 2 we shall develop a generalization of Gerschgorin's theorem and in Section 3 apply it to develop a perturbation theory for multiple eigenvalues.

Throughout the paper we shall use Householder's notational conventions [2]. We shall use the symbols $||\cdot||_p$ ($p = 1,2,\infty$) to denote the usual Hölder vector norms.

2. Gerschgorin Theory

Let A and B be matrices of order n. Set

$$a_i^H = (\alpha_{i1}, \ldots, \alpha_{i,i-1}, \alpha_{i,i+1}, \ldots \alpha_{i,n});$$

that is a_i^H is the vector formed from the i-th row of A by deleting its i-th component. Define the vectors b_i similarly. The following theorem generalizes the Gerschgorin exclusion theorem [2, p. 65 ff].

Theorem 2.1. Let λ be an eigenvalue of the problem $Ax = \lambda Bx$. Then λ lies in the union of the regions $\mathcal{D}_i$ defined by

$$\mathcal{D}_i = \{[\alpha_{ii} + a_i^H \tilde{x}, \beta_{ii} + b_i^H \tilde{x}]: ||\tilde{x}||_\infty \le 1\},$$

$$(i = 1, 2, \ldots, n)$$

Proof: Let $Ax = \lambda Bx$ and suppose that the i-th component of x is largest in absolute value. Since $x \ne 0$, we may assume without loss of generality that $|\xi_i| = 1$. Form $\tilde{x}$ from x by deleting its i-th component. Then $||\tilde{x}|| \le 1$ and

$$\alpha_{ii} + a_i^H \tilde{x} = \lambda(\beta_{ii} + b_i^H \tilde{x}),$$

which says that $\lambda \in \mathcal{D}_i$. ∎

In stating Theorem 2.1 we have used the usual identification of the complex projective line with the Riemann sphere. The proof of the theorem in addition exhibits the region in which λ must lie; namely the region corresponding to a maximal component at the eigenvector. When $B = I$, the regions $\mathcal{D}_i$ become the disks in the complex plane defined by

$$\mathscr{G}_i = \{\lambda: \ |\lambda - \alpha_{ii}| \le ||a_i||_1\},$$

which are the usual Gerschgorin disks.

As they are defined, the sets $\mathscr{G}_i$ are not convenient to work with. Instead we shall work with neighborhoods, defined in terms of the chordal metric, that are generally larger. Specifically, we have

$$\chi([\alpha_{ii}, \beta_{ii}], \ [\alpha_{ii} + a_i^H \tilde{x}, \ \beta_{ii} + b_i^H \tilde{x}])$$

$$= \frac{|\alpha_{ii} b_i^H - \beta_{ii} a_i^H \tilde{x}|}{\sqrt{|\alpha_{ii}|^2 + |\beta_{ii}|^2} \ \sqrt{|\alpha_{ii} + a_i^H \tilde{x}|^2 + |\beta_{ii} + b_i^H \tilde{x}|^2}}$$

$$\le \frac{||a_{ii} b_i^H - \beta_{ii} a_i^H||_1}{\sqrt{|\alpha_{ii}|^2 + |\beta_{ii}|^2} \ \sqrt{\alpha_{ii}'^2 + \beta_{ii}'^2}}$$

$$\equiv \rho_i,$$

where

$$\alpha_{ii}' = \max\{0, \ |\alpha_{ii}| - ||a_i||_1\}$$

and

$$\beta_{ii}' = \max\{0, \ |\beta_{ii}| - ||b_i||_1\}.$$

Hence if we set

$$\mathscr{G}_i' = \{\lambda: \ \chi(\frac{\alpha_i}{\beta_i}, \lambda) \le \rho_i\},$$

the eigenvalues of $Ax = \lambda Bx$ lie in the union of the $\mathscr{G}_i'$. Note that if $B = I$ and the elements of A are less than unity, then the $\mathscr{G}_i'$ give a fair approximation of the usual Gerschgorin disks.

The utility of the Gerschgorin theorem in its applications to the eigenvalue problem is enhanced by the fact that we can often localize a specific

number of eigenvalues in a union of some of the $\mathscr{G}_i$. The same is true of our generalization.

Theorem 2.2. If the union of k at the Gerschgorin regions is disjoint from the remaining regions and is not the entire sphere, then exactly k eigenvalues lie in the union.

Proof. Suppose, without loss of generality, that the regions in question are $\mathscr{G}_1, \mathscr{G}_2, \ldots, \mathscr{G}_k$. Write

$$A = A_1 + A_2$$

where $A_1 = \text{diag}(\alpha_{11}, \alpha_{22}, \ldots, \alpha_{nn})$ and similarly write

$$B = B_1 + B_2$$

where $B_1 = \text{diag}(\beta_{11}, \beta_{22}, \ldots, \beta_{nn})$. For $0 \leq \epsilon \leq 1$ set

$$A(\epsilon) = A_1 + \epsilon A_2$$

and

$$B(\epsilon) = B_1 + \epsilon B_2,$$

and let $\mathscr{G}_i(\epsilon)$ be the corresponding Gerschgorin regions, which of course satisfy $\mathscr{G}_i(\epsilon) \subset \mathscr{G}_i$.

Now the hypotheses of the theorem insure that the characteristic polynomial $\rho_\epsilon(\lambda) = \det(A(\epsilon) - \lambda B(\epsilon))$ cannot vanish identically; for otherwise any number λ would be an eigenvalue and $\bigcup_{i=1}^{n} \mathscr{G}_i$ would be the entire sphere. This means that the eigenvalues of $A(\epsilon)x = \lambda B(\epsilon)x$ must vary continuously in the sphere, or, what is equivalent, there are continuous functions $\lambda_i(\epsilon)$ $(i = 1, 2, \ldots, n)$ such that $\lambda_1(\epsilon), \lambda_2(\epsilon), \ldots, \lambda_n(\epsilon)$ comprise all the

eigenvalues of $A(\varepsilon)x = \lambda B(\varepsilon)x$. We may choose $\lambda_i(0) = \alpha_{ii}/\beta_{ii}$. Then since

$\lambda_i(0) \in \bigcup_{j=1}^{k} \mathscr{G}_j(0)$ $(i = 1,2,\ldots,k)$ and $\bigcup_{j=1}^{k} \mathscr{G}_j(\varepsilon) \cap \bigcup_{j=k+1}^{n} \mathscr{G}_j(\varepsilon) \neq \emptyset$, we must

have $\lambda_i(1) \in \bigcup_{j=1}^{k} \mathscr{G}_j(1)$ for $(i = 1,2,\ldots,k)$ and $\lambda_i(1) \in \bigcup_{j=k+1}^{n} \mathscr{G}_j(1)$ for

$(i = k+1,\ldots,n)$. ∎

3. Perturbation Theory

In [4] Wilkinson has applied the Gerschgorin theorem to produce a perturbation theory for the eigenvalue problem. The heart of his theory is a technique in which off diagonal elements of order ϵ are reduced to order ϵ^2 by diagonal similarity transformations. In this section we shall show that the same technique can be applied to the generalized eigenvalue problem. Since it is the technique rather than its specific applications that is of chief interest, we shall confine ourselves to one of the several cases treated by Wilkinson; the case where the problem has a complete set of eigenvectors.

Specifically we shall be concerned with the case where there are non-singular matrices X and Y such that

$$Y^H AX = \text{diag}(\alpha_1, \alpha_2, \ldots, \alpha_n)$$

and

$$Y^H BX = \text{diag}(\beta_1, \beta_2, \ldots, \beta_n).$$

For definiteness we shall take X and Y to have columns of 2-norm unity. Of course the eigenvalues of $Ax = \lambda Bx$ are $\lambda_i = \alpha_i / \beta_i$ (i = 1,2,...,n). We shall suppose that the first p eigenvalues are equal and distinct from the remaining $q = n-p$ eigenvalues, and we shall apply the Gerschgorin theory of the last section to obtain perturbation bounds for this set of multiple eigenvalues.

Let E and F be matrices of order n and let ϵ be an upper bound on the elements of the matrices $Y^H EX$ and $Y^H FX$. Then $Y^H(A+E)X$ and $Y^H(B+F)X$ have the generic forms (n = 5)

$$\begin{pmatrix} \alpha_1+\varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \alpha_2+\varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \alpha_3+\varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & \alpha_4+\varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & \varepsilon & \alpha_5+\varepsilon \end{pmatrix} \quad , \quad \begin{pmatrix} \beta_1+\varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \beta_2+\varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \beta_3+\varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & \beta_4+\varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & \varepsilon & \beta_5+\varepsilon \end{pmatrix} \quad .$$

If ε is small enough, the first p Gerschgorin regions will be disjoint from the others, and their union will contain exactly p eigenvalues which are necessarily near λ_1. However we can obtain an even better result.

Let $\tau \neq 0$ be given. The eigenvalues of $(A+E)X = \lambda(B+F)X$ will be unchanged if the last q columns of $Y^H(A+E)X$ and $Y^H(B+F)X$ are multiplied by τ. If this is done the new matrice $Y^H(A+E)X$ has the form (n=5, p=2)

$$\begin{pmatrix} \alpha_1+\varepsilon & \varepsilon & \tau\varepsilon & \tau\varepsilon & \tau\varepsilon \\ \varepsilon & \alpha_2+\varepsilon & \tau\varepsilon & \tau\varepsilon & \tau\varepsilon \\ \varepsilon & \varepsilon & \tau(\alpha_3+\varepsilon) & \tau\varepsilon & \tau\varepsilon \\ \varepsilon & \varepsilon & \tau\varepsilon & \tau(\alpha_4+\varepsilon) & \tau\varepsilon \\ \varepsilon & \varepsilon & \tau\varepsilon & \tau\varepsilon & \tau(\alpha_5+\varepsilon) \end{pmatrix}$$

and $Y^H(B+F)X$ has a similar form. We shall attempt to choose $\tau \leq 1$ so that the first p Gerschgorin regions are disjoint.

Let

$$\delta = \min\{\chi(\lambda_i,\lambda_j): i = 1,2,\ldots,p; \ j = p+1,\ldots,n\}.$$

Let

$$\nu_i = \sqrt{|\alpha_i|^2+|\beta_i|^2} \qquad (i = 1,2,\ldots,n)$$

and

$$\nu = \min\{\nu_i: i = 1,2,\ldots,n\}.$$

Then the first p Gerschgorin regions are contained in the disk whose center is λ_1 and whose chordal radius is

$$\frac{|\alpha_i| + |\beta_i|}{\nu_i} \frac{(p+\tau q)\epsilon}{\nu_i - \sqrt{2}(p+\tau q)\epsilon} \leq \frac{\sqrt{2}(p+\tau q)\epsilon}{\nu - \sqrt{2}(p+\tau q)\epsilon} \; .$$

Likewise the remaining q regions will be contained in disks whose centers are λ_i and whose chordal radius is

$$\frac{\sqrt{2}(p+\tau q)\epsilon}{\tau\nu - \sqrt{2}(p+\tau q)\epsilon} \; .$$

Thus if we can find τ satisfying $0 < \tau < 1$ such that

$$\frac{\sqrt{2}(p+\tau q)\epsilon}{\tau\nu - \sqrt{2}(p+\tau q)\epsilon} \leq \frac{\delta}{2} \; ,$$

the appropriate Gerschgorin regions will be disjoint. This will surely be true if

$$(3.1) \qquad\qquad \tau =_{df} \frac{\sqrt{2}\,n\epsilon}{\nu} \frac{(2+\delta)}{\delta} < 1.$$

Thus we have shown that if (3.1) is satisfied, there are exactly p eigenvalues in the region defined by

$$\chi(\lambda_1, \lambda) \leq \frac{\sqrt{2}(p+\tau q)\epsilon}{\nu - \sqrt{2}(p+\tau q)\epsilon},$$

where $\bar\nu = \min\{\nu_i: i = 1, 2, \ldots, p\}$. For small ϵ this radius is asymptotic to

$$\frac{\sqrt{2}\,p\epsilon}{\bar\nu},$$

which shows that $\sqrt{2}p/\bar\nu$ is a condition number for the multiple eigenvalue λ_1.

When λ_1 is simple and (3.1) is satisfied, then the perturbed eigenvalue must lie in the disk defined by

$$X\left(\frac{\alpha_1+y_1^H E x_1}{\beta_1+y_1^H F x_1}, \ \lambda\right) \leq \frac{(|\alpha_1|+|\beta_1|+2\varepsilon)\tau q\varepsilon}{\nu_1[\nu_1-\sqrt{2}(1+\tau q)\varepsilon]}.$$

This bound approaches zero quadratically with ε, which makes rigorous the observations of Section 1.

The treatment sketched above of course does not exhaust all possible cases. Various defective cases can be treated by applying the Gerschgorin theorem to canonical forms, as has been done by Wilkinson for the ordinary eigenvalue problem. Alternatively one can use the techniques of [3] to split off a set of multiple eigenvalues and treat these separately by means of the Gerschgorin theorem. With either approach, the generalized Gerschgorin theorem is required to deal with multiple eigenvalues, and the use of the chordal metric simplifies the final results.

4. Acknowledgments

Part of this work was done while I was a visitor at the IBM Thomas J. Watson Research Center.   I should also like to thank W. Kahan for his comments.

References

1. C. Carathéodory, Theory of Functions, v. I, Chelsea, 1958, New York.

2. A. S. Householder, The Theory of Matrices in Numerical Analysis, Blaisdell, 1964, New York.

3. G. W. Stewart, Error and perturbation bounds for subspaces associated with certain eigenvalue problems, to appear SIAM Review.

4. J. H. Wilkinson, The Algebraic Eigenvalue Problem, Clarendon, 1965, Oxford.