The Algebraic Eigenproblem

Variational principles

If x_j is an eigenvector corresponding to eigenvalue λ_j of symmetric matrix A, then $\lambda_j = x_j^T A x_j / x_j^T x_j$. The rational function

$$\lambda(x) = \frac{x^T A x}{x^T B x} \tag{6.223}$$

where $A = A^T$, and where B is positive definite and symmetric is *Rayleigh's quotient.* Apart from the obvious $\lambda(x_j) = \lambda_j$, Rayleigh's quotient has remarkable properties that we shall discuss here for the special, but not too restrictive, case $B = I$.

Theorem (Rayleigh) 6.50. *Let the eigenvalues of $A = A^T$ be arranged in the ascending order $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$, with orthogonal eigenvectors $x_1, x_2, \ldots, x_n$. Then*

$$\lambda_{k+1} \le \frac{x^T A x}{x^T x} \le \lambda_n \ \text{if} \ x^T x_1 = x^T x_2 = \cdots = x^T x_k = 0, \ x \ne o \tag{6.224}$$

with the lower equality holding if and only if $x = x_{k+1}$, and the upper inequality holding if and only if $x = x_n$. Also

$$\lambda_1 \le \frac{x^T A x}{x^T x} \le \lambda_{n-k} \ \text{if} \ x^T x_n = x^T x_{n-1} = \cdots = x^T x_{n-k+1} = 0, \ x \ne o \tag{6.225}$$

with the lower equality holding if and only if $x = x_1$, and the upper if and only if $x = x_{n-k}$.

The two inequalities reduce to

$$\lambda_1 \le \frac{x^T A x}{x^T x} \le \lambda_n \tag{6.226}$$

for arbitrary $x \in R^n$.

Proof. Vector $x \in R^n$, orthogonal to $x_1, x_2, \ldots, x_k$ has the unique expansion

$$x = \alpha_{k+1} x_{k+1} + \alpha_{k+2} x_{k+2} + \cdots + \alpha_n x_n \tag{6.227}$$

with which

$$x^T A x = \lambda_{k+1}\alpha_{k+1}^2 + \lambda_{k+2}\alpha_{k+2}^2 + \cdots + \lambda_n \alpha_n^2. \tag{6.228}$$

We normalize x by

$$x^T x = \alpha_{k+1}^2 + \alpha_{k+2}^2 + \cdots + \alpha_n^2 = 1 \tag{6.229}$$

and use this equation to eliminate α_{k+1}^2 from $x^T A x$ so as to have

$$\lambda(x) = x^T A x = \lambda_{k+1} + \alpha_{k+2}^2(\lambda_{k+2} - \lambda_{k+1}) + \cdots + \alpha_n^2(\lambda_n - \lambda_{k+1}). \tag{6.230}$$

By assumption $\lambda_j - \lambda_{k+1} \geq 0$ if $j > k+1$ and hence

$$\lambda(x) = \lambda_{k+1} + \quad \text{non-negative quantity} \tag{6.231}$$

or $\lambda(x) \geq \lambda_{k+1}$, with equality holding if and only if

$$\alpha_{k+2}^2(\lambda_{k+2} - \lambda_{k+1}) + \cdots + \alpha_n^2(\lambda_n - \lambda_{k+1}) = 0. \tag{6.232}$$

In case of distinct eigenvalues, $\lambda_j - \lambda_{k+1} \neq 0 \ j = k+2, \ldots, n$, equality holds if and only if $\alpha_{k+2} = \alpha_{k+3} = \cdots = \alpha_n = 0$, and $\lambda(x_{k+1}) = \lambda_{k+1}$. If eigenvalues repeat and $\lambda_j - \lambda_{k+1} = 0$, then α_j need not be zero, but equality still holds if and only if x is in the invariant subspace spanned by the eigenvectors of λ_{k+1}.

To prove the upper bound we use

$$\alpha_n^2 = 1 - \alpha_{k+1}^2 - \alpha_{k+2}^2 - \cdots - \alpha_{n-1}^2 \tag{6.233}$$

to eliminate it from $\lambda(x)$, so as to be left with

$$\lambda(x) = \lambda_n - \alpha_{k+1}^2(\lambda_n - \lambda_{k+1}) - \cdots - \alpha_{n-1}^2(\lambda_n - \lambda_{n-1}) \tag{6.234}$$

and $\lambda(x) \leq \lambda_n$ with equality holding if and only if $x = x_n$.

The proof to the second part of the theorem is the same. End of proof.

Corollary 6.51. *If $A = A^T$, then the $(k+1)$th and $(n-k)$th eigenvalues of A are variationally given by*

$$\lambda_{k+1} = \min_{x \neq o} \lambda(x), \ x^T x_1 = x^T x_2 = \cdots = x^T x_k = 0 \tag{6.235}$$

$$\lambda_{n-k} = \max_{x \neq o} \lambda(x),\ \ x^T x_n = x^T x_{n-1} = \cdots = x^T x_{n-k+1} = 0.$$

The two extremum statements reduce to

$$\lambda_1 = \min_{x \neq o} \lambda(x), \quad \lambda_n = \max_{x \neq o} \lambda(x) \tag{6.236}$$

for arbitrary $x \in R^n$.

Proof. This is an immediate consequence of the previous theorem. If λ_j is isolated, then the minimizing (maximizing) element of $\lambda(x)$ is unique, but if λ_j repeats, then the minimizing (maximizing) element of $\lambda(x)$ is any vector in the invariant subspace corresponding to λ_j. End of proof.

Minimization of $\lambda(x)$ may be subject to the k linear constraints $x^T p_1 = x^T p_2 = \cdots = x^T p_k = 0$, where $p_1, p_2, \ldots, p_k$ are any k constant vectors in R^n. Because of the constraints the minimum of $\lambda(x)$ is *raised*, and the maximum of $\lambda(x)$ is *lowered*. The question is by how much.

Theorem (Fischer) 6.52. *If $A = A^T$, then*

$$\left.\begin{array}{l} \displaystyle\min_{x \neq o} \frac{x^T A x}{x^T x} \leq \lambda_{k+1} \\ \displaystyle\max_{x \neq o} \frac{x^T A x}{x^T x} \geq \lambda_{n-k} \end{array}\right. \qquad x^T p_1 = x^T p_2 = \cdots = x^T p_k = 0. \tag{6.237}$$

Proof. We order the eigenvalues of A in the ascending order $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$, with the corresponding orthogonal eigenvectors $x_1, x_2, \ldots, x_n$. Any vector $x \in R^n$ is uniquely expanded in the form $x = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n$. We shall start with the one constraint $x^T p_1 = 0$ that in terms of $\alpha_1, \alpha_2, \ldots, \alpha_n$ is

$$0 = \alpha_1 x_1^T p_1 + \alpha_2 x_2^T p_1 + \cdots + \alpha_n x_n^T p_1. \tag{6.238}$$

This is one homogeneous equation in the n unknowns $\alpha_1, \alpha_2, \ldots, \alpha_n$ and possesses a nontrivial solution. We may even set $\alpha_3 = \alpha_4 = \cdots = \alpha_n = 0$ and still be left with $\alpha_1 x_1^T p_1 + \alpha_2 x_2^T p_1 = 0$ that has a nontrivial solution. Thus, when $\alpha_3 = \alpha_4 = \cdots = \alpha_n = 0$, $\lambda(x) = (\lambda_1 \alpha_1^2 + \lambda_2 \alpha_2^2)/(\alpha_1^2 + \alpha_2^2)$, by Rayleigh's theorem $\lambda(x) \leq \lambda_2$, and obviously $\min \lambda(x) \leq \lambda_2$.

On the other hand if we choose $\alpha_1 = \alpha_2 = \cdots = \alpha_{n-2} = 0$, then we are left with the constraint equation $\alpha_{n-1}x_{n-1}^Tp_1 + \alpha_n x_n^T p_1 = 0$, which we know possesses a nontrivial solution. Now $\lambda(x) = (\lambda_{n-1}\alpha_{n-1}^2 + \lambda_n\alpha_n^2)/(\alpha_{n-1}^2+\alpha_n^2)$, by Rayleigh's theorem $\lambda(x) \geq \lambda_{n-1}$, and obviously $\max \lambda(x) \geq \lambda_{n-1}$.

Extension of the proof to k constraints is straightforward and is left as an exercise. End of proof.

The following *interlace* theorem is the first important consequence of Fischer's theorem.

Theorem 6.53. *Let the eigenvalues of $A = A^T$ be $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ with corresponding eigenvectors $x_1, x_2, \ldots, x_n$. If*

$$\lambda_k' = \min_{x \neq o} \lambda(x), \begin{cases} x^Tx_1 = x^Tx_2 = \cdots = x^Tx_{k-1} = 0 \\ x^Tp_1 = \cdots = x^Tp_m = 0 \end{cases}, \quad 1 \leq k \leq n-m \tag{6.239}$$

then

$$\lambda_k \leq \lambda_k' \leq \lambda_{k+m} \tag{6.240}$$

In particular, for $m = 1$

$$\lambda_1 \leq \lambda_1' \leq \lambda_2, \quad \lambda_2 \leq \lambda_2' \leq \lambda_3, \ \cdots, \ \lambda_{n-1} \leq \lambda_n' \leq \lambda_n. \tag{6.241}$$

Proof. The lower bound on λ_k' is a consequence of Rayleigh's theorem, and the upper bound of Fischer's with $k+m-1$ constraints. End of proof.

Theorem (Cauchy) 6.54. *Let $A = A^T$ with eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$, be partitioned as*

$$A = \begin{array}{cc} \begin{array}{cc} n-m & m \end{array} & \\ \begin{bmatrix} A' & C \\ C^T & B \end{bmatrix} & \begin{array}{l} n-m \\ m. \end{array} \end{array} \tag{6.242}$$

If $\lambda_1' \leq \lambda_2' \leq \cdots \leq \lambda_{n-m}'$ are the eigenvalues of A' then

$$\lambda_k \leq \lambda_k' \leq \lambda_{k+m}, \ k = 1, 2, \ldots, n-m. \tag{6.243}$$

Proof. $\min x'^T A'x'/x'^T x', \ x' \in R^{n-m}$ can be interpreted as minimization of x^TAx/x^Tx, $x \in R^n$ under the m constraints $x^Te_{n-m+1} = \cdots = x^Te_n = 0$. Theorem 6.53 then assures the inequalities. End of proof.

Theorem 6.55. *Let x be a unit vector, λ a real scalar variable, and define for $A = A^T$ the residual vector $r(\lambda) = r = Ax - \lambda x$. Then $\lambda = \lambda(x) = x^T Ax$ minimizes $r^T r$.*

Proof. If x happens to be an eigenvector, then $r^T r = 0$ if and only if λ is the corresponding eigenvalue. Otherwise

$$r^T r(\lambda) = r^T r = (x^T A - \lambda x^T)(Ax - \lambda x) = \lambda^2 - 2\lambda x^T Ax + x^T A^2 x. \tag{6.244}$$

The vertex of this parabola is at $\lambda = x^T Ax$ and $\min_{\lambda} r^T r = x^T A^2 x - (x^T Ax)^2$. End of proof.

If x is given as an approximation to an eigenvector, then Rayleigh's quotient $\lambda = \lambda(x)$ is the best approximation, in the sense of $\min r^T r$, to the corresponding eigenvalue. We shall look more closely at this approximation.

Theorem 6.56. *Let λ_j be an eigenvalue of $A = A^T$ with corresponding unit eigenvector x_j. Consider unit vector x as an approximation to x_j and $\lambda = \lambda(x)$ as an approximation to λ_j. Then*

$$|\lambda_j - \lambda| \leq (\lambda_n - \lambda_1) 4 \sin^2 \frac{\phi}{2} \tag{6.245}$$

where ϕ is the angle between x_j and x, and where λ_1 and λ_n are the extreme eigenvalues of A.

Proof. Decompose x into $x = x_j + e$. Since $x^T x = x_j^T x_j = 1,\ e^T e + 2e^T x_j = 0$, and

$$\lambda = (x_j + e)^T A(x_j + e) = \lambda_j + e^T (A - \lambda_j I)e. \tag{6.246}$$

But

$$|e^T (A - \lambda_j I)e| \leq \max_k |\lambda_k - \lambda_j| e^T e \leq (\lambda_n - \lambda_1) e^T e \tag{6.247}$$

and therefore

$$|\lambda_j - \lambda| \leq e^T e(\lambda_n - \lambda_1) \tag{6.248}$$

which with $e^T e = 2(1 - \cos\phi) = 4\sin^2 \frac{\phi}{2}$ establishes the inequality. End of proof.

To see that the factor $\lambda_n - \lambda_1$ in Theorem 6.56 is realistic take $x = x_1 + \epsilon x_n, x^T x = 1 + \epsilon^2$, so as to have

$$\lambda - \lambda_1 = \frac{\epsilon^2}{1+\epsilon^2}(\lambda_n - \lambda_1). \tag{6.249}$$

Theorem 6.56 is theoretical. It tells us that a reasonable approximation to an eigenvector should produce an excellent Rayleigh quotient approximation to the corresponding eigenvalue. To actually know how good the approximation is requires yet a good deal of hard work.

Theorem 6.57. *Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be the eigenvalues of $A = A^T$ with corresponding orthonormal eigenvectors $x_1, x_2, \ldots, x_n$. Given unit vector x and scalar λ, then*

$$\min_j |\lambda_j - \lambda| \leq \|r\| \tag{6.250}$$

if $r = Ax - \lambda x$.

Proof. In terms of the n eigenvectors of A, $x = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n$ so that

$$r = \alpha_1(\lambda_1 - \lambda)x_1 + \alpha_2(\lambda_2 - \lambda)x_2 + \cdots + \alpha_n(\lambda_n - \lambda)x_n. \tag{6.251}$$

Consequently

$$r^T r = \alpha_1^2(\lambda_1 - \lambda)^2 + \alpha_2^2(\lambda_2 - \lambda)^2 + \cdots + \alpha_n^2(\lambda_n - \lambda)^2 \tag{6.252}$$

and

$$r^T r \geq \min_j (\lambda_j - \lambda)^2(\alpha_1^2 + \alpha_2^2 + \cdots + \alpha_n^2). \tag{6.253}$$

Recalling that $\alpha_1^2 + \alpha_2^2 + \cdots + \alpha_n^2 = 1$, and taking the positive square root on both sides yields the inequality. End of proof.

Theorem 6.57 does not refer specifically to $\lambda = \lambda(x)$, but it is reasonable to choose this λ, that we know minimizes $r^T r$. It is of considerable computational interest because of its numerical nature. The theorem states that given λ and $\|r\|$ there is at least one eigenvalue λ_j in the interval $\lambda - \|r\| \leq \lambda_j \leq \lambda + \|r\|$.

At first sight Theorem 6.57 appears disappointing in having a right-hand side that is only $\|r\|$. Theorem 6.56 raises the expectation of a power to $\|r\|$ higher than 1, but as we shall see in the example below, if an eigenvalue repeats, then the bound in Theorem 6.57 is sharp; equality does actually happen with it.

Example. For

$$A = \begin{bmatrix} 1 & \epsilon \\ \epsilon & 1 \end{bmatrix} \quad x_1 = \frac{\sqrt{2}}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \lambda_1 = 1 + \epsilon, \ x_2 = \frac{\sqrt{2}}{2}\begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \lambda_2 = 1 - \epsilon \tag{6.254}$$

we choose $x = [1\ 0]^T$ and obtain $\lambda(x) = 1$, and $r = \epsilon[0\ 1]^T$. The actual error in both λ_1 and λ_2 is ϵ, and also $\|r\| = \epsilon$.

For

$$A = \begin{bmatrix} 1 & \epsilon \\ \epsilon & 2 \end{bmatrix},\ \lambda_1 = 1 - \epsilon^2,\ \lambda_2 = 2 + \epsilon^2,\ \epsilon^2 << 1 \tag{6.255}$$

we choose $x = [1\ 0]^T$ and get $\lambda(x) = 1$, and $r = \epsilon[0\ 1]^T$. Here $\|r\| = \epsilon$, but the actual error in λ_1 is ϵ^2.

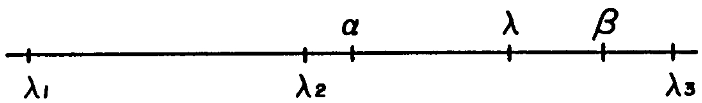

Fig. 6.3

A better inequality can be had, but only at the heavy price in practicality of knowing the eigenvalues' separation. See Fig.6.3 that refers to the following

Theorem (Kato) 6.58. *Let $A = A^T,\ x^Tx = 1,\ \lambda = \lambda(x) = x^TAx$, and suppose that α and β are two real numbers such that $\alpha < \lambda < \beta$ and such that no eigenvalue of A is found in the interval $\alpha \le \lambda \le \beta$.*

Then

$$(\beta - \lambda)(\lambda - \alpha) \le r^Tr = \epsilon^2,\ r = Ax - \lambda x \tag{6.256}$$

and the inequality is sharp.

Proof. Write $x = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \cdots + \alpha_n x_n$ to have

$$\begin{aligned} Ax - \beta x &= (\lambda_1 - \beta)\alpha_1 x_1 + (\lambda_2 - \beta)\alpha_2 x_2 + \cdots + (\lambda_n - \beta)\alpha_n x_n \\ Ax - \alpha x &= (\lambda_1 - \alpha)\alpha_1 x_1 + (\lambda_2 - \alpha)\alpha_2 x_2 + \cdots + (\lambda_n - \alpha)\alpha_n x_n. \end{aligned} \tag{6.257}$$

Then

$$\begin{aligned} (Ax - \beta x)^T(Ax - \alpha x) = (\lambda_1 - \beta)(\lambda_1 - \alpha)\alpha_1^2 + (\lambda_2 - \beta)(\lambda_2 - \alpha)\alpha_2^2 \\ + \cdots + (\lambda_n - \beta)(\lambda_n - \alpha)\alpha_n^2 \ge 0 \end{aligned} \tag{6.258}$$

because $(\lambda_j - \beta)$ and $(\lambda_j - \alpha)$ are either both negative or both positive, or their product is zero.

But

$$\begin{aligned} Ax - \alpha x &= Ax - \lambda x + (\lambda - \alpha)x = r + (\lambda - \alpha)x \\ Ax - \beta x &= Ax - \lambda x + (\lambda - \beta)x = r + (\lambda - \beta)x \end{aligned} \tag{6.259}$$

and therefore

$$(r + (\lambda - \alpha)x)^T(r + (\lambda - \beta)x) \geq 0. \tag{6.260}$$

Since $x^Tr = 0,\ x^Tx = 1$, multiplying out yields

$$r^Tr + (\lambda - \alpha)(\lambda - \beta) \geq 0 \tag{6.261}$$

and the inequality is proved.

To show that equality does occur in Kato's theorem assume that $x = \alpha_1 x_1 + \alpha_2 x_2,\ \alpha_1^2 + \alpha_2^2 = 1$. Then

$$\begin{aligned} &\lambda = \alpha_1^2\lambda_1 + \alpha_2^2\lambda_2,\ \lambda_1 - \lambda = \alpha_2^2(\lambda_1 - \lambda_2), \lambda - \lambda_2 = \alpha_1^2(\lambda_1 - \lambda_2), \\ &\epsilon^2 = \alpha_1^2(\lambda_1 - \lambda)^2 + \alpha_2^2(\lambda_2 - \lambda)^2 = \alpha_1^2\alpha_2^2(\lambda_1 - \lambda_2)^2 \end{aligned} \tag{6.262}$$

and in fact $\epsilon^2 = (\lambda_2 - \lambda)(\lambda - \lambda_1)$. End of proof.

Example. The three eigenvalues of matrix

$$A = \begin{bmatrix} 1 & -1 & \\ -1 & 2 & -1 \\ & -1 & 2 \end{bmatrix} \tag{6.263}$$

are $\lambda_j = 4\sin^2(\theta_j/2),\ \theta_j = (2j-1)\pi/7\ j = 1, 2, 3$, or numerically

$$\lambda_1 = 0.1980623,\ \lambda_2 = 1.5549581,\ \lambda_3 = 3.2469796. \tag{6.264}$$

We take

$$x'_1 = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix},\ x'_2 = \begin{bmatrix} 2 \\ -1 \\ -2 \end{bmatrix},\ x'_3 = \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix} \tag{6.265}$$

as approximations to the three eigenvectors of A and compute the corresponding Rayleigh quotients

$$\lambda'_1 = \frac{3}{14} = 0.2143,\ \lambda'_2 = \frac{14}{9} = 1.5556,\ \lambda'_3 = \frac{29}{9} = 3.2222. \tag{6.266}$$

These are seen to be excellent approximations, and we expect them to be so in view of Theorem 6.56, even with eigenvectors that are only crudely approximated. But we shall not know how good $\lambda_1', \lambda_2', \lambda_3'$ are until the approximations to the eigenvalues are separated.

We write $r_j = Ax_j' - \lambda_j' x_j'$, compute the three relative residuals

$$\epsilon_1 = \frac{\|r_1\|}{\|x_1'\|} = \frac{\sqrt{5}}{14} = 0.1597,\ \epsilon_2 = \frac{\|r_2\|}{\|x_2'\|} = \frac{\sqrt{2}}{9} = 0.1571,\ \epsilon_3 = \frac{\|r_3\|}{\|x_3'\|} = \frac{\sqrt{5}}{9} = 0.2485 \quad (6.267)$$

and have from Theorem 6.57 that

$$0.0546 \le \lambda_1 \le 0.374,\ 1.398 \le \lambda_2 \le 1.713,\ 2.974 \le \lambda_3 \le 3.471. \quad (6.268)$$

Figure 6.4 has the exact $\lambda_1, \lambda_2, \lambda_3$, the approximate $\lambda_1', \lambda_2', \lambda_3'$, and the three intervals marked on it.

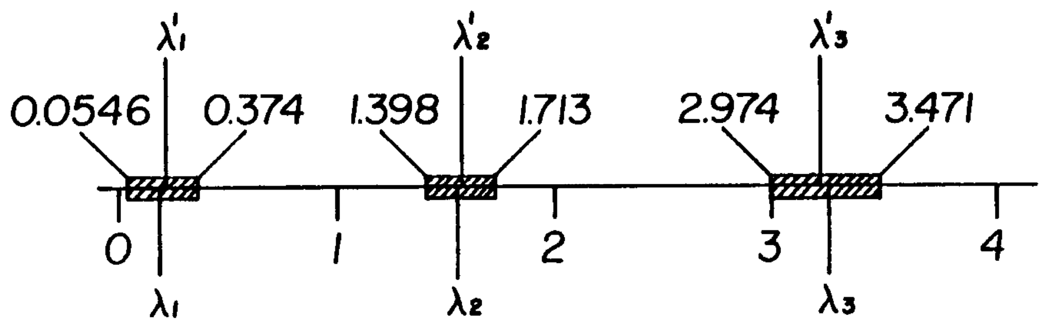

Fig. 6.4

Even if the bounds on $\lambda_1, \lambda_2, \lambda_3$ are not very tight, they at least separate the eigenvalue approximations. Rayleigh and Kato's theorems will help us do much better than this.

Rayleigh's theorem assures us that $\lambda_1 \le \lambda_1'$, and hence we select $\alpha = \lambda_1,\ \beta = 1.398$ in Kato's inequality so as to have

$$(\lambda_1' - \lambda_1)(1.398 - \lambda_1') \le \epsilon_1^2 \quad (6.269)$$

and

$$0.1927 \le \lambda_1 \le 0.2143. \quad (6.270)$$

If $\lambda_2' \le \lambda_2$, then we select $\alpha = \lambda_1',\ \beta = \lambda_2$ in Kato's inequality and obtain

$$\lambda_2' \le \lambda_2 \le \lambda_2' + \frac{\epsilon_2^2}{\lambda_2' - \lambda_1'} \quad (6.271)$$

while if $\lambda_2' \geq \lambda_2$, then we select $\alpha = \lambda_2,\ \beta = \lambda_3'$ in Kato's inequality and

$$\lambda_2' - \frac{\epsilon_2^2}{\lambda_3' - \lambda_2'} \leq \lambda_2 \leq \lambda_2'. \tag{6.272}$$

Hence, wherever the location of λ_2' relative to λ_2

$$\lambda_2' - \frac{\epsilon_2^2}{\lambda_3' - \lambda_2'} \leq \lambda_2 \leq \lambda_2' + \frac{\epsilon_2^2}{\lambda_2' - \lambda_1'} \tag{6.273}$$

or numerically

$$1.5407 \leq \lambda_2 \leq 1.5740. \tag{6.274}$$

The last approximate λ_3' is, by Rayleigh's theorem, less than the exact, $\lambda_3' \leq \lambda_3$, and we select $\alpha = 1.5740,\ \beta = \lambda_3$ in Kato's inequality,

$$(\lambda_3 - \lambda_3')(\lambda_3' - 1.5740) \leq \epsilon_3^2 \tag{6.275}$$

to obtain

$$3.222 \leq \lambda_3 \leq 3.260. \tag{6.276}$$

Now that better approximations to the eigenvalues are available to us, can we use them to improve the approximations to the eigenvectors? Consider λ_1, x_1 and λ_1', x_1'. Assuming the approximations are good we write

$$x_1 = x_1' + dx_1,\ \ \lambda_1 = \lambda_1' + d\lambda_1 \tag{6.277}$$

and, upon neglecting the product $d\lambda_1 dx_1$, obtain

$$(A - \lambda_1 I)x_1 = (A - \lambda_1' I)(x_1' + dx_1) - d\lambda_1 x_1' = o$$

from which the approximation

$$x_1 = d\lambda_1 (A - \lambda_1' I)^{-1} x_1' \tag{6.278}$$

readily results. Factor $d\lambda_1$ is irrelevant, but its smallness is a warning that $(A - \lambda_1' I)^{-1} x_1'$ can be of a considerable magnitude because $(A - \lambda_1 I)$ may well be nearly singular.

The enterprising reader should undertake the numerical correction of x_1', x_2', x_3'.

Now that supposedly better eigenvector approximations are available, they can be used in turn to produce better Rayleigh approximations to the eigenvalues, and the corrective cycle may be repeated, even without recourse to the complicated Rayleigh-Kato bound tightening. This is in fact the essence of the method of shifted inverse iterations, or linear corrections, described in Sec. 8.5.

Error bounds on the eigenvectors are discussed next.

Theorem 6.59. *Let the eigenvalues of $A = A^T$ be $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$, with corresponding orthonormal eigenvectors $x_1, x_2, \ldots, x_n$, and x a unit vector approximating x_j. If $e_j = x - x_j$, and $\lambda = x^T A x$, then*

$$\|e_j\| \le \left(2 - 2\left(1 - \left(\frac{\epsilon_j}{\alpha}\right)^2\right)^{1/2}\right)^{1/2}, \ \frac{\epsilon_j}{\alpha} < 1 \tag{6.279}$$

where $\epsilon_j = \|r_j\|$, $r_j = Ax - \lambda x$, and where

$$\alpha = \min_{k \neq j} |\lambda_k - \lambda|. \tag{6.280}$$

If $|\epsilon_j/\alpha| << 1$, then

$$\|e_j\| \le \frac{\epsilon_j}{\alpha}. \tag{6.281}$$

Proof. Write

$$x = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_j x_j + \cdots + \alpha_n x_n, \ \alpha_1^2 + \alpha_2^2 + \cdots + \alpha_n^2 = 1 \tag{6.282}$$

so as to have

$$e_j = x - x_j = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + (\alpha_j - 1)x_j + \cdots + \alpha_n x_n \tag{6.283}$$

and

$$e_j^T e_j = \alpha_1^2 + \alpha_2^2 + \cdots + (\alpha_j - 1)^2 + \cdots + \alpha_n^2. \tag{6.284}$$

Because $x^T x = 1$

$$e_j^T e_j = 2(1 - \alpha_j) \ , \ \alpha_j = 1 - \frac{1}{2} e_j^T e_j. \tag{6.285}$$

Also,

$$\epsilon_j^2 = r_j^T r_j = \alpha_1^2(\lambda_1 - \lambda)^2 + \alpha_2^2(\lambda_2 - \lambda)^2 + \cdots + \alpha_j^2(\lambda_j - \lambda) + \cdots + \alpha_n^2(\lambda_n - \lambda)^2 \tag{6.286}$$

and

$$\epsilon_j^2 \geq \alpha_1^2(\lambda_1-\lambda)^2+\alpha_2^2(\lambda_2-\lambda)^2+\cdots+0+\cdots+\alpha_n^2(\lambda_n-\lambda)^2. \tag{6.287}$$

Moreover

$$\epsilon_j^2 \geq \min_{k\neq j}(\lambda_k-\lambda)^2(\alpha_1^2+\alpha_2^2+\cdots 0+\cdots+\alpha_n^2) \tag{6.288}$$

or

$$\epsilon_j^2 \geq \alpha^2(1-\alpha_j^2). \tag{6.289}$$

But $\alpha_j = 1-\frac{1}{2}e_j^Te_j$ and therefore

$$(1-\frac{1}{2}e_j^Te_j)^2 \geq 1-\left(\frac{\epsilon_j}{\alpha}\right)^2. \tag{6.290}$$

With the proper sign choice for $x, \frac{1}{2}e_j^Te_j < 1$, and taking the positive square root on both sides yields the first inequality. The simpler inequality comes from

$$\left(1-\left(\frac{\epsilon_j}{\alpha}\right)^2\right)^{\frac{1}{2}} = 1-\frac{1}{2}\left(\frac{\epsilon_j}{\alpha}\right)^2 \tag{6.291}$$

for $(\epsilon_j/\alpha) << 1$. End of proof.

Notice that Theorem 6.57 does not require λ to be x^TAx, but in view of Theorem 6.55 it is reasonable to choose it this way. Notice also that as $x \rightarrow x_j,\ \lambda$ may be replaced with λ_j, and α becomes the least of $\lambda_{j+1}-\lambda_j$ and $\lambda_j-\lambda_{j-1}$. To compute a good bound on $\|x-x_j\|$ we need to know how well λ_j is separated from its left and right neighbors. To see that the bounds are sharp take $x = x_1+\epsilon x_2,\ \epsilon^2 << 1$, so as to get $\|x-x_1\| = \|r\|/(\lambda_2-\lambda_1)$.

Lemma 6.60. *If $x \in R^n$ and $x^Tx = 1$, then*

$$|x_1|^2+|x_2|^2+\cdots+|x_n|^2 = 1 \quad \textit{and} \quad |x_1|+|x_2|+\cdots+|x_n| \leq \sqrt{n}. \tag{6.292}$$

proof. Select vector s with components ± 1 so that $s^Tx = |x_1|+|x_2|+\cdots+|x_n|$. Obviously $\|s\| = \sqrt{n}$. By the Cauchy–Schwarz inequality

$$s^Tx \leq \|s\|\|x\| = \sqrt{n} \tag{6.293}$$

since $\|x\| = 1$, and hence the inequality of the lemma. Equality occurs in eq.(6.292) for vector x with all components equal in magnitude. End of proof.

Theorem (Hirsch) 6.61. *Let matrix $A = A(n \times n)$ have a complex eigenvalue $\lambda = \alpha + i\beta$. Then*

$$|\lambda| \leq n \max_{i,j} |A_{ij}|, \; |\alpha| \leq n \max_{i,j} \frac{1}{2}|A_{ij} + A_{ji}|, \; |\beta| \leq n \max_{i,j} \frac{1}{2}|A_{ij} - A_{ji}|. \tag{6.294}$$

proof. Let x be a unit, $x^H x = 1$, eigenvector corresponding to eigenvalue λ so that $Ax = \lambda x$. Then

$$\lambda = x^H Ax = A_{11}x_1\overline{x}_1 + A_{12}\overline{x}_1 x_2 + A_{21}x_1\overline{x}_2 + \cdots + A_{nn}x_n\overline{x}_n \tag{6.295}$$

and

$$|\lambda| \leq \max_{i,j} |A_{ij}|(|x_1|^2 + 2|x_1||x_2| + \cdots + |x_n|^2) \tag{6.296}$$

or

$$|\lambda| \leq \max_{i,j} |Aij|(|x_1| + |x_2| + \cdots + |x_n|)^2 \tag{6.297}$$

and since $|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2 = 1$, Lemma 6.60 guarantees the first inequality of the theorem. To prove the other two inequalities we write $x = u + iv, u^T u = 1 \; v^T v = 1$, and separate the eigenproblem into

$$Au = \alpha u - \beta v, \quad Av = \alpha v + \beta u \tag{6.298}$$

from which we get through premultiplication by u^T and v^T

$$2\alpha = \frac{1}{2}u^T(A + A^T)u + \frac{1}{2}v^T(A + A^T)v, \quad 2\beta = u^T(A - A^T)v. \tag{6.299}$$

From the second equation we derive the inequality

$$2\beta \leq \max_{i,j} |A_{ij} - A_{ji}|(|u_1||v_1| + |u_1||v_2| + |u_2||v_1| + \cdots + |u_n||v_n|) \tag{6.300}$$

or

$$2\beta \leq \max_{i,j} |A_{ij} - A_{ji}|(|u_1| + |u_2| + \cdots + |u_n|)(|v_1| + |v_2| + \cdots + |v_n|). \tag{6.301}$$

Recalling lemma 6.60 we ascertain the third inequality of the theorem. The second inequality of the theorem is proved likewise. End of proof.

For matrix $A = A(n \times n), A_{ij} = 1$, the estimate $|\lambda| \leq n$ of Theorem 6.61 is sharp; here in fact $\lambda_n = n$. For upper-triangular matrix $U, U_{ij} = 1, \; |\lambda| \leq n$ is a terrible over estimate;

all eigenvalues of U are here only 1. Theorem 6.61 is nevertheless of theoretical interest. It informs us that a matrix with small entries has small eigenvalues, and that a matrix only slightly asymmetric has eigenvalues that are only slightly complex.

We close this section with a monotonicity theorem and an application.

Theorem (Weyl) 6.62. *Let A and B in $C = A + B$ be symmetric. If $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_n$ are the eigenvalues of A, $\beta_1 \leq \beta_2 \leq \cdots \leq \beta_n$ the eigenvalues of B, and $\gamma_1 \leq \gamma_2 \leq \cdots \leq \gamma_n$ the eigenvalues of C, then*

$$\alpha_i + \beta_j \leq \gamma_{i+j-1}, \;\; \gamma_{i+j-n} \leq \alpha_i + \beta_j. \tag{6.302}$$

In particular

$$\alpha_i + \beta_1 \leq \gamma_i \leq \alpha_i + \beta_n. \tag{6.303}$$

Proof. Let $a_1, a_2, \ldots, a_n$ be the n orthonormal eigenvectors of A, and $b_1, b_2, \ldots, b_n$ the orthonormal eigenvectors of B. Obviously

$$\min_{\substack{x \\ x^Ta_1 = \cdots = x^Ta_{i-1} = 0 \\ x^Tb_1 = \cdots = x^Tb_{j-1} = 0.}} \frac{x^TCx}{x^Tx} \geq \min_{\substack{x \\ x^Ta_1 = \cdots = x^Ta_{i-1} = 0}} \frac{x^TAx}{x^Tx} + \min_{\substack{x \\ x^Tb_1 = \cdots = x^Tb_{j-1} = 0}} \frac{x^TBx}{x^Tx} \tag{6.304}$$

By Fischer's theorem the left-hand side of the above inequality does not exceed γ_{i+j-1}, while by Rayleigh's theorem the right-hand side is equal to $\alpha_i + \beta_j$. Hence the first set of inequalities.

The second set of inequalities are obtained from

$$\max_{\substack{x \\ x^Ta_{i+1} = \cdots = x^Ta_n = 0 \\ x^Tb_{j+1} = \cdots = x^Tb_n = 0.}} \frac{x^TCx}{x^Tx} \leq \max_{\substack{x \\ x^Ta_{i+1} = \cdots = x^Ta_n = 0}} \frac{x^TAx}{x^Tx} + \max_{\substack{x \\ x^Tb_{j+1} = \cdots = x^Tb_n = 0}} \frac{x^TBx}{x^Tx} \tag{6.305}$$

By Fischer's theorem the left-hand side of the above inequality is not less than γ_{i+j-n}, while by Rayleigh's theorem the right-hand side is equal to $\alpha_i + \beta_j$.

The particular case is obtained with $j = 1$ on the one hand and $j = n$ on the other hand. End of proof.

Theorem 6.62 places no limit on the size of the eigenvalues but it may be put into a perturbation form. Let positive ϵ be such that $-\epsilon \leq \beta_1, \; \beta_n \leq \epsilon$. Then

$$|\gamma_i - \alpha_i| \leq \epsilon \tag{6.306}$$

and if ϵ is small $|\gamma_i - \alpha_i|$ is smaller. The above inequality together with Theorem 6.61 carry an important implication: *if the entries of symmetric matrix A are symmetrically perturbed slightly, then the change in each eigenvalue is slight.*

One of the more interesting applications of Weyl's theorem is the following. If in the symmetric

$$A = \begin{bmatrix} K & R^T \\ R & M \end{bmatrix} \tag{6.307}$$

matrix $R = O$, then A reduces to block diagonal and the eigenvalues of A become those of K together with those of M. We expect that if matrix R is small, then the eigenvalues of K and M will not be far from the eigenvalues of A, and indeed we have

Corollary 6.63. *If*

$$A = \begin{bmatrix} K & R^T \\ R & M \end{bmatrix} = \begin{bmatrix} K & \\ & M \end{bmatrix} + \begin{bmatrix} & R^T \\ R & \end{bmatrix} = A' + E \tag{6.308}$$

then

$$|\lambda_i - \lambda'_i| \leq |\rho_n| \tag{6.309}$$

where λ_i and λ'_i are the ith eigenvalue of A and A', respectively, and where ρ_n^2 is the largest eigenvalue of R^TR, or RR^T.

Proof. Write

$$\begin{bmatrix} & R^T \\ R & \end{bmatrix} \begin{bmatrix} x \\ x' \end{bmatrix} = \rho \begin{bmatrix} x \\ x' \end{bmatrix}. \tag{6.310}$$

Then $R^TRx = \rho^2 x$ or $RR^Tx' = \rho^2 x'$, provided that $\rho \neq 0$. If ρ_n^2 is the largest eigenvalue of R^TR (or equally RR^T), then the eigenvalues of E are between $-\rho_n$ and $+\rho_n$, and the inequality in the corollary follows from the previous theorem. End of proof.

exercises

6.16.1. Let $A = A^T$. Show that if $Ax - \lambda x = r,\ \lambda = x^T Ax/x^T x$, then $x^T r = 0$. Also, that $Bx = \lambda x$ for

$$B = A - (xr^T + rx^T)/x^T x.$$

6.16.2. Use Fischer's and Rayleigh's theorems to show that

$$\lambda_2 = \max_p(\min_{x \perp p} \lambda(x)),\ \ \lambda_{n-1} = \min_p(\max_{x \perp p} \lambda(x))$$

where $\lambda(x) = x^T Ax/x^T x$.

6.16.3. Let A and B be symmetric positive definite. Show that

$$\lambda_n(AB) \leq \lambda_n(A)\lambda_n(B)$$

and

$$\lambda_1(A+B) \geq \lambda_1(A) + \lambda_1(B)\ \ ,\ \ \lambda_n(A+B) \leq \lambda_n(A) + \lambda_n(B).$$

6.16.4. Show that for square A

$$\alpha_1 \leq \frac{x^T Ax}{x^T x} \leq \alpha_n$$

where α_1 and α_n are the extremal eigenvalues of $\frac{1}{2}(A + A^T)$.

6.16.5. Let $A = A^T$ and $A' = A'^T$ have eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ and $\lambda'_1 \leq \lambda'_2 \leq \cdots \leq \lambda'_n$ such that $\lambda'_j \geq \lambda_j$ for all j. Is it true that $x^T A'x \geq x^T Ax$ for any x? Consider

$$A = \begin{bmatrix} -1 & -1 \\ -1 & -1 \end{bmatrix} \text{ and } A' = \begin{bmatrix} -1.1 & \sqrt{0.88} \\ \sqrt{0.88} & -0.8 \end{bmatrix}.$$

6.16.6. Prove that if for symmetric A' and A, $x^T A'x \geq x^T Ax$ for any x, then pairwise $\lambda_i(A') \geq \lambda_i(A)$.

6.16.7. Let $A = A^T$ be such that $A_{ij} \geq 0$. Show that for any $x, x_i \geq 0$,

$$(x^T Ax)^2 \leq (x^T x)(x^T A^2 x).$$

6.16.8. Use corollary 6.63 to prove that symmetric

$$A = \begin{bmatrix} \alpha & a^T \\ a & A' \end{bmatrix}$$

has an eigenvalue in the interval

$$|\alpha - \lambda| \leq (a^T a)^{1/2}$$

Generalize the bound to other diagonal elements of A using a symmetric interchange of rows and columns.

6.16.9. Let $\sigma_i = (\lambda_i(A^T A))^{1/2}$ be the singular values of $A = A(n \times n)$, and let σ_i' be the singular values of A' obtained from A through the deletion of one row (column). Show that

$$\sigma_i \leq \sigma_i' \leq \sigma_{i+1} \; i = 1, 2, \ldots, n-1.$$

Generalize to more deletions.

6.16.10. Let $\sigma_i = (\lambda_i(A^T A))^{1/2}$ be the singular values of $A = A(n \times n)$. Show, after Weyl, that

$$\sigma_1 \sigma_2 \cdots \sigma_k \leq |\lambda_1||\lambda_2| \cdots |\lambda_k|, \text{ and } \sigma_k \ldots \sigma_{n-1} \sigma_n \geq |\lambda_k| \ldots |\lambda_{n-1}||\lambda_n|, \; k = 1, 2, \ldots, n$$

where $\lambda_k = \lambda_k(A)$ are such that $|\lambda_1| \leq |\lambda_2| \leq \cdots \leq |\lambda_n|$.

6.16.11. Recall that

$$\|A\|_F = (\sum_{i,j} A_{ij}^2)^{1/2}$$

is the Frobenius norm of A. Show that among all symmetric matrices, $S = (A + A^T)/2$ minimizes $\|A - S\|_F$.

6.16.12. Let nonsingular A have the polar decomposition $A = (AA^T)^{1/2} Q$. Show that among all orthogonal matrices, $Q = (AA^T)^{-1/2} A$ is the unique minimizer of $\|A - Q\|_F$. Discuss the case of singular A.

Bounds and perturbations

Computation of even approximate eigenvalues and their accuracy assessment is a serious computational affair and we appreciate any quick procedure for their enclosure. Gerschgorin's theorem on eigenvalue bounds is surprisingly simple, yet general and practical.

Theorem (Gerschgorin) 6.64. *Let $A = A(n \times n)$. If $A = D + A'$, where D is the diagonal $D_{ii} = A_{ii}$, then every eigenvalue of A lies in at least one of the discs*

$$|\lambda - A_{ii}| \leq |A'_{i1}| + |A'_{i2}| + \cdots + |A'_{in}| \quad i = 1, 2, \ldots, n \tag{6.311}$$

in the complex plane.

Proof. Even if A is real its eigenvalues and eigenvectors may be complex. Let λ be any eigenvalue of A and $x = [x_1\ x_2\ \ldots\ x_n]^T$ the corresponding eigenvector so that $Ax = \lambda x\ x \neq o$. Assume that the kth component of x, x_k, is largest in magnitude (modulus) and normalize x so that $|x_k| = 1$ and $|x_i| \leq 1$. The kth equation of $Ax = \lambda x$ then becomes

$$A_{k1}x_1 + A_{k2}x_2 + \cdots + A_{kk} + \cdots + A_{kn}x_n = \lambda \tag{6.312}$$

and

$$\begin{aligned} |\lambda - A_{kk}| &= |A'_{k1}x_1 + A'_{k2}x_2 + \cdots + A'_{kn}x_n| \\ &\leq |A'_{k1}|\ |x_1| + |A'_{k2}|\ |x_2| + \cdots + |A'_{kn}|\ |x_n| \\ &\leq |A'_{k1}| + |A'_{k2}| + \cdots + |A'_{kn}|. \end{aligned} \tag{6.313}$$

We do not know what k is, but we are sure that λ lies in one of these discs. End of proof.

Example. Matrix

$$A = \begin{bmatrix} 2 & -3 & 1 \\ -2 & 1 & 3 \\ 1 & -4 & 2 \end{bmatrix} \tag{6.314}$$

has the characteristic equation

$$-\lambda^3 + 5\lambda^2 - 13\lambda + 14 = 0 \tag{6.315}$$

with the three roots

$$\lambda_1 = \frac{3}{2} + \frac{\sqrt{19}}{2}i,\ \lambda_2 = \overline{\lambda}_1 = \frac{3}{2} - \frac{\sqrt{19}}{2}i,\ \lambda_3 = 2. \tag{6.316}$$

Gerschgorin's theorem encloses the eigenvalues in the three discs

$$\delta_1 : |2-\lambda| \leq 4, \ \delta_2 : |1-\lambda| \leq 5, \ \delta_3 : |2-\lambda| \leq 5 \tag{6.317}$$

shown in Fig. 6.5. Not even a square root is needed to have these bounds.

Corollary 6.65. *If λ is an eigenvalue of symmetric A, then*

$$\min_k (A_{kk} - |A'_{k1}| - \cdots - |A'_{kn}|) \leq \lambda \leq \max_k (A_{kk} + |A'_{k1}| + \cdots + |A'_{kn}|) \tag{6.318}$$

where $A'_{ij} = A_{ij}$ and $A'_{ii} = 0$.

Proof. When A is symmetric λ is real and the Gerschgorin discs become intervals on the real axis. End of proof.

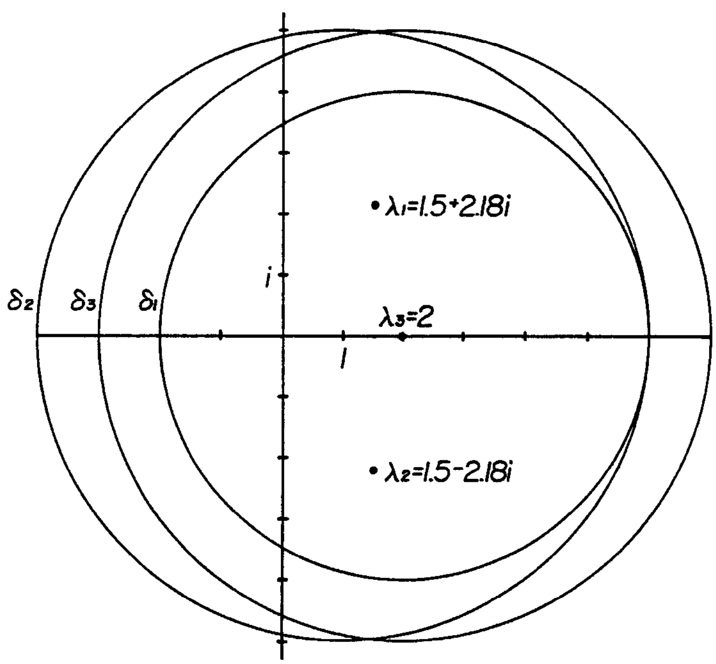

Fig. 6.5

Gerschgorin's eigenvalue bounds are utterly simple, but on difference matrices the theorem fails where we need it most. The difference matrices of mathematical physics are, as we noticed in Chapter 3, most commonly symmetric and positive definite. We know that for these matrices all eigenvalues are positive, $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ but we would like to have a *lower* bound λ_1 in order to secure an *upper* bound on λ_n/λ_1. In this respect Gerschgorin's theorem is a disappointment.

For matrix

$$A = \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix} \tag{6.319}$$

Gerschgorin's theorem yields the eigenvalue interval $0 \le \lambda \le 4$ for any n, failing to predict the positive definiteness of A. For matrix

$$A^2 = \begin{bmatrix} 5 & -4 & 1 & & \\ -4 & 6 & -4 & 1 & \\ 1 & -4 & 6 & -4 & 1 \\ & 1 & -4 & 6 & -4 \\ & & 1 & -4 & 5 \end{bmatrix} \tag{6.320}$$

Gerschgorin's theorem yields $-4 \le \lambda \le 16$, however large n is, where in fact $\lambda > 0$.

Similarity transformations can save Gerschgorin's estimates for these matrices. First we notice that $D^{-1}AD$ and $D^{-1}A^2D$, with the diagonal $D, D_{ii} = (-1)^i$ turns all entries of the transformed matrices nonnegative. Matrices with nonnegative or positive entries are common; A^{-1} and $(A^2)^{-1}$ are with entries that are all positive.

Definition. *Matrix A is nonnegative, $A \ge O$, if $A_{ij} \ge 0$ for all i and j. It is positive, $A > O$, if $A_{ij} > 0$.*

Discussion of good similarity transformations to improve the lower bound on the eigenvalues is not restricted to the finite difference matrix A, and we shall look at a broader class of these matrices.

Theorem 6.66. *Let symmetric tridiagonal matrix*

$$T = \begin{bmatrix} \alpha_1 + \alpha_2 & -\alpha_2 & & \\ -\alpha_2 & \alpha_2 + \alpha_3 & -\alpha_3 & \\ & -\alpha_3 & \ddots & -\alpha_n \\ & & -\alpha_n & \alpha_n + \alpha_{n+1} \end{bmatrix} \tag{6.321}$$

be such that $\alpha_1 \ge 0,\ \alpha_2 > 0,\ \alpha_3 > 0, \ldots, \alpha_n > 0,\ \alpha_{n+1} \ge 0$. Then eigenvector x corresponding to its minimal eigenvalue is positive, $x > o$.

Proof. If $x = [x_1\ x_2\ \dots\ x_n]^T$ is the eigenvector corresponding to the lowest eigenvalue λ, then

$$\lambda(x) = \frac{\alpha_1 x_1^2 + \alpha_2(x_2 - x_1)^2 + \alpha_3(x_3 - x_2)^2 + \cdots + \alpha_n(x_n - x_{n-1})^2 + \alpha_{n+1}x_n^2}{x_1^2 + x_2^2 + \cdots + x_n^2} \tag{6.322}$$

and matrix T is seen to be positive semidefinite. Matrix T is singular only if both $\alpha_1 = \alpha_{n+1} = 0$, and then $x = [1\ 1\ \dots\ 1]^T$. Suppose therefore that α_1 and α_{n+1} are not both zero.

Looking at the equation $Ax = \lambda x$ we readily observe that no two consecutive components of x, including x_1 and x_2, may be zero, for this would imply $x = o$. No interior component of x can be zero either; speaking physically the string may have no interior node, for this would contradict the fact, by Theorem 6.49, that x is the unique minimizer of $\lambda(x')$. Say $n = 4$ and $x_2 = 0$. Then the numerator of $\lambda(x)$ is $(\alpha_1 + \alpha_2)x_1^2 + \alpha_3 x_3^2 + \alpha_4(x_3 - x_4)^2 + \alpha_5 x_4^2$, and replacing x_1 by $-x_1$ leaves $\lambda(x)$ un affected. The components of x cannot be of different signs because sign reversals would lower the numerator of $\lambda(x)$ without changing the denominator contradicting the assumption that x is a minimizer of $\lambda(x')$. Hence we may choose all components of x positive. End of proof.

For the finite difference matrix A of eq.(6.319), or for that matter for any symmetric matrix A such that $A_{ii} > 0$ and $A_{ij} \leq 0$, the lower Gerschgorin bound on first eigenvalue λ_1 may be written as

$$\lambda_1 \geq \min_i (Ae)_i \tag{6.323}$$

for $e = [1\ 1\ 1\ \dots, 1]^T$. If D is a positive diagonal matrix, $D > O$, then also

$$\lambda_1 \geq \min_i (D^{-1}ADe)_i \tag{6.324}$$

where equality holds for $De = x_1$ if $x_1 > o$.

Matrix A of eq.(6.319) has a first eigenvector with components that are all positive, that is, approximately $x'_1 = [0.50\ 0.87\ 1.00\ 0.87\ 0.50]^T$. Taking the diagonal matrix D with $D_{ii} = (x'_1)_i$ yields

$$D^{-1}AD = \begin{bmatrix} 2 & -1.740 & & & \\ -0.575 & 2 & -1.15 & & \\ & -0.87 & 2 & -0.87 & \\ & & -1.15 & 2 & -0.575 \\ & & & -1.74 & 2 \end{bmatrix} \tag{6.325}$$

and from its five rows we obtain the five, almost equal, inequalities

$$\lambda_1 \geq 0.260,\ \lambda_1 \geq 0.275,\ \lambda_1 \geq 0.260,\ \lambda_1 \geq 0.275,\ \lambda_1 \geq 0.260 \tag{6.326}$$

so that certainly $\lambda_1 \geq 0.260$, whereas actually $\lambda_1 = 4\sin^2\ 15^o = 0.26795$.

On the other hand according to Rayleigh's theorem $\lambda_1 \leq \lambda_1(x'_1) = 0.26797$, and $0.260 \leq \lambda_1 \leq 0.26797$.

Gerschgorin's theorem does not require the knowledge that x'_1 is a good approximation to x_1, but suppose that we know that $\lambda'_1 = \lambda(x'_1) = 0.26797$ is nearest to λ_1. Then from $r = Ax'_1 - \lambda'_1 x'_1 = 10^{-3}[3.984\ 6.868\ 7.967\ 6.868\ 3.984]^T$ we get that $0.260 \leq \lambda_1 \leq 0.276$.

Similarly, if symmetric A is nonnegative $A \geq O$, then Gerschgorin's upper bound on the eigenvalues of A becomes

$$\lambda_n \leq \max_i (D^{-1}ADe)_i \tag{6.327}$$

for any $D > O$.

The following is a symmetric version of Perron's theorem on positive matrices.

Theorem (Perron) 6.67. *If A is a symmetric positive matrix, then the eigenvector corresponding to the largest (positive) eigenvalue of A is positive and unique.*

Proof. If x_n is a unit eigenvector corresponding to λ_n, and $x \neq x_n$ is such that $x^T x = 1$, then

$$x^T Ax < \lambda_n = \lambda(x_n) = x_n^T Ax_n \tag{6.328}$$

and λ_n is certainly positive. Moreover, since $A_{ij} > 0$ the components of x_n cannot have different signs, for this would contradict the assumption that x_n maximizes $\lambda(x)$. Say then that $(x_n)_i \geq 0$. But none of the $(x_n)_i$ components can be zero since $Ax_n = \lambda_n x_n$, and obviously $Ax_n > o$. Hence $x_n > o$.

There can be no other positive vector orthogonal to x_n, and hence the eigenvector, and also the largest eigenvalue λ_n, are unique. End of proof.

Theorem 6.68. *Suppose that A has a positive inverse, $A^{-1} > O$. Let x be any vector satisfying $Ax - e = r, e = [1\ 1\ \ldots\ 1]^T$, $\|r\|_\infty < 1$. Then*

$$\frac{\|x\|_\infty}{1 + \|r\|_\infty} \leq \|A^{-1}\|_\infty \leq \frac{\|x\|_\infty}{1 - \|r\|_\infty}. \tag{6.329}$$

Proof. Obviously $x = A^{-1}e + A^{-1}r$ so that

$$\|x\|_\infty \le \|A^{-1}e\|_\infty + \|A^{-1}r\|_\infty \le \|A^{-1}\|_\infty + \|A^{-1}\|_\infty \|r\|_\infty \tag{6.330}$$

and

$$\frac{\|x\|_\infty}{1 + \|r\|_\infty} \le \|A^{-1}\|_\infty. \tag{6.331}$$

To prove the other bound write $x = A^{-1}e - (-A^{-1}r)$, observe that $\|A^{-1}e\|_\infty = \|A^{-1}\|_\infty$, and have that

$$\begin{aligned} \|x\|_\infty &\ge \|A^{-1}e\|_\infty - \|A^{-1}r\|_\infty \\ &\ge \|A^{-1}\|_\infty - \|A^{-1}\|_\infty \|r\|_\infty. \end{aligned} \tag{6.332}$$

Hence, if $\|r\|_\infty < 1$, then

$$\frac{\|x\|_\infty}{1 - \|r\|_\infty} \ge \|A^{-1}\|_\infty. \tag{6.333}$$

End of proof.

Gerschgorin's theorem has some additional interesting consequences.

Theorem 6.69. *The eigenvalues of a symmetric matrix depend continuously on its entries.*

Proof. Let matrix $B = B^T$ be such that $|B_{ij}| < \epsilon$. The theorems of Gerschgorin and Hirsch assure us that the eigenvalues of B are in the interval $-n\epsilon \le \beta \le n\epsilon$. If $C = A + B$, then according to Theorem 6.62 $|\gamma_i - \alpha_i| \le n\epsilon$ where $\alpha_1 \le \alpha_2 \le \cdots \le \alpha_n$ are the eigenvalues of A and $\gamma_1 \le \gamma_2 \le \cdots \le \gamma_n$ are the eigenvalues of C. As $\epsilon \to 0$ so does $|\gamma_i - \alpha_i|$, and $|\gamma_i - \alpha_i|/\epsilon$ is finite for all $\epsilon > 0$. End of proof.

The eigenvalues of any matrix depend continuously on its entries. It is a basic result of polynomial equation theory that the roots of the equation depend continuously on the coefficients (which does not mean that roots cannot be very sensitive to small changes in the coefficients.) We shall not prove it here, but will accept this fact to prove the second Gerschgorin theorem on the distribution of the eigenvalues in the discs. It is this theorem that makes Gerschgorin's theorem invaluable for nearly diagonal symmetric matrices.

Theorem (Gerschgorin) 6.70. *If k Gerschgorin discs of matrix A are disjoint from the other discs, then precisely k eigenvalues of A are found in the union of the k discs.*

Proof. Write $A = D + A'$ with diagonal $D_{ii} = A_{ii}$, and consider matrix $A(\tau) = D + \tau A'$ $0 \le \tau \le 1$. Obviously $A(0) = D$ and $A(1) = A$. For clarity we shall continue the proof for a real 3×3 matrix, but the argument is general.

Suppose that the three Gerschgorin discs $\delta_1 = \delta_1(1),\ \delta_2 = \delta_2(1),\ \delta_3 = \delta_3(1)$ for $A = A(1)$ are as shown in Fig. 6.6 For $\tau = 0$ the three circles contract to points $\lambda_1(0) = A_{11},\ \lambda_2(0) = A_{22},\ \lambda_3(0) = A_{33}$. As τ is increased the three discs $\delta_1(\tau),\ \delta_2(\tau),\ \delta_3(\tau)$ for $A(\tau)$ expand and the three eigenvalues $\lambda_1(\tau), \lambda_2(\tau), \lambda_3(\tau)$ of A vary inside them. Never is an eigenvalue of $A(\tau)$ outside the union of the three discs. Disc $\delta_1(\tau)$ is disjoint from the other two discs for any $0 \le \tau \le 1$. Since $\lambda_1(\tau)$ varies continuously with τ it cannot jump over to the other two discs, and the same is true for $\lambda_2(\tau)$ and $\lambda_3(\tau)$. Hence δ_1 contains one eigenvalue of A and $\delta_1 \cup \delta_2$

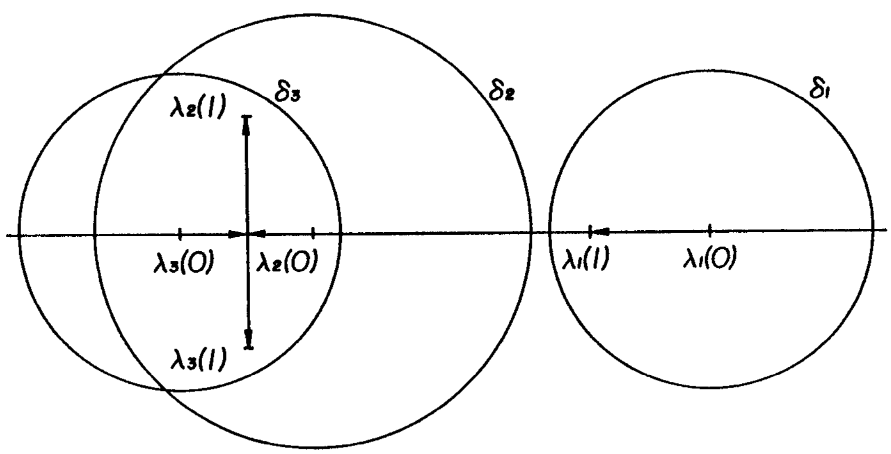

Fig. 6.6

Example. Straightforward application of Gerschgorin's theorem to

$$A = \begin{bmatrix} 1 & -10^{-2} & 2\,10^{-2} \\ -5\,10^{-3} & 2 & 10^{-2} \\ 10^{-2} & -10^{-2} & 3 \end{bmatrix} \tag{6.334}$$

yields

$$|\lambda_1 - 1| \le 3.0\,10^{-2},\ |\lambda_2 - 2| \le 1.5\,10^{-2},\ |\lambda_3 - 3| \le 2.0\,10^{-2} \tag{6.335}$$

and we conclude that the three eigenvalues are real. A better bound on, say, λ_1 is obtained with a similarity transformation that maximally contracts the disc around λ_1 but leaves it disjoint of the other discs. Multiplication of the first row of A by 10^{-2} and the first column of A by 10^2 amounts to the similarity transformation

$$D^{-1}AD = \begin{bmatrix} 1 & -10^{-4} & 210^{-4} \\ -0.5 & 2 & 10^{-2} \\ 1 & -10^{-2} & 3 \end{bmatrix} \tag{6.336}$$

from which we obtain the better $|\lambda_1 - 1| \leq 3.0\ 10^{-4}$.

Corollary 6.71. *A disjoint Gerschgorin disc of a real matrix contains one real eigenvalue.*

Proof. For a real matrix all discs are centered on the real axis and there are no two disjoint discs that contain $\lambda = \alpha + i\beta$ and $\overline{\lambda} = \alpha - i\beta$, $\beta \neq 0$. Hence $\beta = 0$. End of proof.

With a good similarity transformation Gerschgorin's theorem may be made to do well even on a triangular matrix. Consider the upper-triangular $U, U_{ij} = 1$. Using diagonal $D, D_{ii} = \epsilon^{n-i}$ we have

$$DUD^{-1} = \begin{bmatrix} 1 & \epsilon & \epsilon^2 & \epsilon^3 \\ & 1 & \epsilon & \epsilon^2 \\ & & 1 & \epsilon \\ & & & 1 \end{bmatrix} \tag{6.337}$$

and we can make the discs have arbitrarily small radii around $\lambda = 1$.

Gerschgorin's theorem does not know to distinguish between a matrix that is only slightly asymmetric and a matrix that is grossly asymmetric, and it is might be desirable to decouple the real and imaginary parts of the eigenvalue bounds. For this we have

Theorem (Bendixon) 6.72. *If real $A = A(n \times n)$ has complex eigenvalue $\lambda = \alpha + i\beta$, then α is neither more nor less than any eigenvalue of $\frac{1}{2}(A + A^T)$, and β is neither more nor less than any eigenvalue of $\frac{1}{2i}(A - A^T)$.*

Proof. As we did in the proof to Theorem 6.61 we write $Ax = \lambda x$ with $x = u + iv, u^Tu = v^Tv = 1$, and decouple the complex eigenproblem into the pair of equations

$$2\alpha = \frac{1}{2}u^T(A + A^T)u + \frac{1}{2}v^T(A + A^T)v, \quad 2\beta = u^T(A - A^T)v. \tag{6.338}$$

Now we think of u and v as being variable unit vectors. Matrix $A + A^T$ is symmetric, and it readily results from Rayleigh's Theorem 6.50 that α in eq.(6.338) can neither dip lower than the minimum nor can it rise higher than the maximum eigenvalues of $\frac{1}{2}(A + A^T)$. Matrix $A - A^T$ is skew-symmetric and has purely imaginary eigenvalues of the form $\lambda = \pm i2\sigma$. Also, $u^T(A - A^T)u = 0$ whatever u is. Therefore we restrict v to being orthogonal to u, and propose to accomplish this by $v = -1/2\beta(A - A^T)$, with factor $-1/2\beta$ guaranteeing $v^Tv = 1$. Presently,

$$4\beta^2 = -u^T(A - A^T)^2u. \tag{6.339}$$

Matrix $-(A - A^T)^2$ is symmetric and has nonnegative eigenvalues all of the form $\lambda = \sigma^2$. Rayleigh's theorem assures us again that $4\beta^2$ is invariably located between the least and most values of $4\sigma^2$, and the proof is done.

exercises

6.17.1. Show that the roots of $\lambda^2 - a_1\lambda + a_0 = 0$ depend continuously on the coefficients. Give a geometrical interpretation to $\lambda\overline{\lambda}$.

6.17.2. Use Gerschgorin's theorem to show that the $n \times n$

$$A = \begin{bmatrix} \alpha & 1 & 1 & 1 \\ 1 & \alpha & 1 & 1 \\ 1 & 1 & \alpha & 1 \\ 1 & 1 & 1 & \alpha \end{bmatrix}$$

is positive definite if $\alpha > n - 1$. Compute all eigenvalues of A.

6.17.3. Use Gerschgorin's theorem to show that

$$A = \begin{bmatrix} 5 & -1 & & \\ -1 & 4 & 2 & \\ & 1 & -3 & 1 \\ & & 1 & -2 \end{bmatrix}$$

is nonsingular.

6.17.4. Does

$$A = \begin{bmatrix} 2 & 1 & & & \\ -1 & 6 & 1 & & \\ & -1 & 10 & 1 & \\ & & -1 & 14 & 1 \\ & & & -1 & 18 \end{bmatrix}$$

have complex eigenvalues?

6.17.5. Consider

$$A = \begin{bmatrix} 1 & -1 & & \\ -1 & 4 & -3 & \\ & -3 & 8 & -5 \\ & & -5 & 12 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 1 & & & \\ & \alpha & & \\ & & 0.7 & \\ & & & 0.4 \end{bmatrix}.$$

Show that the spectrum of A is nonnegative. Form $D^{-1}AD$ and apply Gerschgorin's theorem to this matrix. Determine α so that the lower bound on the lowest eigenvalue of A is as high as possible.

6.17.6. Nonnegative matrix A with row sums all being equal to 1 is said to be a *stochastic* matrix. Positive, $A_{ij} > 0$, stochastic matrix A is said to be a transition matrix. Obviously $e = [1\ 1\ \ldots\ 1]^T$ is an eigenvector of transition matrix A for eigenvalue $\lambda = 1$. Use the Gerschgorin circles of Theorem 6.64 to show that all eigenvalues of transition matrix A are such that $|\lambda| \leq 1$, with equality holding only for $\lambda = 1$. The proof that eigenvalue $\lambda = 1$ is of algebraic multiplicity 1 is more difficult, but it establishes a crucial property of A that assures, by Theorem 6.46, that the *Markov* process $A_n = A^n$ has a limit as $n \to \infty$.

6.17.7. Let $S = S(3 \times 3)$ be a stochastic matrix with row sums all equal to λ. Show that elementary operations matrix

$$E = \begin{bmatrix} 1 & & 1 \\ & 1 & 1 \\ & & 1 \end{bmatrix}, \quad E^{-1} = \begin{bmatrix} 1 & & -1 \\ & 1 & -1 \\ & & 1 \end{bmatrix}$$

is such that $E^{-1}SE$ *deflates* matrix S to the effect that

$$E^{-1}SE = \begin{bmatrix} A_{11} - A_{31} & A_{12} - A_{32} & 0 \\ A_{21} - A_{31} & A_{22} - A_{32} & 0 \\ A_{31} & A_{32} & \lambda \end{bmatrix}.$$

Apply this to

$$A = \begin{bmatrix} 2 & 3 & 2 \\ 1 & 2 & 4 \\ 5 & 1 & 1 \end{bmatrix}$$

for which $\lambda = 7$. Then apply Gerschgorin's Theorem 6.64 to the leading 2×2 diagonal block of $E^{-1}SE$ to bound the rest of the eigenvalues of S. Explain how to generally deflate a square matrix with a known eigenvalue and corresponding eigenvector.

6.17.8. Referring to Theorem 6.68 take

$$A = \begin{bmatrix} 1 & -1 & & & \\ -1 & 4 & -3 & & \\ & -3 & 8 & -5 & \\ & & -5 & 12 & -7 \\ & & & -7 & 16 \end{bmatrix}$$

and $x = \alpha[5\ 4\ 3\ 2\ 1]^T$. Fix α so that $\|r\|_\infty$ is lowest and make sure it is less than 1. Bound $\|A^{-1}\|_\infty = \|A^{-1}e\|_\infty$ and compare the bounds with the computed $\|A^{-1}\|_\infty$.

6.17.9. The characteristic equation of companion matrix

$$C = \begin{bmatrix} & & & -a_0 \\ 1 & & & -a_1 \\ & 1 & & -a_2 \\ & & 1 & -a_3 \end{bmatrix}$$

is $z^4 + a_3 z^3 + a_2 z^2 + a_1 z + a_0 = 0$. With diagonal matrix $D, D_{ii} = \alpha_i > 0$, obtain

$$D^{-1}CD = \begin{bmatrix} & & & -a_0\alpha_4/\alpha_1 \\ \alpha_1/\alpha_2 & & & -a_1\alpha_4/\alpha_2 \\ & \alpha_2/\alpha_3 & & -a_2\alpha_4/\alpha_2 \\ & & \alpha_3/\alpha_4 & -a_3\alpha_4/\alpha_4 \end{bmatrix}.$$

Recall Gerschgorin's theorem to deduce from it that root z of a polynomial equation of degree n is of modulus

$$|z| \le max(\frac{\alpha_i}{\alpha_{i+1}} + |a_i|\frac{1}{\alpha_{i+1}}), \quad i = 0, 1, \ldots, n-1$$

if $\alpha_0 = 0$ and $\alpha_n = 1$.

6.17.10. For matrix A define $\sigma_i = |A_{ii}| - \sum_{i \neq j} |A_{ij}|$. Show that if $\sigma_i > 0$ for all i, then $A^{-1} = B$ is such that $|B_{ij}| \le \sigma_i^{-1}$.

6.17.11. Prove Schur's inequality:

$$\sum_{i=1}^{n} |\lambda_i|^2 \le \sum_{i,j=1}^{n} |A_{ij}|^2$$

where λ_i $i = 1, 2, \ldots, n$ are the eigenvalues of A.

6.17.12. Prove Browne's theorem: If $A = A(n \times n)$ is real, then $|\lambda(A)|^2$ lies between the smallest and largest eigenvalues of AA^T.

6.17.13. Show that if A is symmetric and positive definite, then its largest eigenvalue is bounded by

$$\max_i |A_{ii}| \leq \lambda_n \leq n \max_i |A_{ii}|.$$

6.17.14. Show that if A is diagonalizable, $A = XDX^{-1}$ with $D_{ii} = \lambda_i$, then for any given scalar λ and unit vector x

$$\min_i |\lambda_i - \lambda| \leq \|X\| \; \|X^{-1}\| \; \|r\|$$

where $r = Ax - \lambda x$. Hint: Write $x = Xx'$.

6.17.15. Prove the Bauer-Fike theorem: If A is diagonalizable, $A = XDX^{-1}, D_{ii} = \lambda_i$ then for any eigenvalue λ' of $A' = A + E$,

$$\min_i |\lambda_i - \lambda'| \leq \|X^{-1}EX\| \leq \|X\| \; \|X^{-1}\| \; \|E\|.$$

6.17.16. Show that if A and B are positive definite, then $C, C_{ij} = A_{ij}B_{ij}$, is also positive definite.

6.17.17. Show that every $A = A(n \times n)$ with $\det(A) = 1$ can be written as $A = (BC)(CB)^{-1}$.

6.17.18. Prove that real $A(n \times n) = -A^T,\ n > 2$, has an even number of zero eigenvalues if n is even and an odd number of zero eigenvalues if n is odd.

6.17.19. Diagonal matrix I' is such that $I'_{ii} = \pm 1$. Show that whatever A, $I'A + I$ is nonsingular for some I'. Show that every orthogonal Q can be written as $Q = I'(I - S)(I + S)^{-1}$, where $S = -S^T$.

6.17.20. Let λ_1 and λ_n be the extreme eigenvalues of positive definite and symmetric matrix A. Show that

$$1 \leq \frac{x^TAx}{x^Tx}\frac{x^TA^{-1}x}{x^Tx} \leq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}.$$

6.18 The Ritz reduction

Matrices raised by such practices as computational mechanics are of immense order n, but usually only few eigenvalues at the lower end of the spectrum are of interest. We may know a subspace of dimension m, much smaller than n, in which good approximations to the first $m' \leq m$ eigenvectors of the symmetric $A = A(n \times n)$ can be found.

The Ritz reduction method tells us how to find optimal approximations to the first m' eigenvalues of A with eigenvector approximations confined to the m-dimensional subspace of R^n, by solving an $m \times m$ eigenproblem only.

Let $v_1, v_2, \ldots, v_m$ be an orthonormal basis for subspace V^m of R^n. In reality the basis for V^m may not be originally orthogonal but in theory we may always assume it to be so. Suppose that we are interested in the lowest eigenvalue λ_1 of $A = A^T$ only, and know that a good approximation to the corresponding eigenvector x_1 lurks in V^m. To find $x \in V^m$ that produces the eigenvalue approximation closest to λ_1 we follow Ritz in writing

$$x = y_1 v_1 + y_2 v_2 + \cdots + y_m v_m = Vy \tag{6.340}$$

where $V = [v_1 v_2 \ldots v_m]$, and where $y = [y_1\ y_2\ \ldots\ y_m]^T$, and seek $y \neq o$ in R^m that minimizes

$$\rho(y) = \frac{x^T A x}{x^T x} = \frac{y^T V^T A V y}{y^T y}. \tag{6.341}$$

Setting grad$\rho(y) = o$ produces

$$(V^T A V) y = \rho y \tag{6.342}$$

which is only an $m \times m$ eigenproblem.

Symmetric matrix $V^T A V$ has m eigenvalues $\rho_1 \leq \rho_2 \leq \cdots \leq \rho_m$ and m corresponding orthogonal eigenvectors $y_1, y_2, \ldots, y_m$. According to Rayleigh's theorem $\rho_1 \geq \lambda_1$ and is as near as it can get to λ_1 with $x \in V^m$. What about the other $m-1$ eigenvalues? The next two theorems clear up this question.

Theorem (Poincaré) 6.73. *Let the eigenvalues of the symmetric $n \times n$ matrix A be $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. If matrix $V = V(n \times m)$, $m \leq n$, is with m orthonormal columns, $V^T V = I$, then the m eigenvalues $\rho_1 \leq \rho_2 \leq \cdots \leq \rho_m$ of the $m \times m$ eigenproblem*

$$V^T A V y = \rho y \tag{6.343}$$

are such that

$$\lambda_1 \leq \rho_1 \leq \lambda_{n-m+1},\ \lambda_2 \leq \rho_2 \leq \lambda_{n-m+2}, \ldots, \lambda_{m-1} \leq \rho_{m-1} \leq \lambda_{n-1},\ \lambda_m \leq \rho_m \leq \lambda_n. \tag{6.344}$$

Proof. Let $V = [v_1 v_2, \ldots, v_m]$ and call V^m the column space of V. Augment the basis for V^m to the effect that $v_1, v_2, \ldots, v_m, \ldots, v_n$ is an orthonormal basis for R^n and start with

$$\rho_m = \max_y \frac{y^T V^T A V y}{y^T y} \ , \ y \in R^m \tag{6.345}$$

or

$$\rho_m = \max_x \frac{x^T A x}{x^T x} \ , \ x \in V^m. \tag{6.346}$$

This, in turn, is equivalent to

$$\rho_m = \max_x \frac{x^T A x}{x^T x} \ , \ x^T v_{m+1} = \cdots = x^T v_n = 0 \tag{6.347}$$

and Fischer's theorem tells us that $\rho_m \geq \lambda_m$. The next Ritz eigenvalue ρ_{m-1} is obtained from the maximization under the additional constraint $y^T y_1 = x^T V y_1 = x^T x'_1, \ x \in V^m$,

$$\rho_{m-1} = \max_x \frac{x^T A x}{x^T x} \ , \ x^T x'_1 = x^T v_{m+1} = \cdots = x^T v_n = 0 \tag{6.348}$$

and by Fischer's theorem $\rho_{m-1} \geq \lambda_{m-1}$. Continuing this way we prove the m left-hand inequalities of the theorem.

The second part of the theorem is proved starting with

$$\rho_1 = \min_x \frac{x^T A x}{x^T x} \ , \ x^T v_{m+1} = \cdots = x^T v_n \tag{6.349}$$

and with the assurance by Fischer's theorem that $\rho_1 \leq \lambda_{n-m+1}$, and so on. End of proof.

If subspace V^m is given by the linearly independent $v_1, v_2, \ldots, v_m$ and if a Gram-Schmidt orthogonalization is impractical, then we still write $x = Vy$ and have that

$$\rho(y) = \frac{x^T A x}{x^T x} = \frac{y^T V^T A V y}{y^T V^T V y} \tag{6.350}$$

with a positive definite and symmetric $V^T V$. Setting grad$\rho(y) = o$ yields now the more general

$$(V^T A V) y = \rho (V^T V) y. \tag{6.351}$$

The first Ritz eigenvalue ρ_1 is obtained from the minimization of $\rho(y)$, the last ρ_m from the maximization of $\rho(y)$, and hence the extreme Ritz eigenvalues are optimal in the sense that ρ_1 comes as near as possible to λ_1, and ρ_m comes as close as possible to λ_n. All the Ritz eigenvalues have a similar property and are optimal in the sense of

Theorem 6.74. *Let A be symmetric and have eigenvalues $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$. If $\rho_1 \le \rho_2 \le \cdots \le \rho_m$ are the Ritz eigenvalues with corresponding orthonormal eigenvectors $x'_1, x'_2, \ldots, x'_m$, then for $k = 1, 2, \ldots, m$*

$$\rho_k - \lambda_k = \min_{x \in V^m} \left(\frac{x^T A x}{x^T x} - \lambda_k \right),\ x^T x'_1 = \cdots = x^T x'_{k-1} = 0 \tag{6.352}$$

and

$$\lambda_{n+1-k} - \rho_{m+1-k} = \min_{x \in V^m} \left(\lambda_{n+1-k} - \frac{x^T A x}{x^T x} \right),\ x^T x'_m = \cdots = x^T x'_{m+2-k} = 0. \tag{6.353}$$

Proof. For a proof to the first part of the theorem we consider the Ritz eigenvalues as obtained through the minimization

$$\rho_k = \min_y \frac{y^T V^T A V y}{y^T y}\ ,\ y^T y_1 = \cdots = y^T y_{k-1} = 0 \tag{6.354}$$

where $V = V(n \times m) = [v_1 v_2, \ldots, v_m]$ has m orthonormal columns. Equivalently

$$\rho_k = \min_{x \in V^m} \frac{x^T A x}{x^T x}\ ,\ x^T x'_1 = \cdots = x^T x'_{k-1} = 0 \tag{6.355}$$

where $x = Vy$, and $x'_j = V y_j$. By Poincaré's theorem $\rho_k \ge \lambda_k$ and hence the minimization lowers ρ_k as much as possible to bring it as close as possible to λ_k under the restriction that $x \in V^m$ and $x^T x_1 = \cdots = x^T x'_{k-1} = 0$.

The second part of the theorem is proved similarly by considering the Ritz eigenvalues as obtained by the maximization

$$\rho_{m+1-k} = \max_{x \in V^m} \frac{x^T A x}{x^T x}\ ,\ x^T x'_m = \cdots = x^T x'_{m+2-k} \tag{6.356}$$

the details of which are left as an exercise. End of proof.

For any given Ritz eigenvalue ρ_j and corresponding approximate eigenvector x'_j we may compute the residual vector $r_j = A{x_j}' - \rho_j {x_j}'$ and are assured that the interval $|\rho_j - \lambda| \le \|r_j\| / \|{x_j}'\|$ contains an eigenvalue of A. The bounds are not sharp but they require no knowledge of the eigenvalue distribution, nor that ${x_j}'$ be any special vector and ρ_j any special number. If such intervals for different Ritz eigenvalues and eigenvectors overlap, then we know that the union of overlapping intervals contain an eigenvalue of A. Whether or not more than one eigenvalue is found in the union is not revealed to us by this simple error analysis.

An error analysis based on Corollary 6.63 involving a residual matrix rather than residual vectors removes the uncertainty on the number of eigenvalues in overlapping intervals.

Let $X' = X'(n \times m) = [x'_1 x'_2 \dots x'_m]$, D the diagonal $D_{ii} = \rho_i$, and define the residual matrix

$$R = AX' - X'D. \tag{6.357}$$

Obviously ${X'}^T R = {X'}^T AX' - D = O$, since the columns of X' are orthonormal. Augment X' so that $Q = [X'X'']$ is an orthogonal matrix and form

$$Q^T AQ = \begin{bmatrix} {X'}^T AX' & {X'}^T AX'' \\ {X''}^T AX' & {X''}^T AX'' \end{bmatrix} = \begin{bmatrix} D & {X''}^T R \\ R^T X'' & {X''}^T AX'' \end{bmatrix}. \tag{6.358}$$

The maximal eigenvalue of ${X''}^T RR^T X''$ is less than the maximal eigenvalue of RR^T or $R^T R$. Hence by Corollary 6.63 if ϵ^2 is the largest eigenvalue of $R^T R$, then the union of intervals $|\rho_i - \lambda| \le |\epsilon|$ $i = 1, 2, \dots, m$ contains m eigenvalues of A.

Example. Let x be an arbitrary vector in R^n, and let $A = A(n \times n)$ be a symmetric matrix. In this example we want to examine the *Krylov* sequence $x, Ax, \dots, A^{m-1}x$ as a basis for V^m. An obvious difficulty with this sequence is that the degree of the minimal polynomial of A can be less than n and the sequence may become linearly dependent for $m-1 < n$. Near-linear dependence among the Krylov vectors is more insidious, and we shall look also at this unpleasant prospect.

To simplify the computation we choose $A = A(100 \times 100)$ to be diagonal, with eigenvalues

$$\lambda_{i,j} = \frac{1}{2.5}(i^2 + 1.5j^2) \quad i = 1, 2, \dots, 10 \quad j = 1, 2, \dots, 10 \tag{6.359}$$

so that the first five are 1., 2.2, 2.8, 4.0, 4.2; and the last one is 100.0. It occurs to us to take $x = \sqrt{n}/n[1\ 1\ \dots\ 1]^T$, and we normalize $Ax,\ A^2x, \dots, A^{m-1}x$ to avoid very large vector magnitudes.

The table below lists the four lowest Ritz eigenvalues computed from V^m with a Krylov basis, as a function of m.

m	ρ_1	ρ_2	ρ_3	ρ_4
4	7.615	31.47	61.83	91.85
6	4.141	17.08	36.78	58.94
8	2.680	10.57	23.42	39.38
12	1.469	5.049	11.19	19.63.

Since the basis of V^m is not orthogonal, the Ritz eigenproblem is here the general $V^T AVy = \rho V^T Vy$ and we solved it with a commercial procedure. For m larger than 12 the eigenvalue procedure returns meaningless results. Computation of the eigenvalues of $V^T V$ itself revealed a spectral condition number $\kappa(V^T V) = (1.5m)!$ which means $\kappa = 6 \cdot 5\ 10^{15}$ for $m = 12$, and all the high accuracy used could not save $V^T V$ from singularity.

exercises

6.18.1. For matrix A

$$A = \begin{bmatrix} 1 & -1 & & \\ -1 & 4 & -3 & \\ & -3 & 8 & -5 \\ & & -5 & 12 \end{bmatrix}$$

determine α_1 and α_2 in $x = \alpha_1[1\ 1\ 0\ 0]^T + \alpha_2[0\ 0\ 1\ 1]^T$ so that $x^T Ax/x^T x$ is minimal.

Answers

section 6.1

6.1.1. $\lambda = 2, \alpha_1 = -\alpha_2 = \pm\sqrt{2}/2$ or $\lambda = 4, \alpha_1 = \alpha_2 = \pm\sqrt{2}/2$.

6.1.2. Yes, $\lambda = -2$.

6.1.3. Yes, $\lambda = 1/3$.

section 6.2

6.2.1. $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 0,\ x = \alpha_1 e_1 + \alpha_3 e_3$ for arbitrary α_1, α_2.

6.2.2.

$$\text{for } A:\ \lambda_1 = 1, x_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix};\ \lambda_2 = -2, x_2 = \begin{bmatrix} -1 \\ 3 \\ 0 \end{bmatrix};\ \lambda_3 = 1, x_3 = \begin{bmatrix} 7 \\ 4 \\ 10 \end{bmatrix}.$$

$$\text{for } B: \ \lambda_1 = \lambda_2 = \lambda_3 = 1, x = e_1.$$

$$\text{for } C: \ \lambda_1 = \lambda_2 = 1, x = \alpha_1 e_1 + \alpha_2 e_2; \ \lambda_3 = 2, x_3 = \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}.$$

section 6.3

6.3.1.

$$\begin{bmatrix} 1-\lambda & 1+\lambda \\ -1+2\lambda & -1-\lambda \end{bmatrix} \to \begin{bmatrix} 1-\lambda & 1+\lambda \\ 1 & 1+\lambda \end{bmatrix} \to \begin{bmatrix} -\lambda & 0 \\ 1 & 1+\lambda \end{bmatrix}.$$

6.3.2. $-\lambda^3 + \alpha_2 \lambda^2 - \alpha_1 \lambda + \alpha_0 = 0.$

6.3.5. $f(A) = \lambda_1^2 + \lambda_2^2.$

6.3.6.

$$\text{for } A: \ \lambda_1 = 1, x_1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}; \ \lambda_2 = 4, x_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

$$\text{for } B: \ \lambda = 1 \pm i, x_1 = \begin{bmatrix} 1 \\ \pm i \end{bmatrix}.$$

$$\text{for } C: \ \lambda_1 = 0, x_1 = \begin{bmatrix} 1 \\ i \end{bmatrix}; \ \lambda_2 = 2, x_2 = \begin{bmatrix} 1 \\ -i \end{bmatrix}.$$

$$\text{for } D: \ \lambda_1 = \lambda_2 = 0, x_1 = \begin{bmatrix} -i \\ 1 \end{bmatrix}.$$

6.3.7.

$$\text{for } A: \ \lambda_1 = -1, x_1 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}; \ \lambda_2 = 0, x_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}; \ \lambda_3 = 1, x_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

$$\text{for } B: \ \lambda_1 = 0, x_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}; \ \lambda_2 = i, x_2 = \begin{bmatrix} 1 \\ 0 \\ i \end{bmatrix}; \ \lambda_3 = -i, x_3 = \begin{bmatrix} 1 \\ 0 \\ -i \end{bmatrix}.$$

6.3.8. $\alpha_1 = \alpha_2 = 2.$

6.3.9 $\alpha^2 < 1/4.$

6.3.10. $\alpha = 1, \lambda = 0.$

6.3.11. $\lambda = 2.$

6.3.12. $\lambda = 1.$

6.3.13. An eigenvector of A for $\lambda = 1$.

section 6.4

6.4.1. Yes. $\alpha_1 = -3 + 4i, \alpha_2 = 2 - 3i, \alpha_3 = 5 - 3i$.

6.4.2. Yes. No, $v_2 = (1 + i)v_1$.

6.4.3. $\alpha = 1 + i$.

6.4.4. $q_2 = [1 \ \ -i \ 2i]$.

section 6.7

6.7.1. $\alpha_1 = \alpha_2 = -1$.

section 6.9

6.9.3. $1 \times 1,\ 2 \times 2,\ 3 \times 3,\ 3 \times 3,\ 3 \times 3,\ 4 \times 4$ blocks.

6.9.5. $(A - I)x_1 = o, (A - I)x_2 = x_1, (A - I)x_3 = x_2, X = [x_1 \ x_2 \ x_3]$.

$$X = \begin{bmatrix} 2 & -1 & -2 \\ 1 & 0 & 0 \\ 0 & -1 & -3 \end{bmatrix}, \quad X^{-1}AX = \begin{bmatrix} 1 & 1 & \\ & 1 & 1 \\ & & 1 \end{bmatrix}.$$

6.9.8.

$$X = \begin{bmatrix} \alpha & \beta & \gamma \\ & \alpha & \beta \\ & & \alpha \end{bmatrix}.$$

section 6.10

6.10.5. $\beta = 0$.

section 6.13

6.13.1. $D^2 - I = O, (R - I)^2(R + I) = O$.

6.13.2. $A^2 - 3A = O$.

6.13.3. $(A - \lambda I)^4 = O,\ (B - \lambda I)^3 = O,\ (C - \lambda I)^2 = O,\ \ D - \lambda I = O$.