6. The Algebraic Eigenproblem

6.1 How independent are linearly independent vectors?

If the three vectors a_1, a_2, a_3 in R^3 lie on one plane, then there are three scalars x_1, x_2, x_3, not all zero, so that

$$x_1 a_1 + x_2 a_2 + x_3 a_3 = o. \tag{6.1}$$

On the other hand, if the three vectors are not coplanar, then for any nonzero scalar triplet x_1, x_2, x_3

$$r = x_1 a_1 + x_2 a_2 + x_3 a_3 \tag{6.2}$$

is a nonzero vector.

To condense the notation we write $x = [x_1 \ x_2 \ x_3]^T$, $A = [a_1 \ a_2 \ a_3]$, and $r = Ax$. Vector r is nonzero but it can be made arbitrarily small by choosing an arbitrarily small x. To sidestep this possibility we add the restriction that $x^T x = 1$ and consider

$$r = Ax, \ x^T x = 1 \ \text{ or } \ r = (x^T x)^{-\frac{1}{2}} Ax. \tag{6.3}$$

Magnitude $\|r\|$ of residual vector r is an algebraic function of x_1, x_2, x_3, and if x can be found that makes r small, then the three vectors a_1, a_2, a_3 will be *nearly* coplanar. In this manner we determine how close the three vectors are to being linearly dependent. But first we must qualify what we mean by small.

Since colinearity is a condition on angle and not on length we assume that the three vectors are normalized, $\|a_i\| = 1, \ i = 1, 2, 3$, and consider r small when $\|r\|$ is small relative to 1. The columns of A are now all of length 1, and we write

$$\rho^2(x) = r^T r = x^T A^T A x, \ x^T x = 1 \ \text{ or } \ \rho^2(x) = \frac{x^T A^T A x}{x^T x}\ , \ x \neq o. \tag{6.4}$$

A basic theorem of analysis assures us that since $x^T x = 1$ and since $\rho(x)$ is a continuous function of x_1, x_2, x_3, $\rho(x)$ possesses both a minimum and a maximum with respect to x, which is also obvious geometrically. Figure 6.1 is drawn for $a_1 = [1\ 0]^T$, $a_2 = \sqrt{2}/2[1\ 1]^T$. Clearly the shortest and longest $r = x_1 a_1 + x_2 a_2$, $x_1^2 + x_2^2 = 1$ are the two angle bisectors $r_1 = \sqrt{2}/2(a_1 + a_2)$ and $r_2 = \sqrt{2}/2(a_1 - a_2)$, respectively, which we also notice to be orthogonal. We compute that $r_1^T r_1 = 1 - \sqrt{2}/2$ and $r_2^T r_2 = 1 + \sqrt{2}/2$.

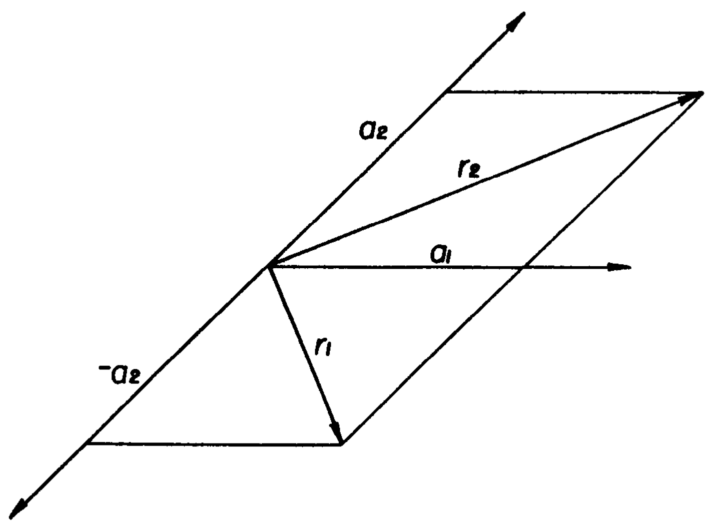

Fig. 6.1

A measure for the degree of the linear independence of the columns of A is provided by $\rho_{min} = \min_x \rho(x)$,

$$\rho_{min}^2 = \min_x \ x^T A^T A x, \ x^T x = 1 \tag{6.5}$$

or equally by

$$\rho_{min}^2 = \min_x \frac{x^T A^T A x}{x^T x}\ , \ x \neq o. \tag{6.6}$$

Clearly, what we argued for R^3 carries over to m vectors in R^n.

If the columns of A are linearly dependent, then $A^T A$ is positive semidefinite and $\rho_{min} = 0$. If the columns of A are linearly independent, then $A^T A$ is positive definite and $\rho_{min} > 0$.

In the case where the columns of A are orthonormal, $A^TA = I$ and $\rho_{min} = 1$. We shall prove that in this sense orthogonal vectors are most independent, ρ_{min} being at most 1.

Theorem 6.1. *If the columns of A are normalized and*

$$\rho^2_{min} = \min_x \frac{x^TA^TAx}{x^Tx} \ , \ x \neq o \tag{6.7}$$

then $0 \leq \rho_{min} \leq 1$.

Proof. Choose $x = e_1 = [1 \ 0 \ \dots \ 0]^T$ for which $r = a_1$ and $\rho^2(x) = a_1^Ta_1 = 1$. Since $\rho_{min} = \min \rho(x)$ it certainly does not exceed 1. End of proof.

Instead of normalizing the columns of A we could find $\min \rho(x)$ with A as given, then divide ρ_{min} by $\rho_{max} = \max_x \rho(x)$, or vice versa. Actually, it is common to measure the degree of linear independence of the columns of A with

$$\kappa^2 = \rho^2_{max}/\rho^2_{min} = \max_x x^TA^TAx/\min_x x^TA^TAx \ , \ x^Tx = 1 \tag{6.8}$$

where $\kappa = \kappa(A)$ is the *spectral condition number* of A. Now $1 \leq \kappa < \infty$. If A is orthogonal, then $\kappa = 1$, while if A is singular $\kappa = \infty$. A matrix with a large κ is said to be ill-conditioned, and that with a small κ, well-conditioned.

A necessary condition for x to minimize (maximize) $\rho^2(x)$ is that

$$\text{grad } \rho^2(x) = \frac{2(x^Tx)A^TAx - 2(x^TA^TAx)x}{(x^Tx)^2} = o \ , \ x \neq o \tag{6.9}$$

or

$$A^TAx = \frac{x^TA^TAx}{x^Tx}x \ , \ x \neq o. \tag{6.10}$$

In short

$$A^TAx = \rho^2 x \quad , \quad x \neq o \tag{6.11}$$

which is an algebraic eigenproblem. It consists of finding scalar ρ^2 and corresponding vector $x \neq o$ that satisfy the homogeneous vector equation $(A^TA - \rho^2 I)x = o$.

An extension of our minimization problem consists of finding the extrema of the ratio

$$\lambda(x) = \frac{x^TAx}{x^TBx} \ , \ x \neq o \tag{6.12}$$

of the two quadratic forms with $A = A^T$ and a positive definite and symmetric B. Here

$$\text{grad } \lambda(x) = \frac{2(x^T Bx)Ax - 2(x^T Ax)Bx}{(x^T Bx)^2} \ , \ x \neq o \tag{6.13}$$

so that

$$Ax = \left(\frac{x^T Ax}{x^T Bx}\right)Bx \quad x \neq o \tag{6.14}$$

or in short

$$Ax = \lambda Bx \tag{6.15}$$

which is the general symmetric algebraic eigenproblem.

The general eigenproblem is more prevalent in mathematical physics then the special $B = I$ case, and we shall deal with the general case in due course. Meanwhile we are satisfied that, at least formally, the general symmetric eigenproblem can be reduced to the special symmetric form by the factorization $B = LL^T$ and the substitution $x' = L^T x$ that turns $Ax = \lambda Bx$ into $L^{-1}AL^{-T}x' = \lambda x'$.

exercises

6.1.1. Let

$$x = \alpha_1 \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \alpha_1^2 + \alpha_2^2 = 1$$

for variable scalars α_1, α_2. Use the Lagrange multipliers method to find the extrema of $x^T x$. In this you set up the Lagrange objective function

$$\phi(\alpha_1, \alpha_2) = x^T x - \lambda(\alpha_1^2 + \alpha_2^2 - 1)$$

and obtain the critical α's and multiplier λ from $\partial\phi/\partial\alpha_1 = 0, \partial\phi/\partial\alpha_2 = 0, \partial\phi/\partial\lambda = 0$.

6.1.2. Is $x = [1 \ 2 \ -1]^T$ an eigenvector of matrix

$$A = \begin{bmatrix} 1 & -1 & 1 \\ -2 & -1 & 0 \\ 2 & 1 & 2 \end{bmatrix}?$$

If yes, compute the corresponding eigenvalue.

6.1.3. Is $x = [1\ 1]^T$ an eigenvector of the general eigenproblem

$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} x = \lambda \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} x?$$

If yes, find the corresponding eigenvalue.

6.2 Diagonal and triangular matrices

There must be something very special and useful about vector x that turns Ax into a scalar times x, and in this chapter we shall give a thorough consideration to the remarkable algebraic eigenproblem $Ax = \lambda x$. Not only for the symmetric A, that has its origin in the minimization of the ratio of two quadratic forms, but also for the more inclusive case where A is merely square.

We may look at the algebraic eigenproblem $Ax = \lambda x$ geometrically, the way we did in Chapter 4, as the search for those vectors x in R^n for which the linear map Ax is colinear with x, with $|\lambda| = \|Ax\|/\|x\|$, or we may write it as $(A - \lambda I)x = o$ and look at the problem algebraically as the search for scalars λ that render matrix $A - \lambda I$ singular, and then the computation of the corresponding nullspace of $A - \lambda I$.

Definition. *Scalar λ that renders $B(\lambda) = A - \lambda I$ singular is an eigenvalue of A. Any nonzero vector x for which $B(\lambda)x = o$ is an eigenvector of A corresponding to eigenvalue λ.*

Because the eigenproblem is homogeneous, if $x \neq o$ is an eigenvector of A, then so is $\alpha x,\ \alpha \neq 0$.

As with linear systems of equations, so too here, diagonal and triangular matrices are the easiest to deal with. Consider

$$A = \begin{bmatrix} 0 & & \\ & 1 & \\ & & 2 \end{bmatrix}. \tag{6.16}$$

We readily verify that A has exactly three eigenvalues, which we write in ascending order of magnitude as

$$\lambda_1 = 0,\ \lambda_2 = 1,\ \lambda_3 = 2 \tag{6.17}$$

affecting

$$A - \lambda_1 I = \begin{bmatrix} 0 & & \\ & 1 & \\ & & 2 \end{bmatrix}, \; A - \lambda_2 I = \begin{bmatrix} -1 & & \\ & 0 & \\ & & 1 \end{bmatrix}, \; A - \lambda_3 I = \begin{bmatrix} -2 & & \\ & -1 & \\ & & 0 \end{bmatrix} \tag{6.18}$$

which are all diagonal matrices of type 0 and hence singular. Corresponding to $\lambda_1 = 0,\ \lambda_2 = 1,\ \lambda_3 = 2$ are the three eigenvectors $x_1 = e_1,\ x_2 = e_2,\ x_3 = e_3$, that we observe to be orthogonal.

Matrix

$$A = \begin{bmatrix} 0 & & \\ & 1 & \\ & & 1 \end{bmatrix} \tag{6.19}$$

also has three eigenvalues $\lambda_1 = 0,\ \lambda_2 = 1,\ \lambda_3 = 1$, but the peculiarity of this eigenproblem consists in the fact that eigenvalue $\lambda = 1$ repeats. Corresponding to $\lambda_1 = 0$ is the unique (up to a nonzero constant multiple) eigenvector $x_1 = e_1$, but corresponding to $\lambda_2 = \lambda_3 = 1$ is any vector in the two-dimensional subspace spanned by e_2 and e_3. Any nonzero $x = \alpha_2 e_2 + \alpha_3 e_3$ is an eigenvector of A corresponding to $\lambda = 1$. The vector is orthogonal to $x_1 = e_1$, whatever α_2 and α_3 are, and we may arbitrarily assign $x_2 = e_2$ to be the eigenvector corresponding to $\lambda_2 = 1$ and $x_3 = e_3$ to be the eigenvector corresponding to $\lambda_3 = 1$ so as to have a set of three orthonormal eigenvectors.

The eigenvalues of a triangular matrix are also written down by inspection, but eigenvector extraction needs computation. A necessary and sufficient condition that a triangular matrix be singular is that it be of type 0, that is, that it has at least one zero diagonal entry. Eigenproblem

$$\begin{bmatrix} 1-\lambda & & \\ 1 & 2-\lambda & \\ 1 & 1 & 3-\lambda \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \; (A - \lambda I)x = 0 \tag{6.20}$$

for instance, is readily seen to have the three distinct eigenvalues $\lambda_1 = 1,\ \lambda_2 = 2,\ \lambda_3 = 3$; *the eigenvalues of a triangular matrix are its diagonal entries.* Corresponding to the three eigenvalues we compute the three eigenvectors

$$x_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \; x_2 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \quad x_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{6.21}$$

that are not orthogonal as in the previous examples, but are nonetheless checked to be linearly independent.

An instance of a tridiagonal matrix with repeating eigenvalues and a multidimensional nullspace for the singular $A - \lambda I$ is

$$A = \begin{bmatrix} 1 & & 3 \\ & 1 & -4 \\ & & 2 \end{bmatrix} \tag{6.22}$$

that is readily verified to have the three eigenvalues $\lambda_1 = 1, \lambda_2 = 1, \lambda_3 = 2$. Taking first the largest eigenvalue $\lambda_3 = 2$ we obtain all its eigenvectors as $x_3 = \alpha_3[3 - 4\ 1]^T\ \alpha_3 \neq 0$, and we elect $x_3 = [3\ \ -4\ 1]^T$ to be the sole eigenvector for $\lambda_3 = 2$.

For $\lambda = 1$ we have

$$A - I = \begin{bmatrix} 0 & & 3 \\ & 0 & -4 \\ & & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0. \tag{6.23}$$

The appearance of two zeroes on the diagonal of $A - \lambda I$ does not necessarily mean a two-dimensional nullspace, but here it does. Indeed, $x = \alpha_1[1\ 0\ 0]^T + \alpha_2[0\ 1\ 0]^T$, and for any choice of α_1 and α_2 that does not render it zero, x is an eigenvector corresponding to $\lambda = 1$, linearly independent of x_3. We may choose any two linearly independent vectors in this nullspace, say $x_1 = [1\ 0\ 0]^T$ and $x_2 = [0\ 1\ 0]^T$ and assign them to be the eigenvectors of $\lambda_1 = 1$ and $\lambda_2 = 1$, so as to have a set of three linearly independent eigenvectors.

On the other hand

$$A = \begin{bmatrix} 1 & 2 & 3 \\ & 1 & -4 \\ & & 2 \end{bmatrix} \tag{6.24}$$

that has the same three eigenvalues $\lambda_1 = 1, \lambda_2 = 1, \lambda_3 = 2$, has eigenvector $x_3 = [3\ \ -4\ 1]^T$ corresponding to $\lambda_3 = 2$ as before, and $x_1 = [1\ 0\ 0]^T$ corresponding to $\lambda_1 = \lambda_2 = 1$. The nullspace of $A - I$ is here only one-dimensional, and the matrix has only *two* linearly independent eigenvectors.

One more instructive triangular example before we move on to the full matrix. Matrix

$$A = \begin{bmatrix} 1 & & \\ 1 & 1 & \\ 1 & 1 & 1 \end{bmatrix} \tag{6.25}$$

has three equal eigenvalues (or one eigenvalue of multiplicity 3) $\lambda_1 = \lambda_2 = \lambda_3 = 1$. To compute the corresponding eigenvectors we solve

$$(A-I)x = \begin{bmatrix} 0 & & \\ 1 & 0 & \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{6.26}$$

and obtain from the system only the one eigenvector $x = e_3$.

It is interesting that an $(n \times n)$ matrix can have n zero eigenvalues and yet be nonzero.

exercises

6.2.1. Compute all eigenvalues and eigenvectors of

$$A = \begin{bmatrix} 0 & 1 & & \\ & 0 & & \\ & & 0 & 1 \\ & & & 0 \end{bmatrix}.$$

6.2.2. Compute all eigenvalues and eigenvectors of

$$A = \begin{bmatrix} 1 & 1 & 1 \\ & -2 & 2 \\ & & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 & 0 \\ & 1 & 2 \\ & & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & 2 \\ & 1 & -3 \\ & & 2 \end{bmatrix}.$$

6.2.3. Give an example of two different 2×2 upper-triangular matrices with the same eigenvalues and eigenvectors. Given that the eigenvectors of A are x_1, x_2, x_3,

$$A = \begin{bmatrix} 1 & \alpha & \beta \\ & 2 & \gamma \\ & & 3 \end{bmatrix}, \ x_1 = \begin{bmatrix} 1 \\ \\ \\ \end{bmatrix}, \ x_2 = \begin{bmatrix} 1 \\ 1 \\ \\ \end{bmatrix}, \ x_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

find α, β, γ.

6.2.4. Show that if matrix A is such that $A^3 - 2A^2 - A + I = O$, then zero is not an eigenvalue of $A + I$.

6.2.5. Matrix $A = 2u_1u_1^T - 3u_2u_2^T$, $u_1^Tu_1 = u_2^Tu_2 = 1, u_1^Tu_2 = u_2^Tu_1 = 0$, is of rank 2. Find β_1 and β_2 so that $A^3 + \beta_2A^2 + \beta_1A = O$.

6.3 The characteristic equation

Our chief conclusion from the previous section is that a triangular (diagonal) matrix of order n has n eigenvalues, some isolated and some repeating. Corresponding to an isolated eigenvalue (eigenvalue of multiplicity one) we computed in all instances a unique (up to length and sense) eigenvector, but for eigenvalues of multiplicity greater than one we occasionally found multiple eigenvectors.

In this section we shall mathematically consolidate these observations and extend them to any square matrix.

Computation of the eigenvalues and eigenvectors of a nontriangular matrix is a considerably harder task than that of a triangular matrix. A necessary and sufficient condition that the homogeneous system $(A - \lambda I)x = o$ has a nontrivial solution x, is that matrix $B(\lambda) = A - \lambda I$ be singular, or equivalent to a triangular matrix of type 0. So, we shall reduce $B(\lambda)$ by elementary operations to triangular form and determine λ that makes the triangular matrix of that type. The operations are elementary but they involve parameter λ and are therefore algebraic rather than numerical. In doing that we shall be careful to avoid λ dependent pivots lest they be zero.

A triangular matrix is of type 0 if and only if the product of its diagonal entries —the determinant of the matrix— is zero, and hence the problem of finding the λ's that singularize $B(\lambda)$ is translated into the single *characteristic equation*

$$\det\,(A - \lambda I) = 0 \tag{6.27}$$

that needs be solved for λ.

The characteristic equation may be written with some expansion rule for $\det\,(B(\lambda))$ or it may be obtained as the end product of a sequence of elementary row and column operations that reduce $B(\lambda)$ to triangular form. For example:

$$A - \lambda I = \begin{bmatrix} 2-\lambda & -1 \\ -1 & 1-\lambda \end{bmatrix} \overset{row}{\longrightarrow} \begin{bmatrix} -1 & 1-\lambda \\ 2-\lambda & -1 \end{bmatrix} \overset{row}{\longrightarrow} \begin{bmatrix} -1 & 1-\lambda \\ 0 & (2-\lambda)(1-\lambda)-1 \end{bmatrix} \tag{6.28}$$

and the characteristic equation of A is

$$\det\,(A - \lambda I) = (2-\lambda)(1-\lambda) - 1 = \lambda^2 - 3\lambda + 1 = 0. \tag{6.29}$$

Notice that since a row interchange was performed, to have the formal det $(B(\lambda))$ the diagonal product is multiplied by -1.

The elementary operations done on $A - \lambda I$ to bring it to equivalent upper-triangular form could have been performed without row interchanges, and still without λ dependent pivots:

$$\begin{bmatrix} 2-\lambda & -1 \\ -1 & 1-\lambda \end{bmatrix} \rightarrow \begin{bmatrix} 2 & -1-\lambda(1-\lambda) \\ -1 & 1-\lambda \end{bmatrix} \rightarrow \begin{bmatrix} 2 & -1-\lambda(1-\lambda) \\ 0 & (1-\lambda)-\frac{1}{2}(1+\lambda(1-\lambda)) \end{bmatrix} \tag{6.30}$$

resulting in the same characteristic equation.

Two real roots,

$$\lambda_1 = (3-\sqrt{5})/2 \quad , \quad \lambda_2 = (3+\sqrt{5})/2 \tag{6.31}$$

the two eigenvalues of matrix A, are extracted from the characteristic equation, with the two corresponding eigenvectors

$$x_1 = \begin{bmatrix} 2 \\ 1+\sqrt{5} \end{bmatrix}, \quad x_2 = \begin{bmatrix} 2 \\ 1-\sqrt{5} \end{bmatrix} \tag{6.32}$$

that we observe to be orthogonal.

Generally, for the 2×2

$$B(\lambda) = \begin{bmatrix} A_{11}-\lambda & A_{12} \\ A_{21} & A_{22}-\lambda \end{bmatrix} \tag{6.33}$$

det $(B(\lambda)) = 0$ expands into

$$(A_{11}-\lambda)(A_{22}-\lambda) - A_{12}A_{21} = \lambda^2 - \lambda(A_{11}+A_{22}) + A_{11}A_{22} - A_{12}A_{21} = 0 \tag{6.34}$$

which is a polynomial equation of degree two and hence has two roots, real or complex. If λ_1 and λ_2 are the two roots of the characteristic equation, then it may be written as

$$\det\,(A - \lambda I) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) = \lambda^2 - \lambda(\lambda_1 + \lambda_2) + \lambda_1\lambda_2 \tag{6.35}$$

and $\lambda_1 + \lambda_2 = A_{11} + A_{22} = \mathrm{tr}(A)$, and $\lambda_1\lambda_2 = A_{11}A_{22} - A_{12}A_{21} = \det\,(A) = \det\,(B(0))$.

As a numerical example to a 3×3 eigenproblem we undertake to carry out the elementary operations on $B(\lambda)$,

$$\begin{bmatrix} 0-\lambda & 1 & 0 \\ 0 & 0-\lambda & 1 \\ 4 & -17 & 8-\lambda \end{bmatrix} \xrightarrow{row} \begin{bmatrix} 4 & -17 & 8-\lambda \\ 0 & -\lambda & 1 \\ -\lambda & 1 & 0 \end{bmatrix} \xrightarrow{row} \begin{bmatrix} 4 & -17 & 8-\lambda \\ 0 & -\lambda & 1 \\ 0 & \frac{1}{4}(4-17\lambda) & \frac{1}{4}\lambda(8-\lambda) \end{bmatrix}$$

$$\xrightarrow{column} \begin{bmatrix} 4 & 8-\lambda & -17 \\ 0 & 1 & -\lambda \\ 0 & \frac{1}{4}\lambda(8-\lambda) & \frac{1}{4}(4-17\lambda) \end{bmatrix} \xrightarrow{row} \begin{bmatrix} 4 & 8-\lambda & -17 \\ 0 & 1 & -\lambda \\ 0 & 0 & \frac{1}{4}(-\lambda^3+8\lambda^2-17\lambda+4) \end{bmatrix} \quad (6.36)$$

and the eigenvalues of $A = B(0)$ are the roots of the characteristic equation

$$-\lambda^3 + 8\lambda^2 - 17\lambda + 4 = 0. \quad (6.37)$$

The same can be accomplished by elementary row operations only:

$$\begin{bmatrix} -\lambda & 1 & \\ & -\lambda & 1 \\ 4 & -17 & 8-\lambda \end{bmatrix} \rightarrow \begin{bmatrix} 4 & -17 & 8-\lambda \\ & -\lambda & 1 \\ -4\lambda & 4 & \end{bmatrix} \rightarrow \begin{bmatrix} 4 & -17 & \\ & -\lambda & 1 \\ & 4-17\lambda & \lambda(8-\lambda) \end{bmatrix} \rightarrow$$

$$\begin{bmatrix} 4 & -17 & \\ & -\lambda & 1 \\ & 4 & \lambda(8-\lambda)-17 \end{bmatrix} \rightarrow \begin{bmatrix} 4 & -17 & \\ & 4 & \lambda(8-\lambda)-17 \\ & -4\lambda & 4 \end{bmatrix}$$

$$\rightarrow \begin{bmatrix} 4 & -17 & \\ & 4 & -\lambda^2+8\lambda-17 \\ & & -\lambda^3+8\lambda^2-17\lambda+4 \end{bmatrix}. \quad (6.38)$$

Generally, for a 3×3 matrix

$$\det\,(A-\lambda I) = -\lambda^3 + \lambda^2(A_{11}+A_{22}+A_{33}) - \lambda(\begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} + \begin{vmatrix} A_{11} & A_{13} \\ A_{31} & A_{33} \end{vmatrix} + \begin{vmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{vmatrix})$$

$$+\det\,(A) = 0 \quad (6.39)$$

which is a polynomial equation in λ of degree 3 that has three roots, at least one of which is real. If $\lambda_1, \lambda_2, \lambda_3$ are the three roots of the characteristic equation, then it may be written as

$$(\lambda_1-\lambda)(\lambda_2-\lambda)(\lambda_3-\lambda) = -\lambda^3+\lambda^2(\lambda_1+\lambda_2+\lambda_3)-\lambda(\lambda_1\lambda_2+\lambda_2\lambda_3+\lambda_3\lambda_1)+\lambda_1\lambda_2\lambda_3 = 0 \quad (6.40)$$

and

$$\lambda_1+\lambda_2+\lambda_3 = A_{11}+A_{22}+A_{33} = \mathrm{tr}(A),\ \ \lambda_1\lambda_2\lambda_3 = \det\,(A) = \det\,(B(0)). \quad (6.41)$$

What we did for the 2×2 and 3×3 matrices results (a formal proof to this is given in Sec. 6.7) in the nth degree polynomial characteristic equation

$$\det\,(A-\lambda I) = p_n(\lambda) = (-\lambda)^n + a_{n-1}(-\lambda)^{n-1} + \cdots + a_0 = 0 \quad (6.42)$$

for $A = A(n \times n)$, and if matrix A is real, so then are coefficients $a_{n-1}, a_{n-2}, \ldots, a_0$ of the equation.

There are some highly important theoretical conclusions that can be immediately drawn from the characteristic equation of real matrices:

1. According to the fundamental theorem of algebra, a polynomial equation of degree n has at least one solution, real or complex, of multiplicity n, and at most n distinct solutions.

2. A polynomial equation of odd degree has at least one real root.

3. The complex roots of a polynomial equation with real coefficients appear in conjugate pairs. If $\lambda = \alpha + i\beta$ is a root of the equation, then so is the conjugate $\overline{\lambda} = \alpha - i\beta$.

Real polynomial equations of odd degree have at least one real root since if n is odd, then $p_n(\infty) = -\infty$ and $p_n(-\infty) = \infty$, and by the reality and continuity of $p_n(\lambda)$ there is at least one real $-\infty < \lambda < \infty$ for which $p_n(\lambda) = 0$.

We prove statement 3 on the complex roots of the real equation

$$p_n(\lambda) = (-\lambda)^n + (-\lambda)^{n-1} a_{n-1} + \cdots + a_0 = 0 \tag{6.43}$$

by writing them as

$$-\lambda = \alpha + i\beta = |\lambda| e^{i\theta} = |\lambda|(\cos\theta + i\sin\theta), \quad i^2 = -1. \tag{6.44}$$

Then, since $e^{in\theta} = \cos n\theta + i \sin n\theta$ and since the coefficients of the equation are real, $p_n(\lambda) = 0$ separates into the real and imaginary parts

$$|\lambda|^n \cos n\theta + a_{n-1} |\lambda|^{n-1} \cos(n-1)\theta + \cdots + a_0 = 0 \tag{6.45}$$

and

$$|\lambda|^n \sin n\theta + a_{n-1} |\lambda|^{n-1} \sin(n-1)\theta + \cdots + a_1 \sin\theta = 0 \tag{6.46}$$

respectively, and because $\cos(-\theta) = \cos\theta$ and $\sin(-\theta) = -\sin\theta$, the equation is satisfied by both $\lambda = |\lambda| e^{i\theta}$ and $\overline{\lambda} = |\lambda| e^{-i\theta}$.

Root λ_1 is of multiplicity m if $\lambda - \lambda_1$ can be factored exactly m times out of $p_n(\lambda)$. Then not only is $p_n(\lambda) = 0$, but also

$$dp_n(\lambda)/d\lambda = 0, \ldots, d^{m-1} p_n(\lambda)/d\lambda^{m-1} = 0 \tag{6.47}$$

at $\lambda = \lambda_1$. Each is a polynomial equation satisfied by λ_1 and $\overline{\lambda_1}$, and hence if λ_1 has multiplicity m so does $\overline{\lambda_1}$.

Finding the eigenvalues of an $n \times n$ matrix entails the solution of an algebraic equation of the nth degree. A quadratic equation is solved algebraically in terms of the coefficients, but already the cubic equation can become difficult. It is best treated numerically by an iterative method such as bisection or that of Newton-Raphson, or any other root enhancing method. In any event an algebraic solution is possible only for polynomial equations of degree less than five. Equations of degree five or higher must be solved by an iterative approximation algorithm. Unlike systems of linear equations, the eigenvalue problem has no finite step solution.

Good numerical polynomial equation solvers yield not only the roots but also their multiplicities. As we shall see the multiplicity of λ has an important bearing on the dimension of the corresponding eigenvector space, and has significant consequences for the numerical solution of the eigenproblem.

Figure 6.2 traces $p_3(\lambda)$ against λ, with $p_3(\lambda) = 0$ having three real roots, one isolated and one double. At $\lambda = \lambda_2 = \lambda_3$ both $p_3(\lambda) = 0$ and $dp_3(\lambda)/d\lambda = 0$, and the Newton-Raphson root-finding method converges linearly to this root whereas it converges quadratically to λ_1. Since close to $\lambda = \lambda_2 = \lambda_3$, $p_3(\lambda)$ has the same sign on both sides of the root, bisection root finding methods must also be carried out here with extra precaution.

On top of this, because a multiple root is just the limiting case between two real roots and no real root, small changes in the coefficients of the characteristic equation may cause drastic changes in these roots. Consider for instance

$$\lambda^2 - 2\lambda + 1 = 0\ ,\quad \lambda^2 - 2.1\lambda + 0.9 = 0\ ,\quad \lambda^2 - 1.9\lambda + 1.1 = 0. \tag{6.48}$$

The first equation has a repeating real root $\lambda_1 = \lambda_2 = 1$, the second a pair of well separated roots $\lambda_1 = 0.6$, $\lambda_2 = 1.5$, while the third equation has two complex conjugate roots $\lambda = 0.95 \pm 0.44i$.

In contrast with the eigenvalues, the coefficients of the characteristic equation can be obtained in a finite number of elementary operations, and we shall soon discuss algorithms

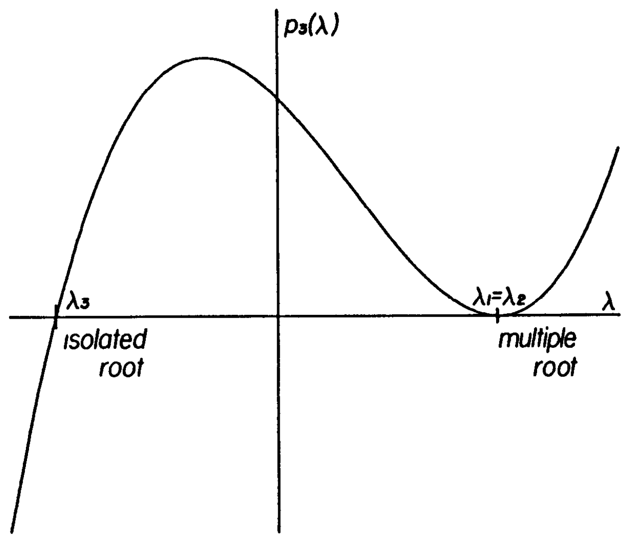

Fig. 6.2

that do that. Writing the characteristic equation of a large matrix in full is generally unrealistic, but for any given value of λ, det $(A - \lambda I)$ can often be computed at reasonable cost by Gauss elimination.

The set of all eigenvalues is the *spectrum* of matrix A, and we shall occasionally denote it by $\lambda(A)$.

exercises

6.3.1. Find λ so that matrix

$$B(\lambda) = \begin{bmatrix} 1-\lambda & 1+\lambda \\ -1+2\lambda & -1-\lambda \end{bmatrix}$$

is singular. Bring the matrix first to equivalent lower-triangular form but be careful not to use a λ-containing pivot.

6.3.2. Write the characteristic equation of

$$C = \begin{bmatrix} & & \alpha_0 \\ -1 & & \alpha_1 \\ & -1 & \alpha_2 \end{bmatrix}.$$

6.3.3. Show that if the characteristic equation of $A = A(3 \times 3)$ is written as

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

then

$$\begin{aligned}\alpha_2 &= trace(A)\\ \alpha_1 &= \frac{1}{2}(-\alpha_2 trace(A) + trace(A^2))\\ \alpha_0 &= \frac{1}{3}(-\alpha_2 trace(A) - \alpha_1 trace(A^2) + trace(A^3)).\end{aligned}$$

6.3.4. What are the conditions on the entries of $A = A(2 \times 2)$ for it to have two equal eigenvalues?

6.3.5. Consider the scalar function $f(A) = A_{11}^2 + 2A_{12}A_{21} + A_{22}^2$ of $A = A(2 \times 2)$. Express it in terms of the two eigenvalues λ_1 and λ_2 of A.

6.3.6. Compute all eigenvalues and eigenvectors of

$$A = \begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & i \\ i & -1 \end{bmatrix}.$$

6.3.7. Write all eigenvalues and eigenvectors of

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}.$$

6.3.8. Fix α_1 and α_2 so that the eigenvalues of

$$A = \begin{bmatrix} \alpha_1 & 1 \\ 1 & \alpha_2 \end{bmatrix}$$

are the prescribed $\lambda_1 = 1$ and $\lambda_2 = 3$.

6.3.9 For what values of real α is $\lambda(A)$ real?

$$A = \begin{bmatrix} 1 & \alpha i \\ \alpha i & 0 \end{bmatrix}.$$

6.3.10. Fix the value of α so that x is an eigenvector of A.

$$x = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \quad A = \begin{bmatrix} 2 & -1 & \alpha \\ 2 & -3 & -1 \\ \alpha & -2 & -1 \end{bmatrix}.$$

What is the corresponding eigenvalue?

6.3.11. Show that vector $u, u^T u = 1$ is an eigenvector of $A = I + uu^T$. What is the corresponding eigenvalue?

6.3.12. Let u and v be two orthogonal unit vectors, $u^T u = v^T v = 1, u^T v = v^T u = 0$. Show that v is an eigenvector of $A = I + uu^T$. What is the corresponding eigenvalue?

6.3.13. Let matrix A be such that $A^2 - I = O$. What is vector $x' = Ax + x \neq o$, where vector x is arbitrary?

6.3.14. Show that if $A^2 x - 2Ax + x = o$ for some $x \neq o$, then $\lambda = 1$ is an eigenvalue of A. Is x the corresponding eigenvector?

6.3.15. Show that if for real A and real $x \neq o$, $A^2 x + x = o$, then $\lambda = \pm i$ are two eigenvalues of A. What are the corresponding eigenvectors?

6.3.16. Show that the eigenvectors of circulant matrix

$$C = \begin{bmatrix} \alpha_0 & \alpha_1 & \alpha_2 \\ \alpha_2 & \alpha_0 & \alpha_1 \\ \alpha_1 & \alpha_2 & \alpha_0 \end{bmatrix}$$

are of the form $x = [1 \ \epsilon \ \epsilon^2]^T$, where $\epsilon^3 = 1$. What are the corresponding eigenvalues?

6.3.17. Let $A = A(n \times n)$ and $B = B(n \times n)$. Show that the characteristic polynomial of

$$C = \begin{bmatrix} A & B \\ B & A \end{bmatrix}$$

is the product of the characteristic polynomials of $A + B$ and $A - B$. Hint: Perform block elementary row and column operations on

$$\begin{bmatrix} A - \lambda I & B \\ B & A - \lambda I \end{bmatrix}.$$

6.3.18. Solve the generalized eigenproblems

$$\begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} x = \lambda \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} x, \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} x = \lambda \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} x$$

$$\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} x = \lambda \begin{bmatrix} 3 & 1 \\ 6 & 2 \end{bmatrix} x, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} x = \lambda \begin{bmatrix} -1 & -1 \\ 2 & 2 \end{bmatrix} x$$

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

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} x = \lambda \begin{bmatrix} 1 & \\ & -1 \end{bmatrix} x.$$

6.4 Complex vector space C^n

Because real matrices can have complex eigenvalues and eigenvectors, we cannot escape discussing vector space C^n, the space of vectors with n *complex* components. Equality of vectors, addition of vectors, and the multiplication of a vector by a scalar is the same in C^n as it is in R^n, but the inner or scalar product in C^n needs to be changed.

In C^n a^Tb is generally complex, and a^Ta can be zero even if $a \neq o$. For instance, if $a = [1\ i]^T,\ b = [i\ 1]^T$, then $a^Tb = 2i$ and $a^Ta = 1 - 1 = 0$. For the inner product theorems of R^n to extend to C^n we must ensure that the inner product of a nonzero vector by itself is a positive real number. Recalling that if $\zeta = \alpha + i\beta$ is a complex number and $\overline{\zeta} = \alpha - i\beta$ is its conjugate, then $\overline{\zeta}\zeta = |\zeta|^2 = \alpha^2 + \beta^2$ and we introduce the

Definition. *Let $u \in C^n$ be of the form $u = a+ib,\ a, b \in R^n$. Then vector $u^H = a^T - ib^T$ is the conjugate transpose of u. Similarly if $C = A+iB$, then $C^H = A^T - iB^T$ is the conjugate transpose of matrix C.*

Now $u^Hu = a^Ta + b^Tb$ is a real positive number that vanishes only if $u = o$

When $C = C^H$, that is, when $A = A^T$ and $B = -B^T$, matrix C is said to be *Hermitian.*

Theorem 6.2.

1. *$u^Hv = \overline{v^Hu},\ |u^Hv| = |\overline{v^Hu}|$*
2. *$u^Hv + v^Hu$ is real*
3. *$\overline{AB} = \overline{A}\ \overline{B}$*
4. *$(\alpha A)^H = \overline{\alpha}A^H$*
5. *$(AB)^H = B^HA^H$*
6. *If $A = A^H$, then $u^HAv = \overline{v^HAu}$*

7. If $A = A^H$, *then* $u^H Au$ *is real*

8. $\|u\| = |u^H u|^{1/2}$ *has the three norm properties:*

8.1 $\|u\| > 0$ *if* $u \neq o$, *and* $\|u\| = 0$ *if* $u = o$

8.2 $\|\alpha u\| = |\alpha| \ \|u\|$

8.3 $\|u + v\| \leq \|u\| + \|v\|$

Proof. Left as an exercise. Notice that $|\ \ |$ means here modulus.

Definition. *Let* u *and* v *be in* C^n. *Vector* u *is a unit vector if* $\|u\| = (u^H u)^{\frac{1}{2}} = 1$. *Vectors* u *and* v *are orthogonal if* $u^H v = \overline{v^H u} = 0$. *If in addition* $\|u\| = \|v\| = 1$, *then* u *and* v *are orthonormal.*

In C^n one must be careful to distinguish between $u^H v$ and $u^T v$ that coexist in the same space. The Cauchy-Schwarz inequality in C^n remains

$$|u^H v| \leq (u^H u)^{\frac{1}{2}} (v^H v)^{\frac{1}{2}} \tag{6.49}$$

except that $|\ \ |$ stands now for modulus.

Example. To compute $x \in C^2$ orthogonal to given vector $a = [2 + i \ \ 3 - i]^T$ we write $a = u + iu'$ and $x = v + iv'$, and have that

$$a^H x = (u^T v + u'^T v') + i(-u'^T v + u^T v') \tag{6.50}$$

so that $a^H x = 0$ separates into

$$u^T v + u'^T v' = 0, \ -u'^T v + u^T v' = 0 \tag{6.51}$$

or in matrix vector form

$$\begin{bmatrix} u^T \\ -u'^T \end{bmatrix} v \ + \ \begin{bmatrix} u'^T \\ u^T \end{bmatrix} v' \ = o. \tag{6.52}$$

The two real vectors u and u' are linearly independent. In case they are not, the given complex vector a becomes a complex number times a real vector and x is real. The matrix multiplying v is invertible, and for the given numbers the condition $a^H x = 0$ becomes

$$v \ = \begin{bmatrix} 1 & 2 \\ -1 & -1 \end{bmatrix} v' \ , \ v = Kv' \tag{6.53}$$

and $x=(K+iI)v',\ v'\neq o$. One solution from among the many is obtained with $v'=[1\ 0]^T$ as $x=[1+i\ \ -1]^T$.

Example: To show that

$$a+ib=\begin{bmatrix}1\\-1\end{bmatrix}+i\begin{bmatrix}1\\1\end{bmatrix}\ \text{and}\ a'+ib'=\begin{bmatrix}-1\\1\end{bmatrix}+i\begin{bmatrix}1\\1\end{bmatrix} \tag{6.54}$$

are linearly dependent. Separation of the real and imaginary parts of

$$(\alpha+i\beta)(a+ib)+(\alpha'+i\beta')(a'+ib')=o \tag{6.55}$$

results in the two real systems

$$\begin{aligned}\alpha a-\beta b+\alpha' a'-\beta' b'&=o\\ \beta a+\alpha b+\beta' a'+\alpha' b'&=o\end{aligned} \tag{6.56}$$

that we write in matrix vector form as

$$\begin{bmatrix}1&-1&-1&-1\\-1&-1&1&-1\\1&1&-1&1\\-1&1&1&1\end{bmatrix}\begin{bmatrix}\alpha\\\beta\\\alpha'\\\beta'\end{bmatrix}=o. \tag{6.57}$$

System (6.57) is solved by $\alpha'=\alpha,\ \beta'=-\beta$.

In the same manner we show that $[1\ 2]^T+i[1\ 1]^T$ and $[-1\ 1]^T+i[-2\ 1]^T$ are linearly independent.

Theorem 6.3. *Let u_1 be a given vector in C^n. Then there are $n-1$ nonzero vectors in C^n orthogonal to u_1.*

Proof. Write $u_1=a_1+ib_1$ and $u_2=x+ix'$. The condition

$$\begin{aligned}u_2^H u_1&=(x^T-i{x'}^T)(a_1+ib_1)\\&=(a_1^Tx+b_1^Tx')+i(b_1^Tx-a_1^Tx')=0\end{aligned} \tag{6.58}$$

separates into

$$\begin{bmatrix}a_1^T&b_1^T\\b_1^T&-a_1^T\end{bmatrix}\begin{bmatrix}x\\x'\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix} \tag{6.59}$$

which are two homogeneous equations in $2n$ unknowns —the components of x and x'. Take any nontrivial solution to the system and let $u_2 = a_2 + ib_2$, where $a_2 = x$ and $b_2 = x'$. Write again $u_3 = x + ix'$ and solve $u_1^H u_3 = u_2^H u_3 = 0$ for x and x'. The two orthogonality conditions separate now into

$$\begin{bmatrix} a_1^T & b_1^T \\ b_1^T & -a_1^T \\ a_2^T & b_2^T \\ b_2^T & -a_2^T \end{bmatrix} \begin{bmatrix} x \\ x' \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{6.60}$$

which are four homogeneous equations in $2n$ unknowns. Take any nontrivial solution to the system and set $u_3 = a_3 + ib_3$ where $a_3 = x$ and $b_3 = x'$.

Suppose u_{k-1} orthogonal vectors have been generated this way in C^n. To compute u_k orthogonal to all the $k-1$ previously computed vectors we need solve $2(k-1)$ homogeneous equations in $2n$ unknowns. Since the number of unknowns is greater than the number of equations for $k = 2, 3, \ldots, n$ there is a nontrivial solution to the system, and consequently a nonzero u_k, for any $k \leq n$. End of proof.

Definition. *Square complex matrix U with orthonormal columns is called unitary. It is characterized by $U^H U = I$, or $U^H = U^{-1}$.*

exercises

6.4.1. Do $v_1 = [1\ i\ 0]^T$, $v_2 = [0\ 1\ -i]^T$, $v_3 = [1\ i\ i]^T$ span C^3? Find complex scalars $\alpha_1, \alpha_2, \alpha_3$ so that $\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 = [2+i\ 1-i\ 3i]^T$.

6.4.2. Are $v_1 = [1\ i\ -i]^T$ and $v_2 = [1\ -i\ i]^T$ linearly independent? What about $v_1 = [2-3i\ 1+i\ 3-i]^T$ and $v_2 = [5-i\ 2i\ 4+2i]^T$?

6.4.3. Fix α so that vectors $u \in C^3$ and $v \in C^3$ are orthogonal, namely so that $u^H v = 0$. $u = [1+i\ 1-i\ 2i], v = [1-2i\ 2+3i\ \alpha]$.

6.4.4. Vector space $V : v_1 = [1\ i\ 0], v_2 = [1\ 0\ i]$, is a two-dimensional subspace of C^3. Use the Gram-Schmidt orthogonalization method to write an orthogonal basis for V. Hint: Write $q_1 = v_1, q_2 = v_2 + \alpha v_1$, and determine α by the condition that $q_1^H q_2 = 0$.

6.4.5. Show that given complex vector $x = u + iv$, $u^T u \neq v^T v$, complex scalar $\zeta = \alpha + i\beta$ can be found so that $\zeta x = u' + iv'$ is with $u'^T u' = v'^T v' = 1$. What are α and β if $u^T v = 0$?

6.5 Basic Theorems

The computational aspects of the algebraic eigenproblem are not dealt with until Chapter 8. The present chapter is all devoted to theory, to the amassing of a wealth of theorems on eigenvalues and eigenvectors.

Theorem 6.4. *If λ is an eigenvalue of matrix A, and x a corresponding eigenvector, then:*

1. *λ^2 is an eigenvalue of A^2, with corresponding eigenvector x.*
2. *$\lambda + \mu$ is an eigenvalue of $A + \mu I$, with corresponding eigenvector x.*
3. *λ^{-1} is an eigenvalue of A^{-1}, with corresponding eigenvector x.*
4. *$\alpha\lambda$ is an eigenvalue of αA, with corresponding eigenvector x.*
5. *λ is also an eigenvalue of $P^{-1}AP$, with corresponding eigenvector $x' = P^{-1}x$.*

Proof. Left as an exercise.

The next theorem extends Theorem 6.4 to include multiplicities.

Theorem 6.5. *If $p_n(\lambda) = \det(A - \lambda I) = 0$ is the characteristic equation of A, then:*

1. $\det(A^T - \lambda I) = p_n(\lambda)$.
2. $\det(A^2 - \lambda^2 I) = p_n(\lambda)p_n(-\lambda)$.
3. $\det(A + \mu I - \lambda I) = p_n(\lambda - \mu)$.
4. $\det(A^{-1} - \lambda I) = \det(-\lambda A^{-1})p_n(\lambda^{-1}), \lambda \neq 0$.
5. $\det(\alpha A - \lambda I) = \det(\alpha I)p_n(\lambda/\alpha)$.
6. $\det(P^{-1}AP - \lambda I) = p_n(\lambda)$.

Proof.

1. $\det(A^T - \lambda I) = \det(A - \lambda I)^T = \det(A - \lambda I)$.
2. $\det(A^2 - \lambda^2 I) = \det(A - \lambda I)\det(A + \lambda I)$.
3. $\det(A + \mu I - \lambda I) = \det(A - (\lambda - \mu))$.

4. $\det(A^{-1} - \lambda I) = \det(-\lambda A^{-1}(A - \lambda^{-1} I))$.

5. $\det(\alpha A - \lambda I) = \det(\alpha I(A - \lambda/\alpha I))$.

6. $\det(P^{-1}AP - \lambda I) = \det(P^{-1}(A - \lambda I)P)$.

End of proof.

The eigenvalues of A and A^T are equal but their eigenvectors are not. However,

Theorem 6.6. *Let $Ax = \lambda x$ and $A^T x' = \lambda' x'$. If $\lambda \neq \lambda'$, then $x^T x' = 0$.*

Proof. Premultiplication of the first equation by x'^T, the second by x^T, and taking their difference yields $(\lambda - \lambda')x^T x' = 0$. By the assumption that $\lambda \neq \lambda'$ it happens that $x^T x' = 0$. End of proof.

Notice that even though matrix A is implicitly assumed to be real, both λ, λ' and x, x' can be complex, and that the statement of the theorem is on $x^T x'$ not $x^H x'$. The word orthogonal is therefore improper here.

A decisively important property of eigenvectors is proved next.

Theorem 6.7. *Eigenvectors corresponding to different eigenvalues are linearly independent.*

Proof. By contradiction. Let $\lambda_1, \lambda_2, \ldots, \lambda_m$ be m distinct eigenvalues with corresponding linearly dependent eigenvectors $x_1, x_2, \ldots, x_m$. Suppose that k is the smallest number of linearly dependent such vectors, and designate them by $x_1, x_2, \ldots, x_k,\ k \leq m$. By our assumption there are k scalars, none of which is zero, such that

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k = o. \tag{6.61}$$

Premultiplying the equation by A and remembering that $Ax_i = \lambda_i x_i$, gives

$$\alpha_1 \lambda_1 x_1 + \alpha_2 \lambda_2 x_2 + \cdots + \alpha_k \lambda_k x_k = o. \tag{6.62}$$

Multiplication of equation (6.61) above by λ_k and its subtraction from eq. (6.62) results in

$$\alpha_1(\lambda_1 - \lambda_k)x_1 + \alpha_2(\lambda_2 - \lambda_k)x_2 + \cdots + \alpha_{k-1}(\lambda_{k-1} - \lambda_k)x_{k-1} = o \tag{6.63}$$

with $\alpha_i(\lambda_i - \lambda_k) \neq 0$. But this implies that there is a smaller set of $k-1$ linearly dependent eigenvectors contrary to our assumption. The eigenvectors $x_1, x_2, \ldots, x_m$ are therefore linearly independent. End of proof.

The multiplicity of eigenvalues is said to be *algebraic*, that of the eigenvectors is said to be *geometric*. Linearly independent eigenvectors corresponding to one eigenvalue of A span an *invariant subspace* of A. The next theorem relates the multiplicity of eigenvalue λ to the largest possible dimension of the corresponding eigenvector subspace.

Theorem 6.8. *Let eigenvalue λ_1 of A have multiplicity m. Then the number k of linearly independent eigenvectors corresponding to λ_1 is at least 1 and at most m; $1 \leq k \leq m$.*

Proof. Let $x_1, x_2, \ldots, x_k$ be all the linearly independent eigenvectors corresponding to λ_1, and write $X = X(n \times k) = [x_1 \ x_2 \ldots x_k]$. Construct $X' = X'(n \times n-k)$ so that $P = P(n \times n) = [X \ X']$ is nonsingular. In partitioned form

$$P^{-1} = \begin{bmatrix} Y(k \times n) \\ Y'(n-k \times n) \end{bmatrix} \tag{6.64}$$

and

$$P^{-1}P = \begin{bmatrix} Y \\ Y' \end{bmatrix} [X \ X'] = \begin{bmatrix} YX & YX' \\ Y'X & Y'X' \end{bmatrix} = \begin{bmatrix} I_k & O \\ O & I_{n-k} \end{bmatrix}. \tag{6.65}$$

Now

$$\begin{aligned} P^{-1}AP = \begin{bmatrix} Y \\ Y' \end{bmatrix} [AX \ AX'] &= \begin{bmatrix} Y \\ Y' \end{bmatrix} [\lambda_1 X \ \ AX'] \\ &= \begin{array}{cc} & \begin{array}{cc} k & n-k \end{array} \\ \begin{array}{c} k \\ n-k \end{array} & \begin{bmatrix} \lambda_1 I_k & YAX' \\ O & Y'AX' \end{bmatrix} \end{array} \end{aligned} \tag{6.66}$$

and

$$\begin{aligned} det(P^{-1}AP - \lambda I) &= det(\lambda_1 I - \lambda I) det(Y'AX' - \lambda I) \\ &= (\lambda_1 - \lambda)^k p_{n-k}(\lambda). \end{aligned} \tag{6.67}$$

On the other hand

$$det(A - \lambda I) = (\lambda_1 - \lambda)^m p_{n-m}(\lambda), \ p_{n-m}(\lambda_1) \neq 0. \tag{6.68}$$

Equating

$$(\lambda_1 - \lambda)^k p_{n-k}(\lambda) = (\lambda_1 - \lambda)^m p_{n-m}(\lambda) \tag{6.69}$$

and taking into consideration the fact that $p_{n-m}(\lambda)$ does not contain the factor $\lambda_1 - \lambda$, but $p_{n-k}(\lambda)$ might, we conclude that $k \leq m$.

At least one eigenvector exists for any λ, and $1 \leq k \leq m$. End of proof.

Theorem 6.9. *Let $A = A(m \times n)$ and $B = B(n \times m)$ be two matrices such that $m \leq n$. Then $AB(m \times m)$ and $BA(n \times n)$ have the same eigenvalues with the same multiplicities, except that the larger matrix BA has in addition $n - m$ zero eigenvalues.*

Proof. Construct the two matrices

$$M = \begin{bmatrix} -\lambda I_m & A \\ -B & I_n \end{bmatrix} \text{ and } M' = \begin{bmatrix} I_m & O \\ B & -\lambda I_n \end{bmatrix} \tag{6.70}$$

so that

$$MM' = \begin{bmatrix} AB - \lambda I_m & -\lambda A \\ O & -\lambda I_n \end{bmatrix} \text{ and } M'M = \begin{bmatrix} -\lambda I_m & A \\ O & BA - \lambda I_n \end{bmatrix}. \tag{6.71}$$

From $\det(MM') = det(M'M)$ it results that

$$(-\lambda)^n \; det(AB - \lambda I_m) = (-\lambda)^m \; det(BA - \lambda I_n) \tag{6.72}$$

or in short $(-\lambda)^n p_m(\lambda) = (-\lambda)^m p_n(\lambda)$. Polynomials $p_n(\lambda)$ and $(-\lambda)^{n-m} p_m(\lambda)$ are the same. All nonzero roots of $p_m(\lambda) = 0$ and $p_n(\lambda) = 0$ are the same and with the same multiplicities, but $p_n(\lambda) = 0$ has extra $n - m$ zero roots. End of proof.

Theorem 6.10. *The geometric multiplicities of the nonzero eigenvalues of AB and BA are equal.*

Proof. Let $x_1, x_2, \ldots, x_k$ be the k linearly independent eigenvectors of AB corresponding to $\lambda \neq 0$. They span a k-dimensional invariant subspace X and for every $x \neq o \in X$, $ABx = \lambda x \neq o$. Hence also $Bx \neq o$. Premultiplication of the homogeneous equation by B yields $BA(Bx) = \lambda(Bx)$, implying that Bx is an eigenvector of BA corresponding to λ. Vectors $Bx_1, Bx_2, \ldots, Bx_k$ are linearly independent since

$$\alpha_1 Bx_1 + \alpha_2 Bx_2 + \cdots + \alpha_k Bx_k = B(\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_k x_k) = Bx \neq o \tag{6.73}$$

and BA has therefore at least k linearly independent eigenvectors $Bx_1, Bx_2, \ldots, Bx_k$. By a symmetric argument for BA and AB we conclude that AB and BA have the same number of linearly independent eigenvectors for any $\lambda \neq 0$. End of proof.

exercises

6.5.1. Show that if

$$A(u+iv) = (\alpha + i\beta)(u+iv),$$

then also

$$A(u-iv) = (\alpha - i\beta)(u-iv)$$

where $i^2 = -1$.

6.5.2. Prove that if A is skew-symmetric, $A = -A^T$, then its spectrum is imaginary, $\lambda(A) = \beta i$, and for every eigenvector x corresponding to a nonzero eigenvalue, $x^T x = 0$. Hint: $x^T A x = 0$.

6.5.3. Show that if A and B are symmetric, then $\lambda(AB - BA)$ is purely imaginary.

6.5.4. Let λ be a distinct eigenvalue of A and x the corresponding eigenvector, $Ax = \lambda x$. Show that if $AB = BA$, and $Bx \neq o$, then $Bx = \lambda' x$ for some $\lambda' \neq 0$.

6.5.5. Specify matrices for which it happens that

$$\lambda_i(\alpha A + \beta B) = \alpha\lambda(A) + \beta\lambda(B)$$

for arbitrary α, β.

6.5.6. Specify matrices A and B for which

$$\lambda(AB) = \lambda(A)\lambda(B).$$

6.5.7. Let $\lambda_1 \neq \lambda_2$ be two eigenvalues of A with corresponding eigenvectors x_1, x_2. Show that if $x = \alpha_1 x_1 + \alpha_2 x_2$, then

$$Ax - \lambda_1 x = \alpha x_2$$

$$Ax - \lambda_2 x = \beta x_1.$$

Write α and β in terms of $\alpha_1, \alpha_2, \lambda_1, \lambda_2$.

6.5.8. Show that if $A^2 = A$, then $\lambda(A) = 0$ or 1.

6.5.9. Show that if $A^2 = -A$, then $\lambda(A) = 0$ or -1.

6.5.10. Show that if $A^2 = I$, then $\lambda(A) = \pm 1$.

6.5.11. What are the eigenvalues of A if $A^2 = 4I,\ A^2 = 4A,\ A^2 = -4A,\ A^2 + A - 2I = O$?

6.5.12. Show that for the nonzero eigenvalues

$$\lambda(X^TAX) = \lambda(AXX^T) = \lambda(XX^TA).$$

6.5.13. Show that if $A^2 = O$, then $\lambda(A) = 0$. Is the converse true? What is the characteristic polynomial of nilpotent $A^2 = O$? Hint: Think about triangular matrices.

6.5.14. Show that if $Q^TQ = I$, then $|\lambda(Q)| = 1$. Hint: If $Qx = \lambda x$, then $x^HQ^T = \overline{\lambda}x^H$.

6.5.15. Write $A = ab^T$ as $A = BC$ with square

$$B = [a\ o\ \ldots\ o] \text{ and } C = \begin{bmatrix} b^T \\ o^T \\ \vdots \\ o^T \end{bmatrix}.$$

Using the fact that the characteristic equation of BC is the same as

$$CB = \begin{bmatrix} b^Ta & o^T \\ o & O \end{bmatrix}$$

show that the characteristic equation of A is

$$\lambda^{n-1}(-\lambda + b^Ta) = 0.$$

What are the eigenvectors of $A = ab^T$? In the same manner show that the characteristic equation of $A = bc^T + de^T$ is

$$\lambda^{n-2}((-\lambda + c^Tb)(-\lambda + e^Td) - (e^Tb)(c^Td)) = 0.$$

6.5.16. Every eigenvector of A is also an eigenvector of A^2. Bring a triangular matrix example to show that the converse need not be true. When is any eigenvector of A^2 an eigenvector of A? Hint: Consider $A \neq I, A^2 = I$.

6.5.17. Matrix A has eigenvalue $\alpha + i\beta$ and corresponding eigenvector $u + iv$. Given u compute α, β and v. Hint: If $Au \neq \alpha u$, then $\beta \neq 0$, and vector v can be eliminated between

$$(A - \alpha I)u = -\beta v \text{ and } (A - \alpha I)v = \beta u.$$

Introduction of vector w such that $w^T u = 0$ is then helpful.

Apply this to

$$A = \begin{bmatrix} 1 & 1 & \\ & 1 & 1 \\ -1 & & 1 \end{bmatrix}, \ \lambda = \frac{3}{2} + \frac{\sqrt{3}}{2}, \ x = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix} + \sqrt{3}i \begin{bmatrix} -1 \\ 1 \\ \end{bmatrix}$$

assuming that only the real part u of x is known.

6.5.18. Let Q be an orthogonal matrix, $Q^T Q = QQ^T = I$. Show that if λ is a complex eigenvalue of $Q, |\lambda| = 1$, with corresponding eigenvector $x = u + iv$, then $u^T v = 0$. Hint: $x^T Q^T = \lambda x^T, x^T x = \lambda^2 x^T x, \lambda^2 \neq 1$.

6.5.19. Matrices A and B are *similar* if $B = P^{-1}AP$ for some invertible P. According to Theorem 6.5 similar matrices have the same characteristic polynomial. Bring a (2×2) upper-triangular matrix example to show that the converse is not true—that matrices having the same characteristic polynomial need not be similar.

6.5.20. Prove that AB and BA are similar if A or B are nonsingular.

6.5.21. Prove that if $AB = BA$, then A and B have at least one common eigenvector.

Hint: Let matrix B have eigenvalue λ with the two linearly independent eigenvectors v_1, v_2 so that $Av_1 = \lambda v_1, Av_2 = \lambda v_2$. From $ABv_1 = BAv_1$ and $ABv_2 = BAv_2$ it follows that $B(Av_1) = \lambda(Av_1)$ and $B(Av_2) = \lambda(Av_2)$, meaning that vectors Av_1 and Av_2 are both in the space spanned by v_1 and v_2. Hence scalars $\alpha_1, \alpha_2, \beta_1, \beta_2$ exist so that

$$\begin{aligned} Av_1 &= \alpha_1 v_1 + \alpha_2 v_2 \\ Av_2 &= \beta_1 v_1 + \beta_2 v_2. \end{aligned}$$

Consequently, for some δ_1, δ_2,

$$A(\delta_1 v_1 + \delta_2 v_2) = (\delta_1 \alpha_1 + \delta_2 \beta_1)v_1 + (\delta_1 \alpha_2 + \delta_2 \beta_2)v_2.$$

You need to show now that scalar μ exists so that

$$\begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} = \mu \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}.$$

6.6 Diagonalization

If matrix $A = A(n \times n)$ has n linearly independent eigenvectors $x_1, x_2, \ldots, x_n$, then they span C^n and any $x \in C^n$ is uniquely written as $x = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n$, so that $Ax = \alpha_1 \lambda_1 x_1 + \alpha_2 \lambda_2 x_2 \cdots + \alpha_n \lambda_n x_n$. Such matrices have special properties.

Definition *Matrix A and matrix $P^{-1}AP$, are similar. If $P = U$, $U^H U = I$, then A and $U^H AU$ are unitarily similar. Matrix A is said to be diagonalizable if a similarity transformation exists that renders $P^{-1}AP$ diagonal.*

Theorem 6.11. *Matrix $A = A(n \times n)$ is diagonalizable if and only if it has n linearly independent eigenvectors.*

Proof. Let A have n linearly independent eigenvectors $x_1, x_2, \ldots, x_n$ and n eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ so that $Ax_i = \lambda_i x_i$ $i = 1, 2, \ldots, n$. With $X = [x_1 x_2 \ldots x_n]$ this is written as $AX = XD$ where D is the diagonal $D_{ii} = \lambda_i$. Because the columns of X are linearly independent X^{-1} exists and $X^{-1}AX = D$.

Conversely if $X^{-1}AX = D$, then $AX = DX$ and the columns of X are the linearly independent eigenvectors corresponding to $\lambda_i = D_{ii}$. End of proof.

An immediate remark we can make about the diagonalization $X^{-1}AX = D$ of A, is that it is not unique since even with distinct eigenvalues the eigenvectors are of arbitrary length.

An important matrix not similar to diagonal is the $n \times n$ Jordan matrix

$$J = \begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & 1 & \\ & & & \lambda & 1 \\ & & & & \lambda \end{bmatrix} \tag{6.74}$$

that has n equal eigenvalues $\lambda_1 = \lambda_2 = \ldots = \lambda_n = \lambda$, but only one single eigenvector $x = e_1$. A matrix that cannot be diagonalized is scornfully named *defective*.

Theorem 6.12. *If matrix A is diagonalized by matrix X, then A^T is diagonalized by matrix $Y = X^{-T}, YX^T = XY^T = I$.*

Proof. If $A = XDX^{-1}$ then $A^T = X^{-T}DX^T$, and $Y = X^{-T}$. End of proof

exercises

6.6.1. Let diagonalizable A have real eigenvalues. Show that the matrix can be written as $A = HS$ with symmetric H and symmetric and positive definite S. Hint: Start with $A = XDX^{-1}$ and recall that symmetric matrix XX^T is positive definite if X is nonsingular.

6.6.2. Prove that if for unitary U both U^HAU and U^HBU are diagonal, then $AB = BA$.

6.6.3. Do n linearly independent eigenvectors and their corresponding eigenvalues uniquely fix $A = A(n \times n)$?

6.6.4. Prove that $A = A(n \times n)$ is similar to A^T, $A^T = P^{-1}AP$.

6.6.5. Show that if $X^{-1}AX$ and $X^{-1}BX$ are both diagonal, then $AB = BA$. Is the converse true?

6.7 Elementary similarity transformations

We do not expect to be able to reduce any square matrix to triangular form by an ending sequence of similarity transformations alone, for this would imply having all the eigenvalues in a finite number of steps, but we should be able to reduce the matrix by such transformations to forms more convenient for the computation of the eigenvalues or the writing of the characteristic equation.

Similarity transformation $B^{-1}AB$, we recall, leaves the characteristic equation of A invariant. Any nonsingular matrix B can be expressed as a product of elementary matrices that we know have simple inverses. We recall from Chapter 2 that the inverse of an elementary operation is an elementary operation, and that premultiplication by an elementary matrix operates on the rows of the matrix, while postmultiplication affects the columns.

Elementary matrices of three types build up matrix B in $B^{-1}AB$, in performing the elementary operations of:

1. permutation of two rows (columns)

$$P = \begin{bmatrix} & 1 & \\ 1 & & \\ & & 1 \end{bmatrix}, \quad P^{-1} = P. \tag{6.75}$$

2. multiplication of a row (column) by a nonzero scalar α

$$E = \begin{bmatrix} 1 & & \\ & \alpha & \\ & & 1 \end{bmatrix}, \; E^{-1} = \begin{bmatrix} 1 & & \\ & \alpha^{-1} & \\ & & 1 \end{bmatrix}. \tag{6.76}$$

3. addition to one row (column) another row (column) times a scalar

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

$$E = \begin{bmatrix} 1 & & \alpha \\ & 1 & \\ & & 1 \end{bmatrix}, \; E^{-1} = \begin{bmatrix} 1 & & -\alpha \\ & 1 & \\ & & 1 \end{bmatrix}. \tag{6.77}$$

The permutation similarity transformation $P^{-1}AP = PAP$ means the interchange of rows k and l of A followed by the interchange of columns k and l of PA. Diagonal entries remain in this row and column permutation on the diagonal.

In the next section we shall need sequences of similarity permutations as described below

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} \lambda_1 & 0 & & & 1 \\ & \lambda_2 & 1 & & \\ & & \lambda_3 & 1 & \\ & & & \lambda_4 & 0 \\ & & & & \lambda_5 \end{bmatrix} \end{array} \rightarrow \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 5 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 5 & 4 \end{array} \\ \begin{bmatrix} \lambda_1 & & & 1 & \\ & \lambda_2 & 1 & & \\ & & \lambda_3 & & 1 \\ & & & \lambda_5 & \\ & & & & \lambda_4 \end{bmatrix} \end{array} \rightarrow$$

$$\begin{array}{c} \\ 1 \\ 2 \\ 5 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 5 & 3 & 4 \end{array} \\ \begin{bmatrix} \lambda_1 & & 1 & & \\ & \lambda_2 & & 1 & \\ & & \lambda_5 & & \\ & & & \lambda_3 & 1 \\ & & & & \lambda_4 \end{bmatrix} \end{array} \rightarrow \begin{array}{c} \\ 1 \\ 5 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 5 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} \lambda_1 & 1 & & & \\ & \lambda_5 & 0 & & \\ & & \lambda_2 & 1 & \\ & & & \lambda_3 & 1 \\ & & & & \lambda_4 \end{bmatrix} \end{array} \tag{6.78}$$

the purpose of which is to bring all the off-diagonal 1's onto the first super-diagonal. It is achieved by performing the row and column permutation $(1,2,3,4,5) \rightarrow (1,5,2,3,4)$ in the sequence $(1,2,3,4,5) \rightarrow (1,2,3,5,4) \rightarrow (1,2,5,3,4) \rightarrow (1,5,2,3,4)$.

Another useful similarity permutation is

$$\begin{array}{cc} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ 0 & A_{22} & A_{23} \\ 1 & A_{32} & A_{33} \end{bmatrix} \end{array} \rightarrow \begin{array}{cc} & \begin{array}{ccc} 1 & 3 & 2 \end{array} \\ \begin{array}{c} 1 \\ 3 \\ 2 \end{array} & \begin{bmatrix} A_{11} & A_{13} & A_{12} \\ 1 & A_{33} & A_{32} \\ 0 & A_{23} & A_{22} \end{bmatrix} \end{array} \tag{6.79}$$

the purpose of which is to have the first column of A start with a nonzero off-diagonal.

Elementary similarity transformation number 2 multiplies row k by $\alpha \neq 0$ and column k by α^{-1}. It leaves the diagonal unchanged, but off-diagonal entries can be modified by it. For instance:

$$\begin{bmatrix} (\alpha\beta\gamma)^{-1} & & & \\ & (\beta\gamma)^{-1} & & \\ & & \gamma^{-1} & \\ & & & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 & \alpha & & \\ & \lambda_2 & \beta & \\ & & \lambda_3 & \gamma \\ & & & \lambda_4 \end{bmatrix} \begin{bmatrix} \alpha\beta\gamma & & & \\ & \beta\gamma & & \\ & & \gamma & \\ & & & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_1 & 1 & & \\ & \lambda_2 & 1 & \\ & & \lambda_3 & 1 \\ & & & \lambda_4 \end{bmatrix}. \tag{6.80}$$

If it happens that some super-diagonal entries are zero, say $\beta = 0$, then we set $\beta = 1$ in the elementary matrix and end up with a zero on this diagonal.

The third elementary similarity transformation that combines rows and columns is of great use in inserting zeroes into $E^{-1}AE$. Schematically, the third similarity transformation is described as

$$\begin{array}{cc} \begin{array}{ccccc} & \_ & & \alpha & \end{array} & \\ \begin{bmatrix} & \times & & \times & \\ \times & \times & \times & \times & \times \\ & \times & & \times & \\ \times & \times & \times & \times & \times \\ & \times & & \times & \end{bmatrix} & \begin{array}{c} \\ \leftarrow \alpha \\ + \\ \\ \\ \end{array} \end{array} \quad \text{or} \quad \begin{array}{cc} \begin{array}{ccccc} & \alpha & & \_ & \end{array} & \\ \begin{bmatrix} & \times & & \times & \\ \times & \times & \times & \times & \times \\ & \times & & \times & \\ \times & \times & \times & \times & \times \\ & \times & & \times & \end{bmatrix} & \begin{array}{c} \\ \\ + \\ \leftarrow \alpha \\ \\ \end{array} \end{array}. \tag{6.81}$$

That is, if row k times α is added to row l, then column l times $-\alpha$ is added to column k; and if row l times α is added to row k, then column k times $-\alpha$ is added to column l.

exercises

6.7.1. Find α so that

$$\begin{bmatrix} 1 & \\ \alpha & 1 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & \\ -\alpha & 1 \end{bmatrix}$$

is upper-triangular.

6.8 Hessenberg and companion matrices

We readily see now how a unique sequence of elementary similarity transformations that uses pivot $p_1 \neq 0$ accomplishes

$$\begin{bmatrix} \times & \times & \times & \times \\ p_1 & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \end{bmatrix} \rightarrow \begin{bmatrix} \times & \times & \times & \times \\ p_1 & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \end{bmatrix}. \tag{6.82}$$

If p_1 is originally zero, then a preliminary elementary similarity permutation is carried out to replace it by another, nonzero, entry from the first column, unless all entries in the column below the diagonal are zero. Doing the same to all columns from the first to the $(n-2)$th reduces the matrix to

$$H = \begin{bmatrix} \times & \times & \times & \times \\ p_1 & \times & \times & \times \\ & p_2 & \times & \times \\ & & p_3 & \times \end{bmatrix} \tag{6.83}$$

which is now in *Hessenberg* form.

In case of an unavoidable zero pivot the matrix reduces to

$$H = \begin{bmatrix} H_{11} & H_{12} \\ & H_{22} \end{bmatrix} \tag{6.84}$$

where H_{11} and H_{22} are Hessenberg submatrices, and $\det(H) = det(H_{11}) det(H_{22})$, effectively decoupling the eigenvalue computations. Assuming that $p_i \neq 0$, and using more elementary similarity transformations of the second kind further reduces the Hessenberg matrix to

$$H = \begin{bmatrix} \times & \times & \times & \times \\ -1 & \times & \times & \times \\ & -1 & \times & \times \\ & & -1 & \times \end{bmatrix}. \tag{6.85}$$

To write the characteristic equation of the Hessenberg matrix we interchange rows so as to have

$$H - \lambda I \rightarrow \begin{bmatrix} -1 & H_{22}-\lambda & H_{23} & H_{24} \\ & -1 & H_{33}-\lambda & H_{34} \\ & & -1 & H_{44}-\lambda \\ p_1 & H_{12} & H_{13} & H_{14} \end{bmatrix}, \ p_1 = H_{11} - \lambda \tag{6.86}$$

and bringing the matrix to upper-triangular form by means of elementary row operations discover the recursive formula

$$\begin{aligned} p_1(\lambda) &= H_{11} - \lambda \\ p_2(\lambda) &= H_{12} + (H_{22} - \lambda)p_1(\lambda) \\ p_3(\lambda) &= H_{13} + H_{23}p_1(\lambda) + (H_{33} - \lambda)p_2(\lambda) \\ &\vdots \\ p_n(\lambda) &= H_{1n} + H_{2n}p_1(\lambda) + H_{3n}p_2(\lambda) + \cdots + (H_{nn} - \lambda)p_{n-1}(\lambda) \end{aligned} \tag{6.87}$$

for the characteristic equation of H.

We summarize it all in

Theorem 6.13. *Any square matrix can be reduced to a Hessenberg form in a finite number of elementary similarity transformations.*

If the Hessenberg matrix

$$H = \begin{bmatrix} \times & q_1 & \times & \times \\ -1 & \times & q_2 & \times \\ & -1 & \times & q_3 \\ & & -1 & \times \end{bmatrix} \tag{6.88}$$

is with $q_i \neq 0$, $i = 1, 2, \ldots, n-2$, then the same elimination done to the lower-triangular part of the matrix may be performed in the upper-triangular part of H to similarly transform it into *tridiagonal* form.

Similarity reduction of matrices to Hessenberg or tridiagonal form precedes most realistic eigenvalue computations. We shall return to this subject in chapter 8, where *orthogonal* similarity transformation will be employed to that purpose. Meanwhile we return to more theoretical matters.

The Hessenberg matrix may be further reduced by elementary similarity transformations to a matrix of simple structure and great theoretical importance. Using the -1's in the first subdiagonal of H to eliminate by elementary similarity transformations all other entries in their column, it is reduced to

$$C = \begin{bmatrix} & & & a_0 \\ -1 & & & a_1 \\ & -1 & & a_2 \\ & & -1 & a_3 \end{bmatrix} \tag{6.89}$$

which is the *companion matrix* of A. We shall show that a_0, a_1, a_2, a_4 in the last column are the coefficients of the characteristic equation of C, and hence of any other matrix similar to it. Indeed, by row elementary operations using the -1's as pivots, and with $n-1$ column interchanges we accomplish the transformations

$$\begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{bmatrix} -\lambda & & & a_0 \\ -1 & -\lambda & & a_1 \\ & -1 & -\lambda & a_2 \\ & & -1 & a_3-\lambda \end{bmatrix} \end{array} \rightarrow \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{bmatrix} & & & p_4 \\ -1 & & & p_3 \\ & -1 & & p_2 \\ & & -1 & p_1 \end{bmatrix} \end{array} \rightarrow \begin{array}{c} \begin{matrix} 4 & 1 & 2 & 3 \end{matrix} \\ \begin{bmatrix} p_4 & & & \\ p_3 & -1 & & \\ p_2 & & -1 & \\ p_1 & & & -1 \end{bmatrix} \end{array} \tag{6.90}$$

where

$$\begin{aligned} & p_4 = \lambda^4 - a_3\lambda^3 + a_2\lambda^2 - a_1\lambda + a_0 \ , \ p_3 = -\lambda^3 + a_3\lambda^2 - a_2\lambda + a_1 \\ & p_2 = \lambda^2 - a_3\lambda + a_2, \quad p_1 = -\lambda + a_3 \end{aligned} \tag{6.91}$$

and

$$det(A-\lambda I) = det(C-\lambda I) = p_4 = \lambda^4 - a_3\lambda^3 + a_2\lambda^2 - a_1\lambda + a_0 \tag{6.92}$$

for any matrix A similar to C.

A Hessenberg matrix with zeroes on the first subdiagonal is transformed into the upper-triangular block form

$$C = \begin{bmatrix} C_1 & \times & \times & \times \\ & C_2 & \times & \times \\ & & \ddots & \times \\ & & & C_k \end{bmatrix} \tag{6.93}$$

where $C_i = C_i(m_i \times m_i), m_1 + m_2 + \cdots m_k = n$, are companion submatrices, and

$$det(C-\lambda I) = det(C_1 - \lambda I)det(C_2-\lambda I)\ldots det(C_k - \lambda I) \tag{6.94}$$

giving the characteristic equation in factored form.

A formal proof is thus given to the long awaited

Theorem 6.14. *If $B(\lambda) = A - \lambda I$ is a real $n \times n$ matrix and λ a scalar, then*

$$det(B(\lambda)) = (-\lambda)^n + a_{n-1}(-\lambda)^{n-1} + \cdots + a_0 \tag{6.95}$$

with real $a_{n-1}, a_{n-2}, \ldots, a_0$.

6.9 Schur to Jordan to Frobenius

The most we can do with elementary similarity transformations is get matrix A into a Hessenberg and then companion matrix form. To further reduce the matrix by similarity transformations to triangular form, we need first to compute all eigenvalues of A.

Theorem (Schur) 6.15. *For any square matrix $A = A(n \times n)$ there exists a unitary matrix U so that $U^{-1}AU = U^H AU = T$ is upper triangular with all the eigenvalues of A on its diagonal, appearing in any specified order.*

Proof. Let λ_1 be an eigenvalue of A that we want to appear first on the diagonal of T, and let u_1 be a unit eigenvector corresponding to it. Even if A is real both λ_1 and u_1 may be complex. In any event there are in C^n $n-1$ unit vectors $u_2, u_3, \ldots, u_n$ orthonormal to u_1. Then $U_1 = [u_1 \ u_2 \ldots u_n]$ is unitary so that $U_1^H U_1 = I$ or $U_1^H = U_1^{-1}$ with which

$$AU_1 = [Au_1 \ Au_2 \ \ldots \ Au_n] = [\lambda_1 u_1 \ u_2' \ \ldots \ u_n'] \tag{6.96}$$

and

$$U_1^H AU_1 = \begin{bmatrix} u_1^H \\ u_2^H \\ \\ u_n^H \end{bmatrix} [\lambda_1 u_1 \ u_2' \ \ldots \ u_n'] = \begin{bmatrix} \lambda_1 & a_1^T \\ & \\ o & A_1 \end{bmatrix} \tag{6.97}$$

with the eigenvalues of A_1 being those of A less λ_1. By the same argument the $(n-1)\times(n-1)$ submatrix A_1 can be similarly transformed into

$$U_2'^H A_1 U_2' = \begin{bmatrix} \lambda_2 & a_2^T \\ & \\ o & A_2 \end{bmatrix} \tag{6.98}$$

where λ_2 is an eigenvalue of A_1, and hence also of A, that we want to appear next on the diagonal of T. Now, if $U_2'(n-1\times n-1)$ is unitary, then so is the $n\times n$

$$U_2 = \begin{bmatrix} 1 & o^T \\ \\ o & U_2' \end{bmatrix} \tag{6.99}$$

and

$$U_2^H U_1^H A U_1 U_2 = \begin{bmatrix} \lambda_1 & \times & \times \times \\ & \lambda_2 & \times \times \\ \\ o & o & A_2 \end{bmatrix}. \tag{6.100}$$

Continuing in this manner we construct $n-1$ unitary matrices $U_1, U_2, \ldots, U_{n-1}$ so that

$$U_{n-1}^H \cdots U_2^H U_1^H A U_1 U_2 \cdots U_{n-1} = \begin{bmatrix} \lambda_1 & \times & \times & \times & \times \\ & \lambda_2 & \times & \times & \times \\ & & \ddots & \times & \times \\ & & & \ddots & \times \\ & & & & \lambda_n \end{bmatrix} \tag{6.101}$$

and since the product of unitary matrices is a unitary matrix, the last equation is concisely written as $U^H A U = T$, where U is unitary, $U^H = U^{-1}$, and T is upper-triangular.

Matrices A and T share the same eigenvalues including multiplicities, and hence $\lambda_1, \lambda_2, \ldots, \lambda_n$, that may be made to appear in any specified order, are the eigenvalues of A. End of proof.

Even if matrix A is real, both U and T in Schur's theorem may be complex, but if we relax the upper-triangular restriction on T and allow 2×2 submatrices on its diagonal, then the Schur decomposition acquires a real counterpart.

Theorem 6.16. *If matrix $A = A(n\times n)$ is real, then there exists a real orthogonal matrix Q so that*

$$Q^T A Q = \begin{bmatrix} S_{11} & \times & \times & \times \\ & S_{22} & \times & \times \\ & & \ddots & \times \\ & & & S_{mm} \end{bmatrix} \tag{6.102}$$

where S_{ii} are submatrices of order either 1×1 or 2×2 with complex conjugate eigenvalues, the eigenvalues of $S_{11}, S_{22}, \ldots, S_{mm}$ being exactly the eigenvalues of A.

Proof. It is enough that we prove the theorem for the first step of the Schur decomposition. Suppose that $\lambda_1 = \alpha + i\beta$ with corresponding unit eigenvector $x_1 = u + iv$. Then $\overline{\lambda} = \alpha - i\beta$ and $\overline{x}_1 = u - iv$ are also an eigenvalue and eigenvector of A, and

$$\begin{aligned} Au &= \alpha u - \beta v \\ Av &= \beta u + \alpha v \end{aligned} \qquad \beta \neq 0. \tag{6.103}$$

This implies that if vector x is in the two-dimensional space spanned by u and v, then so is Ax. The last pair of vector equations are concisely written as

$$AV = VM \ , \ M \ = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix} \tag{6.104}$$

where $V = V(n \times 2) = [u, v]$, and where M is verified to have eigenvalues λ_1 and $\overline{\lambda}_1$.

Let $q_1, q_1, \ldots, q_n$ be a set of orthonormal vectors in R^n with q_1 and q_2 being in the subspace spanned by u and v. Then $q_i^T A q_1 = q_i^T A q_2 = 0 \ i = 3, 4, \ldots, n$, and

$$Q_1^T A Q_1 \ = \begin{bmatrix} S_{11} & A_1' \\ & \\ O & A_1 \end{bmatrix} \tag{6.105}$$

where $Q_1 = [q_1 \ q_2 \ldots q_n]$, and where

$$S_{11} \ = \begin{bmatrix} q_1^T A q_1 & q_1^T A q_2 \\ & \\ q_2^T A q_1 & q_2^T A q_2 \end{bmatrix}. \tag{6.106}$$

To show that the eigenvalues of S_{11} are λ_1 and $\overline{\lambda}_1$ we write $Q = Q(n \times 2) = [q_1 \ q_2]$, and have that $V = QS, \ Q = VS^{-1}, \ S = S(2 \times 2)$, since q_1, q_2 and u, v span the same two-dimensional subspace. With this, $AV = VM$ becomes $AQS = QSM$, and $S_{11} = Q^T AQ = SMS^{-1}$, implying that S_{11} and M share the same eigenvalues. End of proof.

The similarity reduction of A to triangular form T requires the complete eigensolution of $Ax = \lambda x$, but once this is done we expect to be able to further simplify T by means of elementary similarity transformations only. The rest of this section is devoted to elementary similarity transformations designed to bring the upper-triangular T as close as possible to diagonal form, culminating in the Jordan form.

Theorem 6.17. *Suppose that in the partitioning*

$$T \ = \begin{bmatrix} T_{11} & \times & \times & \times \\ & T_{22} & \times & \times \\ & & \ddots & \times \\ & & & T_{mm} \end{bmatrix} \tag{6.107}$$

T_{ii} are upper triangular, each with equal diagonal entries λ_i, but such that $\lambda_i \neq \lambda_j$. Then there exists a similarity transformation

$$X^{-1}TX = \begin{bmatrix} T_{11} & & & \\ & T_{22} & & \\ & & \ddots & \\ & & & T_{mm} \end{bmatrix} \tag{6.108}$$

that annuls all off-diagonal blocks without changing the diagonal blocks.

Proof. The transformation is achieved through a sequence of elementary similarity transformations using diagonal pivots. At first we look at the 2×2 block-triangular matrix

$$T = \begin{bmatrix} T_{11} & T_{12} \\ & \\ & T_{22} \end{bmatrix} = \left[\begin{array}{ccc|cc} \lambda_1 & T_{12} & T_{13} & T_{14} & T_{15} \\ & \lambda_1 & T_{23} & T_{24} & T_{25} \\ & & \lambda_1 & T_{34} & T_{35} \\ \hline & & & \lambda_2 & T_{45} \\ & & & & \lambda_2 \end{array}\right] \rightarrow \left[\begin{array}{ccc|cc} \lambda_1 & T_{12} & T_{13} & T'_{14} & T_{15} \\ & \lambda_1 & T_{23} & T'_{24} & T_{25} \\ & & \lambda_1 & T'_{34} & T'_{35} \\ \hline & & & \lambda_2 & T_{45} \\ & & & & \lambda_2 \end{array}\right] \tag{6.109}$$

where the shown above similarity transformation consists of adding α times column 3 to column 4, and $-\alpha$ times row 4 to row 3, and demonstrate that submatrix T_{12} can be annulled by a sequence of such row-wise elimination of entries T_{34}, T_{35}; T_{24}, T_{25}; T_{14}, T_{15} in that order. An elementary similarity transformation that involves rows and columns 3 and 4 does not affect submatrices T_{11} and T_{22}, but $T'_{34} = T_{34} + \alpha(\lambda_1 - \lambda_2)$, and with $\alpha = -T_{34}/(\lambda_1 - \lambda_2), T'_{34} = 0$. Continuing the elimination in the suggested order leaves created zeroes zero, and does not change the triangular submatrices.

The off-diagonal submatrices of T are eliminated in the same order and we are left with a block-diagonal $X^{-1}TX$. End of proof.

We are at the stage now where square matrix A is similarly transformed into a diagonal block form with triangular diagonal submatrices that have the same λ on their own diagonal. Suppose that T is such a typical triangular matrix with a nonzero first super-diagonal. Then elementary similarity transformations exist to the effect that

$$T = \begin{bmatrix} \lambda & \times & \times & \times & \times \\ & \lambda & \times & \times & \times \\ & & \lambda & \times & \times \\ & & & \lambda & \times \\ & & & & \lambda \end{bmatrix} \rightarrow \begin{bmatrix} \lambda & 1 & \times & \times & \times \\ & \lambda & 1 & \times & \times \\ & & \lambda & 1 & \times \\ & & & \lambda & 1 \\ & & & & \lambda \end{bmatrix} \rightarrow \begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \lambda & 1 & \\ & & & \lambda & 1 \\ & & & & \lambda \end{bmatrix} \tag{6.110}$$

in which elimination is done with the 1's on the first super-diagonal, row by row starting with the first row.

Definition. *The $m \times m$ matrix*

$$J(\lambda) = \begin{bmatrix} \lambda & 1 & & \\ & \lambda & 1 & \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix} \tag{6.111}$$

is a simple Jordan submatrix of order m.

Notice that the simple Jordan matrix may be written as $J = \lambda I + N$ where N is nilpotent, $N^{m-1} \neq O, N^m = O$, and is of nullity 1.

A simple Jordan submatrix of order m has one eigenvalue λ of multiplicity m, one eigenvector e_1, and $m-1$ *generalized eigenvectors* $e_2, e_3, \ldots, e_m$ strung together by

$$\begin{aligned} (J - \lambda I)e_1 &= o \\ (J - \lambda I)e_2 &= e_1 \\ &\vdots \\ (J - \lambda I)e_m &= e_{m-1}. \end{aligned} \tag{6.112}$$

Or $Ne_1 = o, N^2 e_2 = o, \ldots, N^m e_m = o$, implying that the nullspace of N is embedded in its range.

The existence of a sole eigenvector for J implies that nullity$(J - \lambda I) = 1$, rank$(J - \lambda I) = m - 1$, and hence the full complement of super-diagonal 1's in J. Conversely, if matrix T in eq.(1.112) is known to have a single eigenvector, then its corresponding Jordan matrix is assuredly simple. Non-simple Jordan matrix

$$J = \begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \lambda & 0 & \\ & & & \lambda & 1 \\ & & & & \lambda \end{bmatrix} \tag{6.113}$$

is of rank$(J - \lambda I) = m - 2$ and hence of nullity$(J - \lambda I) = 2$, implying the existence of two linearly independent eigenvectors, here e_1 and e_4, for the same repeating eigenvalue λ. There are now two chains of generalized eigenvectors: $e_1, Ne_2 = e_1, Ne_3 = e_2$, and

$e_4, Ne_5 = e_4$; five linearly independent eigenvectors in all.The appearance of two zeroes in the super-diagonal of a non-simple J bespeaks the existence of three linearly independent eigenvectors for the same repeating λ, three chains of generalized eigenvectors, and so on.

Jordan's form is the closest matrix T can get by similarity transformations to diagonal form. Every triangular diagonal submatrix with a nonzero first super-diagonal can be reduced by elementary similarity transformations to a simple Jordan submatrix. Moreover,

Theorem (Jordan) 6.18. *Any matrix $A = A(n \times n)$ is similar to*

$$J = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_k \end{bmatrix} \tag{6.114}$$

where $J_i = (m_i \times m_i)$ are simple Jordan submatrices, and $m_1 + m_2 + \cdots m_k = n$.

Proof. We know that a sequence of elementary similarity transformations exists by which any $n \times n$ matrix is carried into the diagonal block form

$$T = \begin{bmatrix} \lambda_1 & \times & \times & & & & & & & \\ & \lambda_1 & \times & & & & & & & \\ & & \lambda_1 & & & & & & & \\ & & & \lambda_2 & \times & & & & & \\ & & & & \lambda_2 & & & & & \\ & & & & & \lambda_3 & \times & \times & \times & \\ & & & & & & \lambda_3 & \times & \times & \\ & & & & & & & \lambda_3 & \times & \\ & & & & & & & & \lambda_3 & \\ & & & & & & & & & \lambda_4 \end{bmatrix} = \begin{bmatrix} T_{11} & & & \\ & T_{22} & & \\ & & T_{33} & \\ & & & T_{44} \end{bmatrix} \tag{6.115}$$

with $\lambda_1 \neq \lambda_2 \neq \lambda_3 \neq \lambda_4$.

If the first super-diagonal of T_{ii} is nonzero, then a finite sequence of elementary similarity transformations, done in partitioned form reduce T_{ii} to a simple Jordan matrix. Zeroes in the first super-diagonal of T_{ii} complicate the elimination and cause the creation of several Jordan simple submatrices with the same diagonal λ in place of T_{ii}. A constructive proof to this fact is done by induction on the order m of $T = T_{ii}$. A 2×2 upper-triangular matrix with equal diagonal entries can certainly be brought by one elementary similarity transformation to a simple Jordan form of order 2, and if the 2×2 matrix is diagonal, then it consists of two 1×1 Jordan blocks.

Let $T = T(m \times m)$ be upper-triangular with diagonal entries all equal λ, as in eq. (6.110) and suppose that an elementary similarity transformation exists that transforms the leading $m-1 \times m-1$ submatrix of T to Jordan blocks. We shall show then that T itself can be reduced by elementary similarity transformations to a *direct sum* of Jordan blocks with equal diagonal entries.

For clarity's sake we refer in the proof to a particular matrix, but it should be obvious that the argument is general. The similarity transformation that reduces the leading $m-1 \times m-1$ matrix to Jordan blocks leaves a nonzero last column with entries $\alpha_1, \alpha_2, \alpha_3, \alpha_4$, as in the left-hand matrix below

$$\begin{bmatrix} \lambda & 1 & & & \alpha_1 \\ & \lambda & 0 & & \alpha_2 \\ & & \lambda & 1 & \alpha_3 \\ & & & \lambda & \alpha_4 \\ & & & & \lambda \end{bmatrix} \rightarrow \begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 0 & & 1 \\ & & \lambda & 1 & \\ & & & \lambda & 1 \\ & & & & \lambda \end{bmatrix}. \tag{6.116}$$

Entries α_1 and α_3 that have a 1 in their row are eliminated routinely, and if $\alpha_2 = 0$ we are done since then the matrix is in the desired form. If $\alpha_4 = 0$, then α_2 can be brought to the first super-diagonal by the sequence of similarity permutations described in the previous section, and then made 1. Hence the assumption that $\alpha_2 = \alpha_4 = 1$ as in the right-hand matrix above.

One of these 1's can be made to disappear in a finite number of elementary similarity transformations. For a detailed observation of how this is accomplished look at the larger 9×9 submatrices of eq. (6.117).

$$\begin{array}{cc} & \begin{array}{ccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{array} & \begin{bmatrix} \lambda & 1 & & & & & & & \\ & \lambda & 1 & & & 1 & & & \\ & & \lambda & 1 & & & \cdot & & \\ & & & \lambda & 1 & & & \cdot & \\ & & & & \lambda & 0 & & & 1\cdot \\ & & & \cdot & & \lambda & 1 & & \\ & & & & \cdot & & \lambda & 1 & \\ & & & & & & & \lambda & 1\cdot \\ & & & & & & & & \lambda \end{bmatrix} \end{array} \quad \begin{array}{cc} & \begin{array}{ccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{array} & \begin{bmatrix} \lambda & 1 & & & & & \cdot & & \\ & \lambda & 1 & & & & & \cdot & \\ & & \lambda & 0 & & & & & 1\cdot \\ & & & \lambda & 1 & & & & \\ 1 & & & & \lambda & 1 & & & \\ & \cdot & & & & \lambda & 1 & & \\ & & \cdot & & & & \lambda & 1 & \\ & & & & & & & \lambda & 1\cdot \\ & & & & & & & & \lambda \end{bmatrix} \end{array} . \tag{6.117}$$

Look first at the right-hand matrix. If the 1 in row 8 is used to to eliminate by an elementary similarity transformation the 1 above it in row 5, then a new 1 appears at entry (4,8) of the matrix. This new 1 is ousted (it leaves behind a dot) by the super-diagonal 1 below it, but still a new 1 appears at entry (3,7). Repeated, such eliminations push the 1 in a diagonal path across the matrix until it gets stuck in column 6—the column of the super-diagonal zero. Our efforts are yet in vain. If the 1 in row 5 is used to eliminate the 1 below it in row 8, then a new 1 springs up at entry (7,5). Elimination of this new 1 with the super-diagonal 1 above it pushes it up diagonally across the matrix until it is annihilated at row 6, the row just below the row of the super-diagonal zero. We have now succeeded in getting rid of this 1. On the other hand, for the matrix to the right, the upper elimination path is successful since the pushed 1 reaches row 1,where it is zapped, before falling into the column of the zero super-diagonal. The lower elimination path for this matrix does not fare that well. The chased lower 1 comes to the end of its journey in column 1 before having the chance to enter row 4 where it could have been annihilated. In general, if the zero in the super-diagonal happens to be in row k, then the upper elimination path is successful if $n > 2k$, while the lower path is successful if $n \leq 2k+2$. In any event, at least one elimination path always ends in an actual annihilation, and we are essentially done. Only a single 1 remains in the last column and it can be brought upon the super-diagonal, if it is not already on it, by row and column permutations.

In case of several Jordan blocks in one T there are more than two 1's in the last columns and there are several elimination paths to perform.

We have shown that if $T(m-1 \times m-1)$ can be reduced by a finite number of elementary similarity transformations to a Jordan blocks form, then $T(m \times m)$ can also be reduced to this form. Starting from $m = 2$ we have then the result for any m. End of proof.

As an exercise the reader can work out the details of elementary similarity transformations

$$
\begin{bmatrix}
\lambda & 1 & & & & & & & & \\
 & \lambda & 0 & & & & & & & 1 \\
 & & \lambda & 1 & & & & & & \\
 & & & \lambda & 1 & & & & & \\
 & & & & \lambda & 0 & & & & 1 \\
 & & & & & \lambda & 1 & & & \\
 & & & & & & \lambda & 0 & & 1 \\
 & & & & & & & \lambda & 1 & \\
 & & & & & & & & \lambda & 1 \\
 & & & & & & & & & \lambda
\end{bmatrix}
\rightarrow
\begin{bmatrix}
\lambda & 1 & & & & & & & & \\
 & \lambda & 0 & & & & & & & \\
 & & \lambda & 1 & & & & & & \\
 & & & \lambda & 1 & & & & & \\
 & & & & \lambda & 0 & & & & 1 \\
 & & & & & \lambda & 1 & & & \\
 & & & & & & \lambda & 0 & & \\
 & & & & & & & \lambda & 1 & \\
 & & & & & & & & \lambda & 0 \\
 & & & & & & & & & \lambda
\end{bmatrix}
$$

$$
\rightarrow
\begin{bmatrix}
\lambda & 1 & & & & & & & & \\
 & \lambda & & & & & & & & \\
 & & \lambda & 1 & & & & & & \\
 & & & \lambda & 1 & & & & & \\
 & & & & \lambda & 1 & & & & \\
 & & & & & \lambda & & & & \\
 & & & & & & \lambda & 1 & & \\
 & & & & & & & \lambda & & \\
 & & & & & & & & \lambda & 1 \\
 & & & & & & & & & \lambda
\end{bmatrix} . \tag{6.118}
$$

The existence proof given for the Jordan form is constructive and in integer arithmetic the matrix can be set up unambiguously. In floating-point computations the construction is numerically problematic and the Jordan form has not found many practical computational applications. It is nevertheless of considerable theoretical interest, at least in achieving the goal of ultimate systematic reduction of A by means of similarity transformations.

Say matrix A has but one eigenvalue λ, and a corresponding Jordan form as in eq.(6.113). Then nonsingular matrix $X = [x_1 \ x_2 \ x_3 \ x_4 \ x_5]$ exists so that $X^{-1}AX = J$, or $AX = XJ$, where

$$
\begin{aligned}
&(A - \lambda I)x_1 = o, \quad (A - \lambda I)x_2 = x_1, \quad (A - \lambda I)x_3 = x_2 \\
&(A - \lambda I)x_4 = o, \quad (A - \lambda I)x_5 = x_4
\end{aligned} \tag{6.119}
$$

or

$$
\begin{aligned}
&(A - \lambda I)x_1 = o, \quad (A - \lambda I)^2 x_2 = o, \quad (A - \lambda I)^3 x_3 = o \\
&(A - \lambda I)x_4 = o, \quad (A - \lambda I)^2 x_5 = o
\end{aligned} \tag{6.120}
$$

and if A and λ are real, then so are x_1, x_2, x_3, x_4, x_5.

Conversely, computation of the generalized eigenvectors x_1, x_2, x_3, x_4, x_5 furnishes nonsingular matrix X that puts $X^{-1}AX$ into Jordan form.

If A is real but λ complex with complex generalized eigenvectors, then $\overline{\lambda}$ is also an eigenvalue of A with the same algebraic and geometrical multiplicities as λ, and hence with corresponding conjugate complex generalized eigenvectors.

Instead of thinking about Jordan's theorem concretely in terms of elementary operations we may think about it abstractly in terms of vector spaces. Because of the block nature of the theorem we need limit ourselves to just one of the triangular matrices T_{ii} of theorem 6.17. Moreover, because $X^{-1}(\lambda I+N)X = \lambda I + X^{-1}NX$ we may further restrict discussion to nilpotent matrix N only, or equivalently to Jordan matrix J of eq.(6.111) with $\lambda = 0$. First we prove

Lemma 6.19. *If A is nilpotent of index m, then nullity(A^{k+1}) >nullity(A^k) for positive integer $k, m > k > 0$; and $-$nullity(A^{k+1}) $+ 2$nullity(A^k)$-$nullity(A^{k-1}) ≥ 0 for $k > 0$.*

Proof. All eigenvalues of A are zero and hence by Schur's theorem an orthogonal matrix Q exists so that $Q^TAQ = N$ is strictly upper-triangular. Since $Q^TA^kQ = N^k$, and since nullity (Q^TA^kQ) =nullity(A^k), we may substitute N for A. Look first at the case $k = 1$. If E is an elementary operations matrix, then nullity(N^2) =nullity($(EN)N$). To be explicit consider the specific $N(5 \times 5)$ of nullity 2, and assume that row 2 of EN is annulled by the operation so that $(EN)N$ is of the form

$$\overset{EN}{\begin{bmatrix} & \alpha_1 & \times & \times & \times \\ & & & & \\ & & & \alpha_3 & \times \\ & & & & \alpha_4 \\ & & & & \end{bmatrix}} \overset{N}{\begin{bmatrix} & \alpha_1 & \times & \times & \times \\ & & \alpha_2 & \times & \times \\ & & & \alpha_3 & \times \\ & & & & \alpha_4 \\ & & & & \end{bmatrix}} = \overset{EN^2}{\begin{bmatrix} & & \alpha_1\alpha_2 & \times & \times \\ & & & & \\ & & & & \alpha_3\alpha_4 \\ & & & & \\ & & & & \end{bmatrix}}. \tag{6.121}$$

Rows 2 and 5 of EN^2 are zero because rows 2 and 5 of EN are zero, but in addition, row 4 of EN^2 is also zero and the nullity of EN^2 is greater than that of EN. In the event that $\alpha_4 = 0$, row 2 of EN does not vanish if nullity(EN) $= 2$, yet the last three rows of EN^2 are now zero by the fact that $\alpha_3\alpha_4 = 0$. The essence of the proof for $k = 1$ is, then, showing that if corner entry $(EN)_{45} = \alpha_4 = 0$, then also corner entry $(EN^2)_{35} = \alpha_3\alpha_4 = 0$. This is always the case whatever k is. Say $N = N(6 \times 6)$ with $N_{56} = \alpha_5$. Then $N^2_{46} = \alpha_4\alpha_5, N^3_{36} = \alpha_3\alpha_4\alpha_5$, and if $N^2_{46} = 0$, then also $N^3_{36} = 0$. Consequently nullity(N^3) >nullity(N^2). Proof of the second part, which is a special case of the Frobenius rank inequality of Theorem 5.25, is left as an exercise. End of proof.

In the process of setting up matrix X in $X^{-1}AX$ that similarly transforms nilpotent matrix A into the Jordan nilpotent matrix $N = J(0)$ the need arises to solve chains of linear equations such as the typical $Ax_1 = o, Ax_2 = x_1$. Suppose that nullity$(A) = 1$, and nullity$(A^2) = 2$.This means a one dimensional nullspace for A, so that $Ax = o$ is inclusively solved by $x = \alpha_1 x_1$ for any α_1 and $x_1 \neq o$. Premultiplication by A turns the second equation into $A^2x_2 = Ax_1 = o$. Since nullity$(A^2) = 2$, $A^2x = o$ is inclusively solved by $x = \beta_1 x_1 + \beta_2 x_2$, in which β_1 and β_2 are arbitrary and x_1 and x_2 are linearly independent. Now, since $Ax_1 = o$, $Ax = \beta_2 Ax_2$, and since Ax is in the nullspace of A, $A(Ax) = o$, it must so be that $\beta_2 Ax_2 = \alpha_1 x_1$.

Theorem 6.20 *Nilpotent matrix A of nullity k and index m, $A^{m-1} \neq O$, $A^m = O$, is similar to a block diagonal matrix N of k nullity-one Jordan nilpotent submatrices, with the largest diagonal block being of dimension m, and with the dimensions of the other blocks being uniquely determined by the nullities of $A^2, A^3, \ldots, A^{m-1}$; the number of $j \times j$ blocks in the nilpotent Jordan matrix N being equal to* $-\text{nullity}(N^{j+1}) + 2\text{nullity}(N^j) - \text{nullity}(N^{j-1})$ $j = 1, 2, \ldots, m$.

Proof. Let N be a block Jordan nilpotent matrix. If nullity$(N) = k$, then k rows of N are zero and it is composed of k blocks. Raising N to some power amounts to raising each block to that power. If $N_j = N_j(j \times j)$ is one such block, then $N_j^{j-1} \neq O, N_j^j = O$, nullity$(N_j^k) = j$ if $k \geq j$, and the dimension of the largest block in N is m. Also, nullity$(N_j^k)-$nullity$(N_j^{k-1}) = 1$ if $k \leq j$. It results that the number of blocks in N larger than $j \times j$ is equal to nullity$(N^{j+1})-$nullity(N^j). The number of blocks larger than $j-1 \times j-1$ is then equal to nullity$(N^j)-$nullity(N^{j-1}), and the difference is the number of $j \times j$ blocks in N. Doing the same to $A = A(n \times n)$ we determine all the block sizes, and they add up to n.

We shall look at three typical examples that will expose the generality of the contention.

Say matrix $A = A(4 \times 4)$ is such that nullity $(A) = 1$, nullity $(A^2) = 2$, nullity $(A^3) =$

3, $A^4 = O$. The only 4×4 nilpotent Jordan matrix $N = J(0)$ that has these nullities is

$$N = \begin{bmatrix} & 1 & & \\ & & 1 & \\ & & & 1 \\ & & & \end{bmatrix}. \tag{6.122}$$

We show that a nonsingular X exists so that $X^{-1}AX = N$. Indeed, if $X = [x_1\ x_2\ x_3\ x_4]$, then $AX = XN$ is written columnwise as

$$[Ax_1\ Ax_2\ Ax_3\ Ax_4] = [o\ x_1\ x_2\ x_3] \tag{6.123}$$

and $Ax_1 = o,\ Ax_2 = x_1,\ Ax_3 = x_2,\ Ax_4 = x_3$, or $Ax_1 = o,\ A^2x_2 = o,\ A^3x_3 = o,\ A^4x_4 = o$. Because nullity $(A) = 1,\ Ax = o$ possesses a nontrivial solution $x = x_1$. Because nullity $(A^2) = 2,\ A^2x = o$ has two linearly independent solutions of which at least one, call it x_2, is linearly independent of x_1. Because nullity $(A^3) = 3,\ A^3x = o$ has three linearly independent solutions of which at least one, call it x_3, is linearly independent of x_1 and x_2. Because $A^4 = O,\ A^4x = o$ has four linearly independent solutions of which at least one, call it x_4, is linearly independent of x_1, x_2, x_3. Hence $X = [x_1\ x_2\ x_3\ x_4]$ is invertible and $X^{-1}AX = N$.

Say matrix $A = A(5 \times 5)$ is such that nullity $(A) = 2$, nullity $(A^2) = 4,\ A^3 = O$. The only 5×5 compound Jordan nilpotent matrix N that has these nullities is the two-block

$$N = \begin{bmatrix} & 1 & & & \\ & & 1 & & \\ & & & 0 & \\ & & & & 1 \\ & & & & \end{bmatrix}. \tag{6.124}$$

for which $AX = XN$ is

$$[Ax_1\ Ax_2\ Ax_3\ Ax_4\ Ax_5] = [o\ x_1\ x_2\ o\ x_4] \tag{6.125}$$

so that $Ax_1 = 0,\ Ax_2 = x_1,\ Ax_3 = x_2,\ Ax_4 = 0,\ Ax_5 = x_4$, or $Ax_1 = o,\ A^2x_2 = o,\ A^3x_3 = o,\ Ax_4 = o,\ A^2x_5 = o$. Because nullity $(A) = 2,\ Ax = o$ has two linearly independent solutions, $x = x_1$ and $x = x_4$. Because nullity $(A^2) = 4,\ A^2x = o$ has four linearly independent solutions of which at least two, call them x_2, x_5, are linearly independent of x_1 and x_4. Because $A^3 = O,\ A^3x = o$ has four linearly independent solutions of which at least one, call it x_3, is linearly independent of x_1, x_2, x_4, x_5. Hence $X = [x_1\ x_2\ x_3\ x_4\ x_5]$ is invertible and $X^{-1}AX = N$.

Say matrix $A = A(8 \times 8)$ is such that nullity $(A) = 4$, nullity $(A^2) = 7,\ A^3 = O$. Apart from block ordering, the only 8×8 compound nilpotent Jordan matrix that has these nullities is the four-block

$$N = \begin{bmatrix} & 1 & & & & & & \\ & & 1 & & & & & \\ & & & 0 & & & & \\ & & & & 1 & & & \\ & & & & & 0 & & \\ & & & & & & 1 & \\ & & & & & & & 0 \\ & & & & & & & \end{bmatrix} \tag{6.126}$$

for which $AX = XN$ gives rise to $Ax_1 = o,\ A^2x_2 = o,\ A^3x_3 = o,\ Ax_4 = o, A^2x_5 = o,\ Ax_6 = o,\ A^2x_7 = o,\ Ax_8 = o$. Because nullity $(A) = 4,\ Ax = o$ has four linearly independent solutions $x_1,\ x_4,\ x_6,\ x_8$. Because nullity $(A^2) = 7,\ Ax = o$ has seven linearly independent solutions of which at least three, x_2, x_5, x_7, are linearly independent of x_1, x_4, x_6, x_8. Because $A^3 = O,\ A^3x = o$ has eight linearly independent solutions of which at least one, call it x_3, is linearly independent of the other seven x's. Hence $X = [x_1\ x_2\ x_3\ x_4\ x_5\ x_6\ x_7\ x_8]$ is invertible and $X^{-1}AX = N$.

One readily verifies that the blocks as given in the theorem correctly add up in size to just fit in the matrix. End of proof.

We shall now use the Jordan form to prove the remarkable

Theorem (Frobenius) 6.21. *Every complex (real) square matrix is a product of two complex (real) symmetric matrices of which at least one is nonsingular.*

Proof. It is accomplished with the aid of the symmetric permutation submatrix

$$P = \begin{bmatrix} & & & 1 \\ & & 1 & \\ & 1 & & \\ 1 & & & \end{bmatrix},\ P^{-1} = P \tag{6.127}$$

that has the decisive property of turning J into S,

$$PJ = S = \begin{bmatrix} & & & \lambda \\ & & \lambda & 1 \\ & \lambda & 1 & \\ \lambda & 1 & & \end{bmatrix} \tag{6.128}$$

which is symmetric. What is done to the simple Jordan submatrix J by one submatrix P is done to the complete J in block form, and we shall write it as $PJ = S$ and $J = PS$.

To prove the complex case we write $A = XJX^{-1} = XPSX^{-1} = (XPX^T)(X^{-T}SX^{-1})$, and see that A is the product of the two symmetric matrices XPX^T and $X^{-T}SX^{-1}$, of which the first is nonsingular.

The proof to the real case hinges on showing that if A is real, then XPX^T is real; the reality of $X^{-T}SX^{-1}$ follows then from $X^{-T}SX^{-1} = (XPX^T)^{-1}A$. When A is real, whenever complex $J(m \times m)$ appears in the Jordan form, $\overline{J}(m \times m)$ is also there. Since we are dealing with blocks in a partitioned form we permit ourselves to restrict the rest of the argument to the Jordan form and block permutation matrix

$$J \ = \begin{bmatrix} J_1 & & \\ & \overline{J}_1 & \\ & & J_3 \end{bmatrix}, \ P = \begin{bmatrix} P_1 & & \\ & P_1 & \\ & & P_3 \end{bmatrix} \tag{6.129}$$

in which J_3 is real. Accordingly, matrix X in $X^{-1}AX = J$ is partitioned as

$$X \ = [X_1 \ \overline{X}_1 \ X_3] \ = [X_1 \ O \ O] \ + \ [O \ \overline{X}_1 \ O] \ + \ [O \ O \ X_3] \tag{6.130}$$

where X_3 is real, and

$$XPX^T = X_1P_1X_1^T + \overline{X}_1P_1\overline{X}_1^T + X_3P_3X_3^T. \tag{6.131}$$

With $X_1 = R + iR'$, where R and R' are real, XPX^T becomes

$$XPX^T = 2(RP_1R^T - R'P_1R'^T) + X_3P_3X_3^T \tag{6.132}$$

proving that XPX^T is real and symmetric. It is also nonsingular. End of proof.

exercises

6.9.1. Show that the the Jordan form of

$$A = \begin{bmatrix} \lambda & & 1 & \\ & \lambda & & 1 \\ & & \lambda & \\ & & & \lambda \end{bmatrix} \text{ is } J = \begin{bmatrix} \lambda & 1 & & \\ & \lambda & & \\ & & \lambda & 1 \\ & & & \lambda \end{bmatrix}.$$

6.9.2. If matrix A is of the form $A = \lambda I + N$, then $X^{-1}AX = \lambda I + X^{-1}NX$, and we may consider N only. Perform the elementary similarity transformations needed to bring

$$A = \begin{bmatrix} 0 & 1 & & & & & \\ & 0 & 1 & & & & \\ & & 0 & 0 & & & 1 \\ & & & 0 & 0 & & 1 \\ & & & & 0 & 1 & \\ & & & & & 0 & 1 \\ & & & & & & 0 \end{bmatrix} \text{ and } B = \begin{bmatrix} 0 & 1 & & & & & \\ & 0 & 0 & & & & 1 \\ & & 0 & 1 & & & \\ & & & 0 & 0 & & 1 \\ & & & & 0 & 1 & \\ & & & & & 0 & 1 \\ & & & & & & 0 \end{bmatrix}$$

into Jordan form. Discuss the various possibilities.

6.9.3. What is the Jordan form of matrix $A = A(16 \times 16)$ if nullity$(A) = 6$, nullity$(A^2) = 11$, nullity$(A^3) = 15, A^4 = O$.

6.9.4. What is the Jordan form of J^2, if

$$J = \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & 0 & 1 & \\ & & & 0 & 1 \\ & & & & 0 \end{bmatrix}?$$

6.9.5. Write all (generalized) eigenvectors of

$$A = \begin{bmatrix} 6 & 10 & -3 \\ -3 & -5 & 2 \\ -1 & -2 & 2 \end{bmatrix}$$

knowing that $\lambda(A) = 1$.

6.9.6. Show that the eigenvalues of

$$A = \begin{bmatrix} 6 & -2 & -2 & \\ 2 & 2 & -4 & 2 \\ & & 6 & -2 \\ & & 2 & 2 \end{bmatrix}$$

are all 4 with the two (linearly independent) eigenvectors $x_1 = [1\ 0\ 1\ 1]^T$ and $x_2 = [1\ 1\ 0\ 0]^T$. Find generalized eigenvectors x_1' and x_2' such that $Ax_1' = 4x_1' + x_1$ and $Ax_2' = 4x_2' + x_2$ and form matrix $X = [x_1\ x_1'\ x_2\ x_2']$. Show that $J = X^{-1}AX$ is the Jordan form of A.

6.9.7. Instead of the companion matrix of section 6.8 it is occasionally more convenient to use the transpose

$$C = \begin{bmatrix} & -1 & \\ & & -1 \\ \alpha_0 & \alpha_1 & \alpha_2 \end{bmatrix}.$$

Verify that the characteristic equation of C is $-\lambda^3 + \alpha_2\lambda^2 - \alpha_1\lambda + \alpha_0 = 0$. Show that to eigenvalue λ of C corresponds one eigenvector $x = [1\ -\lambda\ \lambda^2]^T$, and that if λ repeats then $Cx = \lambda x$ and $Cx' = \lambda x' + x$, where $x' = [0\ -1\ 2\lambda]^T$ and x are linearly independent. Say the eigenvalues of C are $\lambda_1, \lambda_1, \lambda_2$ with corresponding generalized eigenvectors x_1, x_1', x_2 and write $X = [x_1\ x_1'\ x_2]$. Show that x_1, x_1', x_2 are linearly independent and prove that $J = X^{-1}CX$ is the Jordan form of C.

Use the above argument to similarly transform

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

to Jordan form.

6.9.8. Find all X such that $JX = XJ$,

$$J = \begin{bmatrix} \lambda & 1 & \\ & \lambda & 1 \\ & & \lambda \end{bmatrix}.$$

Prove that if $A = S^{-1}JS$, then also $A = T^{-1}JT$ if $T = XS$ where X is nonsingular and such that $JX = XJ$.

6.9.9. What is wrong with the following? Instead of A we write B,

$$A = \begin{bmatrix} 0 & 1 & & \\ & 0 & 0 & 1 \\ & & 0 & 1 \\ & & & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 & & \\ & 0 & \epsilon & 1 \\ & & 0 & 1 \\ & & & 0 \end{bmatrix},$$

where ϵ is minute. We use ϵ to annul the 1 in its row then set $\epsilon = 0$.

Show that the eigenvectors of matrix

$$A = \begin{bmatrix} 1 & 1 & \\ & 1+\epsilon & 1 \\ & & 1+2\epsilon \end{bmatrix}$$

corresponding to eigenvalues $\lambda_1 = 1, \lambda_2 = 1+\epsilon, \lambda_3 = 1+2\epsilon$, are $x_1 = [1\ 0\ 0]^T, x_2 = [1\ \epsilon\ 0]^T, x_3 = [1\ 2\epsilon\ 2\epsilon^2]$.

6.9.10. Show that if

$$J = \begin{bmatrix} \lambda & 1 & & \\ & \lambda & 1 & \\ & & \lambda & 1 \\ & & & \lambda \end{bmatrix} \text{ then } J^4 = \begin{bmatrix} \lambda^4 & 4\lambda^3 & 6\lambda^2 & 4\lambda \\ & \lambda^4 & 4\lambda^3 & 6\lambda^2 \\ & & \lambda^4 & 4\lambda^3 \\ & & & \lambda^4 \end{bmatrix}$$

and $J^m \rightarrow O$ as $m \rightarrow \infty$ if and only if $|\lambda| < 1$. Show that $A^m \rightarrow O$ as $m \rightarrow \infty$ if and only if $|\lambda(A)| < 1$.

6.9.11. Prove that if zero is the only eigenvalue of $A = A(n \times n)$, then $A^m = O$ for some $m \leq n$.

6.9.12. Show that

$$\begin{bmatrix} \lambda & 1 & \\ & \lambda & 1 \\ & & \lambda \end{bmatrix} X = X \begin{bmatrix} \mu & 1 & \\ & \mu & 1 \\ & & \mu \end{bmatrix}$$

has a nontrivial solution if and only if $\lambda \neq \mu$. Prove that $AX = XB$ has a nontrivial solution if and only if A and B have no common eigenvalues. Hint: Write $A = S^{-1}JS$ and $B = T^{-1}J'T$ for the Jordan matrices J and J'.

6.9.13. Solve

$$\begin{bmatrix} \lambda & 1 & \\ & \lambda & 1 \\ & & \lambda \end{bmatrix} X - X \begin{bmatrix} \mu & 1 & \\ & \mu & 1 \\ & & \mu \end{bmatrix} = C.$$

Prove that $AX - XB = C$ has a unique solution for any C if and only if A and B have no common eigenvalues.

6.9.14. Let $A = I + CC'$, with $C = C(n \times r)$ and $C' = C'(r \times n)$ both of rank r. Show that if A is nonsingular, then

$$A^{-1} = (I + CC')^{-1} = I + \alpha_1(CC') + \alpha_2(CC')^2 + \ldots + \alpha_r(CC')^r.$$

6.10 Hermitian (symmetric) matrices

Complex matrix $A = R + iS$ is Hermitian if its real part is symmetric, $R^T = R$, and its imaginary part is skew-symmetric, $S^T = -S$, $A = A^H$. The algebraic structure of the Hermitian (symmetric) eigenproblem has greater completeness and more certainty to it than the unsymmetric eigenproblem. Real symmetric matrices are also most prevalent in discrete mathematical physics. Nature is very often symmetric.

Theorem 6.22. *All eigenvalues of a Hermitian matrix are real.*

Proof. If $Ax = \lambda x$, then $x^H Ax = \lambda x^H x$, and since by Theorem 6.2 both $x^H Ax$ and $x^H x$ are real so is λ. End of proof.

Corollary 6.23. *All eigenvalues and eigenvectors of a real symmetric matrix A are real.*

Proof. The eigenvalues are real because A is Hermitian. The eigenvectors x are real because both A and λ in $(A - \lambda I)x = o$ are real. End of proof.

Theorem 6.24 *The eigenvectors of a Hermitian matrix corresponding to different eigenvalues are orthogonal.*

Proof. Let $Ax_1 = \lambda_1 x_1$ and $Ax_2 = \lambda_2 x_2$ be with $\lambda_1 \neq \lambda_2$. Premultiplying the first equation by x_2^H produces $x_2^H Ax_1 = \lambda_1 x_2^H x_1$. Since $A = A^H$, λ is real, $(x_2^H Ax_1)^H = \lambda_1 (x_2^H x_1)^H$, and $x_1^H Ax_2 = \lambda_1 x_1^H x_2$. With $Ax_2 = \lambda_2 x_2$ this becomes $\lambda_2 x_1^H x_2 = \lambda_1 x_1^H x_2$, and since $\lambda_1 \neq \lambda_2$ $x_1^H x_2 = 0$. End of proof.

Theorem (spectral) 6.25. *If A is Hermitian (symmetric), then there exists a unitary (orthogonal) matrix U so that $U^H AU = D$, where D is diagonal with the real eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of A on its diagonal.*

Proof. By Schur's theorem there is a unitary matrix U so that $U^H AU = T$ is upper-triangular. Since A is Hermitian, $A^H = A$, $T^H = (U^H AU)^H = U^H AU = T$, and T is diagonal, $T = D$. Then $AU = UD$, the columns of U consisting of the n eigenvectors of A, and $D_{ii} = \lambda_i$. End of proof.

Corollary 6.26. *If λ is an eigenvalue of symmetric matrix A of multiplicity k, then there are k linearly independent eigenvectors corresponding to it.*

Proof. Let the diagonalization of A be $Q^TAQ = D$ with $D_{ii} = \lambda \;\; i = 1, 2, \ldots, k$ and $D_{ii} \neq \lambda \;\; i = k+1, \ldots, n$. Then $Q^T(A - \lambda I)Q = D'$, $D'_{ii} = 0 \;\; i = 1, 2, \ldots, k$, and $D'_{ii} \neq 0 \;\; i > k$. The rank of D' is $n - k$, and because Q is nonsingular this is also the rank of $A - \lambda I$. The nullity of $A - \lambda I$ is k. End of proof.

Symmetric matrices have a complete set of orthogonal eigenvectors whether or not their eigenvalues repeat. Let $u_1, u_2, \ldots, u_n$ be the n orthonormal eigenvectors of the symmetric $A = A(n \times n)$. Then

$$I = u_1u_1^T + u_2u_2^T + \cdots + u_nu_n^T, \;\; A = \lambda_1u_1u_1^T + \lambda_2u_2u_2^T + \cdots + \lambda_nu_nu_n^T \tag{6.133}$$

and

$$A^{-1} = \lambda_1^{-1}u_1u_1^T + \lambda_2^{-1}u_2u_2^T + \cdots + \lambda_n^{-1}u_nu_n^T. \tag{6.134}$$

Corollary 6.27. *Symmetric matrix $A = A(n \times n)$ is uniquely determined by its n eigenvalues and n orthonormal eigenvectors $u_1, u_2, \ldots, u_n$.*

Proof. In case the eigenvalues are distinct the eigenvectors are unique (up to sense) and, unambiguously, $A = \lambda_1u_1u_1^T + \lambda_2u_2u_2^T + \ldots + \lambda_nu_nu_n^T$. Say $\lambda_1 = \lambda_2 \neq \lambda_3$ so that

$$A = \lambda_1(u_1u_1^T + u_2u_2^T) + \lambda_3u_3u_3^T = \lambda_1UU^T + \lambda_3u_3u_3^T \tag{6.135}$$

with $U = [u_1 \;\, u_2]$. According to Corollary 6.26 any nonzero vector confined to the space spanned by u_1, u_2 is an eigenvector of A for eigenvalue $\lambda_1 = \lambda_2$. Let u'_1, u'_2 be an orthonormal pair in the plane of u_1, u_2, and write

$$B = \lambda_1(u'_1{u'_1}^T + u'_2{u'_2}^T) + \lambda_3u_3u_3^T = \lambda_1U'{U'}^T + \lambda_3u_3u_3^T \tag{6.136}$$

for $U' = [u'_1 \;\, u'_2]$. Matrices U and U' are related by $U' = UQ$, where Q is orthogonal, and hence

$$B = \lambda_1UQQ^TU^T + \lambda_3u_3u_3^T = \lambda_1UU^T + \lambda_3u_3u_3^T = A. \tag{6.137}$$

End of proof.

Example. Consider the symmetric eigenproblem

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \lambda \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \tag{6.138}$$

transformed by elementary row operations into

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ & & & \\ & & & \\ & & & \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \lambda \begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ -1 & & 1 & \\ -1 & & & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}. \tag{6.139}$$

The first equation of the homogeneous system is

$$(1-\lambda)x_1 + x_2 + x_3 + x_4 = 0 \tag{6.140}$$

while the last three are

$$\lambda(x_4 - x_1) = 0,\ \lambda(x_3 - x_1) = 0,\ \lambda(x_2 - x_1) = 0. \tag{6.141}$$

We verify that $\lambda = 0$ is an eigenvalue with eigenvector x that has its components related by $x_1 + x_2 + x_3 + x_4 = 0$, so that

$$x = x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}. \tag{6.142}$$

Eigenvalue $\lambda = 0$, we conclude, is of multiplicity three. Assuming next that $\lambda \neq 0$, we have from the last three homogeneous equations that $x_1 = x_2 = x_3 = x_4$, and from the first that $(4-\lambda)x_1 = 0$. To have a nonzero eigenvector, x_1 must be nonzero, and $\lambda = 4$ is an eigenvalue with the corresponding eigenvector $x = [1\ 1\ 1\ 1]^T$.

Let the eigenvalues be listed as $\lambda_1 = 0, \lambda_2 = 0, \lambda_3 = 0, \lambda_4 = 4$. To set up an orthogonal eigenvector system we choose the first eigenvector to be $x_1 = [1\ 0\ 0\ -1]^T$, and the second $x_2 = \alpha x_1 + [0\ 1\ 0\ -1]^T$. The condition $x_2^T x_1 = 0$ determines that $\alpha = -1/2$, and $x_2 = [-1\ 2\ 0\ -1]^T$. The third eigenvector corresponding to $\lambda = 0$ is written as $x_3 = \alpha x_1 + \beta x_2 + [0\ 0\ 1\ -1]^T$, and the conditions $x_1^T x_3 = x_2^T x_3 = 0$ determine that $\alpha = -1/2,\ \beta = -1/6$, and $x_3 = [-1\ -1\ 3\ -1]^T$. Now

$$x_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix},\ x_2 = \begin{bmatrix} -1 \\ 2 \\ 0 \\ -1 \end{bmatrix},\ x_3 = \begin{bmatrix} -1 \\ -1 \\ 3 \\ -1 \end{bmatrix}, \qquad x_4 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \tag{6.143}$$

with x_1, x_2, x_3 being the eigenvectors of $\lambda = 0$, and x_4 the eigenvector of $\lambda = 4$, and they constitutes now an orthogonal set that spans R^4.

Symmetric matrices have a complete set of orthogonal eigenvectors whether or not their eigenvalues repeat, but they are not the only matrices that possess this property. Our next task is to characterize the class of complex matrices that are unitarily similar to a diagonal matrix.

Definition. *Complex matrix A is normal if and only if $AA^H = A^H A$.*

Hermitian, skew-Hermitian, unitary, and diagonal are normal matrices.

Lemma 6.28. *An upper-triangular normal matrix is diagonal.*

Proof. Let N be the upper triangular normal matrix

$$N = \begin{bmatrix} N_{11} & N_{12} & N_{13} & N_{14} \\ & N_{22} & N_{23} & N_{24} \\ & & N_{33} & N_{34} \\ & & & N_{44} \end{bmatrix}, \ N^H = \begin{bmatrix} \overline{N}_{11} & & & \\ \overline{N}_{12} & \overline{N}_{22} & & \\ \overline{N}_{13} & \overline{N}_{23} & \overline{N}_{33} & \\ \overline{N}_{14} & \overline{N}_{24} & \overline{N}_{34} & \overline{N}_{44} \end{bmatrix}. \tag{6.144}$$

Then

$$(NN^H)_{11} = N_{11}\overline{N}_{11} + N_{12}\overline{N}_{12} + \cdots + N_{1n}\overline{N}_{1n} = |N_{11}|^2 + |N_{12}|^2 + \cdots + |N_{12}|^2 \tag{6.145}$$

while

$$(N^H N)_{11} = |N_{11}|^2. \tag{6.146}$$

The condition $N^H N = NN^H$ dictates that $N_{12} = N_{13} = \cdots N_{1n} = 0$. Next

$$(NN^H)_{22} = |N_{22}|^2 + |N_{33}|^2 + \cdots + |N_{24}|^2 \tag{6.147}$$

and

$$(N^H N)_{22} = |N_{22}|^2 \tag{6.148}$$

so that $N_{23} = N_{24} = \cdots N_{24} = 0$. Proceeding in this manner, we conclude that $N_{ij} = 0,\ i \neq j$. End of proof.

Lemma 6.29. *A matrix that is unitarily similar to a normal matrix is normal.*

Proof. If N is normal, U unitary, and $N' = U^H N U$, then ${N'}^H = U^H N^H U$ and

$$N'{N'}^H = U^H N U U^H N^H U = U^H N N^H U = U^H N^H N U = (U^H N^H U)(U^H N U) = {N'}^H N'. \tag{6.149}$$

End of proof.

Theorem (Toeplitz) 6.30. *Complex matrix A is unitarily similar to a diagonal matrix if and only if A is normal.*

Proof. If $U^H A U = D$, then $A = U D U^H$, $A^H = U D^H U^H$,

$$AA^H = UDU^H U D^H U^H = U D D^H U^H = U D^H D U^H = (U D^H U^H)(U D U^H) = A^H A \tag{6.150}$$

and A is normal.

Conversely, let A be normal. By Schur's theorem there exists a unitary U so that $U^H A U = T$ is upper triangular. By Lemma 6.29 T is normal, and by Lemma 6.28 it is diagonal. End of proof.

We have characterized now all matrices that can be diagonalized by a unitary similarity transformation, but one question still lingers in our mind. Are there real unsymmetric matrices with a full complement of n orthogonal eigenvectors and n real eigenvalues? The answer is no.

First we notice that a square real matrix may be normal without being symmetric or skew-symmetric. For instance

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} + \alpha \begin{bmatrix} & 1 & -1 \\ -1 & & 1 \\ 1 & -1 & \end{bmatrix}, \; A^T A = A A^T = 3 \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} + \alpha^2 \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} \tag{6.151}$$

is such a matrix for arbitrary α. Real unsymmetric matrices exist that have n orthogonal eigenvectors but their eigenvalues are complex. Indeed, let $U + iV$ be unitary so that $(U + iV)D(U^T - iV^T) = A$, where D is a real diagonal matrix and A a real square matrix. Equating real parts we find that $A = UDU^T + VDV^T$ which can happen only if A is symmetric.

Only real symmetric matrices have n real orthogonal eigenvectors.

exercises

6.10.1. If symmetric A and B have the same eigenvalues does this mean that $A = B$?

6.10.2. Prove the eigenvalues of $A = A^T$ are all equal to λ if and only if $A = \lambda I$.

6.10.3. Show that if $A = A^T$ and $A^2 = A$, then $\text{rank}(A) = \text{trace}(A) = A_{11} + A_{22} + \ldots + A_{nn}$.

6.10.4. Let $A = A^T(m \times m)$ have eigenvalues $\alpha_1, \alpha_2, \ldots, \alpha_m$ and corresponding orthonormal eigenvectors $u_1, u_2, \ldots, u_m$. Let $B = B^T(n \times n)$ have eigenvalues $\beta_1, \beta_2, \ldots, \beta_n$ and corresponding orthonormal eigenvectors $v_1, v_2, \ldots, v_n$. Show that the eigenvalues of

$$C = \begin{bmatrix} A & \gamma u_1 v_1^T \\ \gamma v_1 u_1^T & B \end{bmatrix}$$

are $\alpha_2, \ldots, \alpha_m$, $\beta_2, \ldots, \beta_n, \gamma_1, \gamma_2$, where γ_1 and γ_2 are the eigenvalues of

$$T = \begin{bmatrix} \alpha_1 & \gamma \\ \gamma & \beta_1 \end{bmatrix}.$$

Hint: Look for an eigenvector of the form $x = [\delta_1 u_1^T \ \delta_2 v_1^T]^T$.

6.10.5. Let matrix $A + iB$ be Hermitian so that $A = A^T \neq O, B = -B^T \neq O$. What are the conditions on α and β so that $C = (\alpha + i\beta)(A + iB)$ is Hermitian?

6.10.6. Show that the inverse of a Hermitian matrix is Hermitian.

6.10.7. Is circulant matrix

$$A = \begin{bmatrix} \alpha_0 & \alpha_1 & \alpha_2 \\ \alpha_2 & \alpha_0 & \alpha_1 \\ \alpha_1 & \alpha_2 & \alpha_0 \end{bmatrix}$$

normal?

6.10.8. Prove that if A, B, and AB are normal, then so is BA. Also that if A is normal, then so are A^2 and A^{-1}.

6.10.9. Show that if A is normal, then so is A^m, for any integer m, positive or negative.

6.10.10. Show that $A = I + \alpha Q$ is normal if Q is orthogonal.

6.10.11. Show that the sum of two normal matrices need not be normal, nor the product.

6.10.12. Show that if A and B are normal and $AB = O$, then also $BA = O$.

6.10.13. Show that if normal A is with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ and corresponding orthonormatl eigenvectors $x_1, x_2, \ldots, x_n$, then $A = \lambda_1 x_1 x_1^H + \lambda_2 x_2 x_2^H + \cdots + \lambda_n x_n x_n^H$ and $A^H = \overline{\lambda}_1 x_1 x_1^H + \overline{\lambda}_2 x_2 x_2^H + \cdots + \overline{\lambda}_n x_n x_n^H$.

6.10.14. Show that if A is normal and $Ax = \lambda x$, then $A^H x = \overline{\lambda} x$.

6.10.15. Prove that A is normal if and only if

$$trace\ (A^H A) = \sum_{i=1}^{n} |\lambda_i|^2.$$

Otherwise

$$trace\ (A^H A) \geq \sum_{i=1}^{n} |\lambda_i|^2.$$

6.10.16. Show that if A and B are normal and $AB = O$, then also $BA = O$.

6.10.17. Let A and B in $C = A + iB$ be both real. Show that the eigenvalues of

$$C' = \begin{bmatrix} A & -B \\ B & A \end{bmatrix}$$

are those of C and $\overline{C}$.

6.10.18. A unitary matrix is normal and is hence unitarily similar to a diagonal matrix. Show that real orthogonal matrix A can be transformed by real orthogonal Q into

$$Q^T A Q = \begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_n \end{bmatrix}$$

where

$$A_i = \begin{bmatrix} c & s \\ -s & c \end{bmatrix}\ ,\ c^2 + s^2 = 1$$

unless it is 1 or -1. Show that if the eigenvalues of an orthogonal matrix are real, then it is necessarily symmetric.

6.11 Positive definite matrices

Matrices that have their origin in physical problems where conservation of energy holds are positive (semi) definite, and hence their special place in linear algebra. In this section we consider some of the most important theorems on the positive (semi) definite eigenproblem.

Theorem 6.31. *Matrix $A = A^T$ is positive (semi) definite if and only if all its eigenvalues are (non-negative) positive.*

Proof. Since $A = A(n \times n)$ is symmetric it possesses a complete orthogonal set of eigenvectors $x_1, x_2, \ldots, x_2$, and any $x \in R^n$ may be expanded as $x = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n$. With this

$$x^T A x = \lambda_1 \alpha_1^2 + \lambda_2 \alpha_2^2 + \cdots + \lambda_n \alpha_n^2 \tag{6.152}$$

and $x^T A x > 0$ for any $x \neq o$ if and only if $\lambda_i > 0$. If some eigenvalues are zero, then $x^T A x \geq 0$ even with $x \neq o$. End of proof.

Theorem 6.32. *To every positive semidefinite and symmetric $A = A(n \times n)$ there corresponds a unique positive semidefinite and symmetric matrix $B = A^{\frac{1}{2}}$, the positive square root of A, such that $A = BB$.*

Proof. If such B exists, then it must be of the form

$$B = \mu_1 u_1 u_1^T + \mu_2 u_2 u_2^T + \cdots + \mu_n u_n u_n^T \tag{6.153}$$

where $\mu_i \geq 0$ and where $u_1, u_2, \ldots, u_n$ are orthonormal. Then

$$B^2 = \mu_1^2 u_1 u_1^T + \mu_2^2 u_2 u_2^T + \cdots + \mu_n^2 u_n u_n^T = A \tag{6.154}$$

implying that $\mu_1^2, \ldots, \mu_n^2$ are the eigenvalues of A and $u_1, u_2, \ldots, u_n$ are the eigenvectors of A. According to Corollary 6.27 there is no other B. End of proof.

Theorem (polar decomposition) 6.33. *Matrix $A = A(m \times n)$ with linearly independent columns admits the unique factorization $A = QS$, in which $Q = Q(m \times n)$ has orthonormal columns, and where $S = S(n \times n)$ is symmetric and positive definite.*

Proof. The factors are $S = (A^T A)^{1/2}$ and $Q = AS^{-1}$, and the factorization is unique by the uniqueness of S. End of proof.

Every matrix of rank r can be written (Corollary 2.36) as the sum of r rank one matrices. The spectral decomposition theorem gives such a sum for a symmetric A as $A = \lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T + \cdots + \lambda_n u_n u_n^T$, where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the (real) eigenvalues of A and where $x_1, x_2, \ldots, x_n$ are the corresponding orthonormal eigenvectors. If nonsymetric A has distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ with corresponding eigenvectors $u_1, u_2, \ldots, u_n$, then A^T has the same eigenvalues with corresponding eigenvectors $v_1, v_2, \ldots, v_n$ such that ${v_i}^T u_j = 0$ if $i \neq j$ and $v_i^T u_i = 1$. Then $A = \lambda_1 u_1 v_1^T + \lambda_2 u_2 v_2^T + \cdots + \lambda_n u_n v_n^T$. The next theorem describes a similar *singular value decomposition* for rectangular matrices.

Theorem (singular value decomposition) 6.34. *Let matrix* $A = A(m \times n)$ *be of rank* $r \leq \min(m, n)$. *Then* A *may be decomposed as*

$$A = \sigma_1 v_1 u_1^T + \sigma_2 v_2 u_2^T + \cdots + \sigma_r v_r u_r^T \tag{6.155}$$

where $\sigma_i > 0$, *where* $u_1, u_2, \ldots, u_r$ *are orthonormal in* R^n, *and where* $v_1, v_2, \ldots, v_r$ *are orthonormal in* R^m.

Proof. Matrices $A^T A(n \times n)$ and $AA^T(m \times m)$ are both positive semidefinite. Matrix $A^T A$ has r nonzero eigenvalues. Let the eigenvalues of $A^T A$ be $0 < \sigma_1^2 \leq \sigma_2^2 \leq \cdots \leq \sigma_r^2$, $\sigma_i^2 = 0 \ i = r+1, \ldots, n$, with corresponding orthonormal eigenvectors $u_1, u_2, \ldots, u_r, \ldots u_n$. According to Theorem 6.9, the eigenvalues of AA^T are also $0 < \sigma_1^2 \leq \sigma_2^2 \leq \cdots \leq \sigma_r^2$, and $\sigma_i^2 = 0 \ i = r+1, \ldots, m$. Denote the corresponding orthonormal eigenvectors by $v_1, v_2, \ldots, v_r \ldots v_m$.

Premultiplying $A^T A u_i = \sigma_i^2 u_i$ by A we resolve that $AA^T(Au_i) = \sigma_i^2(Au_i)$. Since $\sigma_i^2 > 0$ for $i \leq r$, $Au_i \neq o$, $i \leq r$, are the orthogonal eigenvectors of AA^T corresponding to σ_i^2, $u_i^T A^T A u_j = 0 \ i \neq j$, $u_i^T A^T A u_i = \sigma_i^2$. Hence $Au_i = \sigma_i v_i \ i \leq r$. Also $Au_i = o$ if $i > r$.

Postmultiplying $Au_i = \sigma_i v_i$ by $u_i^T \ i = 1, 2, \ldots, n$ and adding we obtain

$$A(u_1 u_1^T + u_2 u_2^T + \cdots + u_n u_n^T) = \sigma_1 v_1 u_1^T + \sigma_2 v_2 u_2^T + \cdots + \sigma_n v_n u_n^T \tag{6.156}$$

and since $u_1u_1^T + u_2u_2^T + \cdots + u_nu_n^T = I$, and $\sigma_i = 0 \; i > r$, the equation in the theorem is established. End of proof.

Differently put, Theorem 6.34 states that $A = A(m \times n)$ may be written as $A = VDU$, for $V = V(m \times r)$ with orthonormal columns, for a positive definite diagonal $D = D(r \times r)$, and for $U = U(r \times n)$ with orthonormal rows.

The singular value decomposition of A reminds us of the full rank factorization of A, and in fact the one can be deduced from the other. According to Corollary 2.37 matrix $A = A(m \times n)$ of rank r can be factored as $A = BC$ with $B = B(m \times r)$ and $C = C(r \times n)$, of full column rank and full row rank, respectively. Then according to Theorem 6.33, $B = Q_1S_1, \; C = S_2Q_2$, and $A = Q_1S_1S_2Q_2$. Matrix S_1S_2 is nonsingular and admits the factorization $S_1S_2 = QS$. Matrix $S = S(r \times r)$ is symmetric positive definite and is diagonalized as $S = X^TDX, \; X^TX = XX^T = I$, so that finally $A = Q_1QX^TDXQ_2 = VDU$.

exercises

6.11.1. Matrix $A = A(n \times n)$ is of rank r. Can it have less than $n - r$ zero eigenvalues? More? Consider

$$A = \begin{bmatrix} 0 & 1 & & \\ & 0 & 1 & \\ & & 0 & 1 \\ & & & 0 \end{bmatrix}.$$

What if A is symmetric?

6.11.2. Prove that if $A = A^T$ is positive semidefinite and $x^TAx = 0$,then $Ax = o$.

6.11.3. Let $A = P + S$ be with a positive definite and symmetric P, and a skew-symmetric S, $S = -S^T$. Show that $\lambda(A) = \alpha + i\beta$ with $\alpha > 0$.

6.11.4. Let u_1, u_2, u_3, and v_1, v_2, v_3 be two orthonormal vector systems in R^3. Show that

$$Q = \delta_1u_1v_1^T + \delta_2u_2v_2^T + \delta_3u_3v_3^T$$

is orthogonal provided that $\delta_i^2 = 1$. Is this expansion unique? What happens when Q is symmetric?

6.11.5. Prove that if $A = A^T$ and $B = B^T$, and at least one of them is also positive definite, then the eigenvalues of AB are real. Moreover, if both matrices are positive definite then $\lambda(AB) > 0$. Hint: Consider $Ax = \lambda B^{-1}x$.

6.11.6. Prove that if A and B are positive semidefinite, then the eigenvalues of AB are real and nonnegative, and that consequently $\epsilon I + AB$ is nonsingular for any $\epsilon > 0$.

6.11.7. Prove that if positive definite and symmetric A and B are such that $AB = BA$, then BA is also positive definite and symmetric.

6.11.8. Show that if matrix $A = A^T$ is positive definite, then all coefficients of its characteristic equation are nonzero and alternate in sign. Proof of the converse is more difficult.

6.11.9. Show that if $A = A^T$ is positive definite, then

$$\det(A) = \lambda_1 \lambda_2 \cdots \lambda_n \leq A_{11} A_{22} \cdots A_{nn}$$

with equality holding for diagonal A only.

6.11.10. Let matrices A and B be symmetric and positive definite. What is the condition on α and β so that $\alpha A + \beta B$ is positive definite.

6.11.11. Bring an upper-triangular matrix example to show that a nonsymmetric matrix with positive eigenvalues need not be positive definite.

6.11.12. Prove that if A is positive semidefinite and B is normal then AB is normal if and only if $AB = BA$.

6.11.13. Let complex $A = B + iC$ be with real A and B. show that matrix

$$R = \begin{bmatrix} B & -C \\ C & B \end{bmatrix}$$

is such that

1. If A is normal, then so is R.

2. If A is Hermitian, then R is symmetric.

3. If A is positive definite, then so is R.

4. If A is unitary, then R is orthogonal.

6.11.14. For

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

compute the eigenvalues and eigenvectors of A^TA and AA^T, and write its singular value decomposition.

6.11.15. Prove that nonsingular $A = A(n \times n)$ can be written (uniquely?) as $A = Q_1DQ_2$ where Q_1 and Q_2 are orthogonal and D diagonal.

6.11.16. Show that if A and B are symmetric, then orthogonal Q exists such that Q^TAQ and Q^TBQ are both diagonal if and only if $AB = BA$.

6.11.17. Show that if $A = A^T$ and $B = B^T$, then $\lambda(AB)$ is real provided that at least one of the matrices is positive semidefinite.

6.11.18. Let u_1, u_2 be orthonormal in R^n. Prove that $P = u_1u_1^T + u_2u_2^T$ is of rank 2 and that $P = P^T, P^2 = P$.

6.11.19. Let $P = P^2, P^n = P, n \geq 2$, be an idempotent (projection) matrix. Show that the Jordan form of P is diagonal. Show further that if $\text{rank}(P) = r$, then $P = XDX^{-1}$, where diagonal D is such that $D_{ii} = 1$ if $i \leq r$, and $D_{ii} = 0$ if $i > r$.

6.11.20. Show that if P is an orthogonal projection matrix, $P = P^2, P = P^T$, then $\alpha^2P + \beta^2(I - P)$ is positive definite provided α and β are nonzero.

6.11.21. Prove that projection matrix P is normal if and only if $P = P^T$.

6.12 Congruency-Sylvester's law of inertia

Transformation of square matrix A into square matrix $B = P^TAP$ with nonsingular P is common, particularly with symmetric matrices, and it leaves an interesting invariant. Obviously symmetry, rank, nullity and positive definiteness are preserved by the transformation, but considerably more interesting is the fact that if A is symmetric, then A and B have the same number of positive, negative, and zero eigenvalues for any P. In the language of mechanics, matrices A and B have the same *inertia.*

We remain formal.

Definition. *Square matrix B is congruent to square matrix A if $B = P^TAP$ and P is nonsingular.*

Theorem 6.35. *Congruency of matrices is reflexive, symmetric, and transitive.*

Proof. Matrix A is self-congruent since $A = I^TAI$; congruency is reflexive. It is also symmetric since if $B = P^TAP$, then also $A = P^{-T}BP^{-1}$. To show that if A and B are congruent, and if B and C are congruent, then A and C are also congruent we write $A = P^TBP$, $B = Q^TCQ$, and have by substitution that $A = (QP)^TC(QP)$. Congruency is transitive. End of proof.

If congruency of symmetric matrices has invariants, then they are best seen on the diagonal form of P^TAP. The diagonal form itself is, of course, not unique; different P matrices produce different diagonal matrices. We know that if A is symmetric then there exists an orthogonal Q so that $Q^TAQ = D$ is diagonal with $D_{ii} = \lambda_i$, the eigenvalues of A. But diagonalization of a symmetric matrix by a congruent transformation is also possible with symmetric elementary transformations.

Theorem 6.36. *Every symmetric matrix of rank r is congruent to*

$$D = \begin{bmatrix} d_1 & & & & \\ & d_2 & & & \\ & & \ddots & & \\ & & & d_r & \\ & & & & 0 \end{bmatrix} \tag{6.157}$$

where $d_i \neq 0$ if $i \leq r$.

Proof. We shall give a constructive proof to this theorem. Assume that $A \neq O$. If $A_{11} \neq 0$, then it is used as first pivot $d_1 = A_{11}$ in the symmetric elimination

$$E_1AE_1^T = \begin{bmatrix} d_1 & o^T \\ & \\ o & A_1 \end{bmatrix} \tag{6.158}$$

but if $A_{11} = 0$, then it is replaced through symmetric elementary operations by a nonzero pivot. First the diagonal is searched to see if $A_{ii} \neq 0$ can be found on it. Any nonzero diagonal A_{ii} may be symmetrically interchanged with A_{11} by the interchange of rows 1 and i followed by the interchange of columns 1 and i. If, however, $A_{ii} = 0$ for all $i = 1, 2, \ldots, n$, then the whole matrix is searched for a nonzero entry. There is certainly at least one such entry. To bring entry $A_{ij} = \alpha \neq 0$ to the head of the diagonal, row i is added to row j and column i is added to column j so as to have

$$\begin{array}{cc} & \begin{array}{cc} i & j \end{array} \\ \begin{array}{c} i \\ \\ j \end{array} & \begin{bmatrix} & \alpha \\ & \\ \alpha & \end{bmatrix} \end{array} \rightarrow \begin{array}{cc} & \begin{array}{cc} i & j \end{array} \\ \begin{array}{c} i \\ \\ j \end{array} & \begin{bmatrix} & \alpha \\ & \\ \alpha & 2\alpha \end{bmatrix} \end{array} \tag{6.159}$$

after which rows and columns are appropriately interchanged.

If submatrix $A_1 \neq O$, then the procedure is repeated on it, and on all subsequent nonzero diagonal submatrices until a diagonal form is reached, and since P in $D = P^T AP$ is nonsingular, $\text{rank}(A) = \text{rank}(D)$. End of proof.

Corollary 6.37. *Every symmetric matrix of rank r is congruent to the canonical*

$$D = \begin{bmatrix} I(p \times p) & & \\ & -I(r-p \times r-p) & \\ & & O(n-r \times n-r) \end{bmatrix}. \tag{6.160}$$

Proof. Symmetric row and column permutations rearrange the diagonals $d_1, d_2, \ldots, d_r$ so that positive entries come first, negative second, and zeroes last. Multiplication of the ith row and column of D in Theorem 6.36 by $|d_i|^{-\frac{1}{2}}$ produces the desired diagonal matrix. End of proof.

The following important theorem states that not only is r invariant under congruent transformations, but also index p.

Theorem (Sylvester's law of inertia) 6.38. *Index p in Corollary 6.37 is unique.*

Proof. Suppose not, and assume that symmetric matrix $A = A(n \times n)$ of rank r is congruent to both diagonal D_1 with p 1's and $(r-p)\ -1$'s, and to diagonal D_2 with q 1's and $(r-q)\ -1$'s, and let $p > q$.

By the assumption that D_1 and D_2 are congruent to A there exist nonsingular P_1 and P_2 so that $D_1 = P_1^TAP_1$, $D_2 = P_2^TAP_2$, and $D_2 = P^TD_1P$ where $P = P_1^{-1}P_2$.

For any $x = [x_1\ x_2\ \ldots\ x_n]^T$

$$\delta = x^TD_1x = x_1^2 + x_2^2 + \cdots + x_p^2 - x_{p+1}^2 - x_{p+2}^2 - \cdots - x_r^2 \tag{6.161}$$

while if $x = Py,\ y = [y_1\ y_2\ \ldots\ y_n]^T$, then

$$\delta = y^TPD_1Py = y^TD_2y = y_1^2 + y_2^2 + \cdots + y_q^2 - y_{q+1}^2 - \cdots - y_r^2. \tag{6.162}$$

Since P is nonsingular, $y = P^{-1}x$.

Set

$$y_1 = y_2 = \cdots = y_q = 0 \ \text{ and } \ x_{p+1} = x_{p+2} = \cdots = x_n = 0. \tag{6.163}$$

To relate $x' = [x_1\ x_2\ \ldots\ x_p]^T$ and $y' = [y_{q+1}\ y_{q+2}\ \ldots\ y_n]^T$, we solve $y = P^{-1}x$ in the partitioned form

$$\begin{array}{c} q \\ n-q \end{array} \begin{bmatrix} o \\ y' \end{bmatrix} = \begin{bmatrix} P'_{11} & P'_{12} \\ P'_{21} & P'_{22} \end{bmatrix} \begin{bmatrix} x' \\ o \end{bmatrix} \begin{array}{c} p \\ n-p \end{array} \qquad o = P'_{11}x',\ y' = P'_{21}x'. \tag{6.164}$$

The first homogeneous subsystem consists of q equations in p unknowns, and since $p > q$ a nontrivial solution $x' \neq o$ exists for it, and

$$\begin{aligned} \delta &= x_1^2 + x_2^2 + \cdots + x_p^2 > 0 \\ \delta &= -y_{q+1}^2 - y_{q+2}^2 - \cdots - y_r^2 \leq 0 \end{aligned} \tag{6.165}$$

which is absurd, and $p > q$ is wrong. So is an assumption $q > p$, and we conclude that $p = q$. End of proof.

Sylvester's theorem has important theoretical and computational consequences for the algebraic eigenproblem. Here are two corollaries.

Corollary 6.39. *Matrices $A = A^T$ and P^TAP, for any nonsingular P, have the same number of positive eigenvalues, the same number of negative eigenvalues, and the same number of zero eigenvalues.*

Proof. By virtue of being symmetric, both A and P^TAP are diagonalizable by an orthogonal similarity transformation. Let Q_1 and Q_2 be orthogonal matrices with which $Q_1^TAQ_1 = D_1$ and $Q_2^TP^TAPQ_2 = (PQ_2)^TA(PQ_2) = D_2$. Diagonal matrix D_1 holds on its diagonal the eigenvalues of A, and diagonal matrix D_2 holds on its diagonal the eigenvalues of P^TAP. By Sylvester's theorem D_1 and D_2 have the same rank r and the same index of positive entries p. End of proof.

Corollary 6.40. *If A is symmetric and B is symmetric and positive definite, then the particular eigenproblem $Ax = \lambda x$, and the general eigenproblem $Ax = \lambda Bx$ have the same number of positive, negative and zero eigenvalues.*

Proof. Since B is positive definite it may be factored as $B = LL^T$. With $x = L^{-T}x'$, $Ax = \lambda Bx$ becomes $L^{-1}AL^{-T}x' = \lambda x'$, and since A and $L^{-1}AL^{-T}$ are congruent the previous corollary guarantees the result. End of proof.

A good use for Sylvester's theorem is to count the number of eigenvalues of a symmetric matrix that are larger than a given value.

example. To determine the number of eigenvalues of symmetric matrix

$$A = \begin{bmatrix} 2 & -1 & \\ -1 & 2 & -1 \\ & -1 & 2 \end{bmatrix} \tag{6.166}$$

larger than 1, and the number of those less than 1.

Matrix $A-I$ has eigenvalues $\lambda_1-1, \lambda_2-1, \lambda_3-1$. A sequence of symmetric elementary operations is performed below on $A-I$ until it is transformed into a diagonal matrix

$$\begin{array}{cc} & \begin{matrix}1 & 2 & 3\end{matrix} \\ \begin{matrix}1\\2\\3\end{matrix} & \begin{bmatrix} 1 & -1 & \\ -1 & 1 & -1 \\ & -1 & 1 \end{bmatrix} \end{array} \rightarrow \begin{array}{cc} & \begin{matrix}1 & 2 & 3\end{matrix} \\ \begin{matrix}1\\2\\3\end{matrix} & \begin{bmatrix} 1 & -1 & \\ & & -1 \\ & -1 & 1 \end{bmatrix} \end{array} \rightarrow \begin{array}{cc} & \begin{matrix}1 & 2 & 3\end{matrix} \\ \begin{matrix}1\\2\\3\end{matrix} & \begin{bmatrix} 1 & & \\ & & -1 \\ & -1 & 1 \end{bmatrix} \end{array} \rightarrow$$

$$\begin{array}{cc} & \begin{matrix}1 & 2 & 3\end{matrix} \\ \begin{matrix}1\\3\\2\end{matrix} & \begin{bmatrix} 1 & & \\ & -1 & 1 \\ & & -1 \end{bmatrix} \end{array} \rightarrow \begin{array}{cc} & \begin{matrix}1 & 3 & 2\end{matrix} \\ \begin{matrix}1\\3\\2\end{matrix} & \begin{bmatrix} 1 & & \\ & 1 & -1 \\ & -1 & \end{bmatrix} \end{array} \rightarrow \begin{array}{cc} & \begin{matrix}1 & 3 & 2\end{matrix} \\ \begin{matrix}1\\3\\2\end{matrix} & \begin{bmatrix} 1 & & \\ & 1 & -1 \\ & & -1 \end{bmatrix} \end{array} \rightarrow \begin{array}{cc} & \begin{matrix}1 & 3 & 2\end{matrix} \\ \begin{matrix}1\\3\\2\end{matrix} & \begin{bmatrix} 1 & & \\ & 1 & \\ & & -1 \end{bmatrix} \end{array}. \tag{6.167}$$

Looking at the last diagonal matrix we conclude that two eigenvalues of A are larger than 1 and one eigenvalue is less than 1. Actually $\lambda_1 = 2 - \sqrt{2},\ \lambda_2 = 2,\ \lambda_3 = 2 + \sqrt{2}$.

exercises

6.12.1. Use Corollary 6.37 to assure that only one eigenvalue of

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

is less than 1.

6.12.2. Show that every skew-symmetric, $A = -A^T$, matrix of rank r is congruent to

$$B = \begin{bmatrix} S & & & & \\ & S & & & \\ & & \ddots & & \\ & & & S & \\ & & & & O \end{bmatrix}, \quad S = \begin{bmatrix} & 1 \\ -1 & \end{bmatrix}$$

where the number of S principal submatrices equals $r/2$.

Perform symmetric row and column elementary operations to bring

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

to that form.

6.13 Matrix polynomials

Every polynomial and power series of square matrix A is affected by

Theorem (Cayley-Hamilton) 6.41. *Let the n roots $\lambda_1, \lambda_2, \cdots, \lambda_n$ of characteristic equation*

$$p_n(\lambda) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_n - \lambda) = 0 \tag{6.168}$$

be the eigenvalues of matrix $A = A(n \times n)$. Then also

$$Z = p_n(A) = (\lambda_1 I - A)(\lambda_2 I - A) \cdots (\lambda_n I - A) = O. \tag{6.169}$$

That is, a matrix fulfils its own characteristic equation.

Proof. First we notice that $\lambda_i I - A$ and $\lambda_j I - A$ commute and that the factors of Z can be written in any order.

Let U be the unitary matrix that according to Schur's theorem causes the transformation $U^{-1}AU = T$, where T is upper-triangular with $T_{ii} = \lambda_i$. We use this to write

$$Z = (\lambda_1 I - A)UU^{-1}(\lambda_2 I - A)UU^{-1} \cdots UU^{-1}(\lambda_n I - A) \tag{6.170}$$

and obtain

$$U^{-1}ZU = (\lambda_1 I - T)(\lambda_2 I - T) \cdots (\lambda_n I - T). \tag{6.171}$$

If $U^{-1}ZU = O$, then also $Z = O$.

The last equation has the form

$$U^{-1}ZU = \begin{bmatrix} 0 & \times & \times & \times \\ & \times & \times & \times \\ & & \times & \times \\ & & & \times \end{bmatrix} \begin{bmatrix} \times & \times & \times & \times \\ & 0 & \times & \times \\ & & \times & \times \\ & & & \times \end{bmatrix} \cdots \begin{bmatrix} \times & \times & \times & \times \\ & \times & \times & \times \\ & & \times & \times \\ & & & 0 \end{bmatrix} \tag{6.172}$$

that we concisely write as $U^{-1}ZU = T_1 T_2 \cdots T_n$, T_i being upper-triangular and such that $(T_i)_{ii} = 0$. We prove that $U^{-1}ZU = O$ by showing that $U^{-1}ZUe_1 = o$, $U^{-1}ZUe_2 = o, \ldots, U^{-1}ZUe_n = o$. It is enough that we show it for the last equation. Indeed, if $e_n = [0\ 0\ 0\ \ldots\ 1]^T$, then

$$\begin{array}{c} T_n e_n = [\times\ \ldots\ \times\ \times\ \times\ 0],\ T_{n-1}T_n e_n = [\times\ \ldots\ \times\ \times\ 0\ 0]^T, \\ T_{n-2}T_{n-1}T_n e_n = [\times\ \ldots\ \times\ 0\ 0\ 0]^T, \ldots, T_1 T_2 \cdots T_{n-1} T_n e_n = [0\ \ldots\ 0\ 0\ 0\ 0]^T \end{array}. \tag{6.173}$$

End of proof.

Corollary 6.42. *If $A = A(n \times n)$, then A^k, $k \geq n$ is a polynomial function of A of degree less than n.*

Proof. Matrix A satisfies the nth degree polynomial equation

$$(-A)^n + a_{n-1}(-A)^{n-1} + \cdots + a_0 I = O \tag{6.174}$$

and therefore $A^n = p_{n-1}(A)$, and $A^{n+1} = p_n(A)$. Substitution of A^n into $p_n(A)$ leads back to $A^{n+1} = p_{n-1}(A)$ and then to $A^{n+2} = p_n(A)$. Proceeding in this way we reach $A^k = p_{n-1}(A)$. End of proof.

What this corollary means is that there is no truly infinite power series with matrices. We have encountered before

$$(I-A)^{-1} = I + A + A^2 + A^3 + \cdots + A^m + R_m \tag{6.175}$$

where if $\|A\| < 1$, then $R_m \to O$ as $n \to \infty$. Suppose that A is a 3×3 matrix satisfying

$$-A^3 + a_2A^2 - a_1A + Ia_0 = O,\ \ A^3 = a_2A^2 - a_1A + Ia_0. \tag{6.176}$$

Repeated substitution of this equation into eq. (6.175) results in

$$(I-A)^{-1} = (\alpha_2)_m A^2 + (\alpha_1)_m A + (\alpha_0)_m I + R_m \tag{6.177}$$

for any m.

Another interesting application of the Cayley-Hamilton theorem: If $a_0 \neq 0$, then

$$A^{-1} = \frac{1}{a_0}(A^2 - a_2A + a_1I). \tag{6.178}$$

It may happen that $A = A(n \times n)$ satisfies a polynomial equation of degree less than n. The lowest degree polynomial that A equates to zero is the *minimum polynomial* of A.

Theorem 6.43. *Every matrix $A = A(n \times n)$ satisfies a unique polynomial equation*

$$p_m(A) = (-A)^m + a_{m-1}(-A)^{m-1} + \cdots + a_0I = O \tag{6.179}$$

of minimum degree $m \leq n$.

Proof. To prove uniqueness assume that $p_m(A) = O$ and $p'_m(A) = O$ are different minimal polynomial equations. But then $p_m(A) - p'_m(A) = p_{m-1}(A)$ is in contradiction with the assumption that m is the lowest degree. Hence $p_m(A) = O$ is unique. End of proof.

Theorem 6.44. *The degree of the minimum polynomial of matrix A of rank r is at most $r+1$.*

Proof. Write the minimum rank factorization $A = BC$ of A with $B = B(n \times r), C = C(r \times n)$. It results from $A^{k+1} = BM^kC, M = M(r \times r) = CB$, and the fact that the characteristic equation $\lambda^r + a_{r-1}\lambda^{r-1} + \cdots + a_0 = 0$ of M is of degree r, that

$$B(M^r + a_{r-1}M^{r-1} + \cdots + a_0I)C = A^{r+1} + a_{r-1}A^r + \cdots + a_0A = O. \tag{6.180}$$

End of proof.

We shall leave it as an exercise to prove

Theorem 6.45. *If $\lambda_1, \lambda_2, \ldots, \lambda_k$ are the distinct eigenvalues of $A(n \times n),\ k \leq n$, then the minimum polynomial of A has the form*

$$p_m(\lambda) = (\lambda_1 - \lambda)^{m_1}(\lambda_2 - \lambda)^{m_2} \cdots (\lambda_k - \lambda)^{m_k} \tag{6.181}$$

where $m_i \geq 1,\ i = 1, 2, \ldots, k$.

In other words, the roots of the minimum polynomial of A are exactly the distinct eigenvalues of A, with multiplicities that may differ from those of the corresponding roots of the characteristic equation.

The following is a fundamental theorem of matrix iterative analysis.

Theorem 6.46. *$A^n \to \infty$ as $n \to \infty$ if for some i $|\lambda_i| > 1$, while $A^n \to O$ as $n \to \infty$ only if $|\lambda_i| < 1$ for all i. A^n tends to a limit as $n \to \infty$ only if $|\lambda| \leq 1$, with the only eigenvalue of modulus 1 being 1,and such that the algebraic multiplicity of $\lambda = 1$ equals its geometric multiplicity.*

Proof. Since $(XAX^{-1})^n = XA^nX^{-1}$ we may consider instead of A any other convenient matrix similar to it. Schur's Theorem 6.15 and Theorem 6.17 assure us of the existence of a block diagonal matrix similar to A with diagonal submatrices of the form $\lambda I + N$, where λ is an eigenvalue of A, and where N is strictly upper triangular and hence nilpotent. Since raising the block diagonal matrix to power n amounts to raising each block to that power we need consider only one typical block. Say that nilpotent N is such that $N^4 = O$. Then

$$(\lambda I + N)^n = \lambda^n I + n\lambda^{n-1}N + \frac{n(n+1)}{2!}\lambda^{n-2}N^2 + \frac{n(n+1)(n+2)}{3!}\lambda^{n-3}N^3. \tag{6.182}$$

If $|\lambda| > 1$, then $\lambda^n \to \infty$ as $n \to \infty$ and $(\lambda I + N)^n$ grows out of all bounds. If $|\lambda| < 1$, then $\lambda^n \to 0,\ n\lambda^n \to 0,\ n^2\lambda^n \to 0$ as $n \to \infty$ and $(\lambda I + N)^n \to O$ as $n \to \infty$.

In case $|\lambda| \leq 1$, at least one eigenvalue of A^n is of unit modulus for any n, which is impossible for a matrix with arbitrarily small entries. The reader should carefully work out the details of this last assertion.

For $|\lambda| = 1$ a limit to $(\lambda I + N)^n$ is attained only if $\lambda = 1$ and $N = O$. End of proof.

The modulus of the eigenvalue largest in magnitude is the *spectral radius* of A. Notice that since nonzero matrix A may have a null spectrum, the spectral radius does not qualify to serve as a norm for A.

exercises

6.13.1. Write the minimum polynomials of

$$D = \begin{bmatrix} 1 & & \\ & 1 & \\ & & -1 \end{bmatrix} \text{ and } R = \begin{bmatrix} 1 & 1 & 1 \\ & 1 & 1 \\ & & -1 \end{bmatrix}.$$

6.13.2. Write the minimum polynomial of

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

6.13.3. Write the minimum polynomials of

$$A = \begin{bmatrix} \lambda & 1 & & \\ & \lambda & 1 & \\ & & \lambda & 1 \\ & & & \lambda \end{bmatrix}, \ B = \begin{bmatrix} \lambda & 0 & & \\ & \lambda & 1 & \\ & & \lambda & 1 \\ & & & \lambda \end{bmatrix}, \ C = \begin{bmatrix} \lambda & 1 & & \\ & \lambda & 0 & \\ & & \lambda & 1 \\ & & & \lambda \end{bmatrix}, \ D = \begin{bmatrix} \lambda & 0 & & \\ & \lambda & 0 & \\ & & \lambda & 0 \\ & & & \lambda \end{bmatrix}.$$

6.13.4. Show that Matrix $A = A(n \times n)$ with eigenvalue λ of multiplicity n cannot have n linearly independent eigenvectors if $A - \lambda I \neq O$.

6.13.5. Let matrix A have the single eigenvalue λ_1 with a corresponding eigenvector v_1. Show that if v_1, v_2, v_3 are are linearly independent then so are the two vectors $v_2' = (A - \lambda_1 I)v_2$ and $v_3' = (A - \lambda_1 I)v_3$.

6.13.6. Let matrix $A = A(3 \times 3)$ with eigenvalues λ_1, λ_2 of multiplicities 1 and 2, respectively, be such that $(A - \lambda_1 I)(A - \lambda_2 I) = (A - \lambda_2 I)(A - \lambda_1 I) = O$. Let v_1 be the eigenvector of A corresponding to λ_1, and v_2, v_3 two vectors such that v_1, v_2, v_3 are linearly independent. Show that $(A - \lambda_1 I)v_2$ and $(A - \lambda_1 I)v_3$ are two linearly independent eigenvectors of A for λ_2. Matrix A has thus three linearly independent eigenvectors and is diagonalizable.

6.13.7. Let matrix $A = A(n \times n)$ have two eigenvalues λ_1 and λ_2 repeating k_1 and k_2 times, respectively, so that $k_1 + k_2 = n$. Show that if nonsingular matrix X exists so that XAX^{-1} is diagonal, then

$$X(A - \lambda_1 I)X^{-1}X(A - \lambda_2 I)X^{-1} = O, \text{ and } (A - \lambda_1 I)(A - \lambda_2 I) = O.$$

Consequently show that every nonzero vector $x' = (A - \lambda_2 I)x$ is in the nullspace of $A - \lambda_1 I$, that is, is such that $(A - \lambda_1 I)x' = o$, and hence is an eigenvector of A for λ_1.

Use all this to prove that if the distinct eigenvalues of $A = A(n \times n)$, discounting multiplicities, are $\lambda_1, \lambda_2, \ldots, \lambda_k$ then A is diagonalizable. if

$$(A - \lambda_1 I)(A - \lambda_2 I) \ldots (A - \lambda_k I) = O,$$

and conversely.

6.13.8. Show that if $A^k = I$ for some positive integer k, then A is diagonalizable.

6.13.9. Show that a nonzero nilpotent matrix is not diagonalizable.

6.13.10. Show that matrix A is diagonalizable if and only if its minimum polynomial does not have repeating roots.

6.13.11. Show that A and $P^{-1}AP$ have the same minimum polynomial.

6.13.12. Show that

$$\begin{bmatrix} 1 & \\ & -1 \end{bmatrix}^n, \begin{bmatrix} 1 & 1 \\ & 1 \end{bmatrix}^n, \begin{bmatrix} & 1 \\ 1 & \end{bmatrix}^n, \begin{bmatrix} & 1 \\ -1 & \end{bmatrix}^n$$

tend to no limit as $n \to \infty$.

6.13.13. If $A^2 = I$ $A \neq I$ what is the limit of A^n as $n \to \infty$?

6.13.14. What are the conditions on the eigenvalues of $A = A^T$ for $A^n \to B \neq O$ as $n \to \infty$?

6.13.15. When is the degree of the minimum polynomial of $A = A(n \times n)$ equal to n ?

6.13.16. For matrix

$$A = \begin{bmatrix} 6 & 4 & -1 \\ 3 & 7 & -1 \\ -6 & -8 & 5 \end{bmatrix}$$

find the smallest m such that $I, A, A^2, \ldots, A^m$ are linearly dependent. If for $A = A(n \times n)$ $m < n$, then matrix A is said to be *derogatory*.

6.14. Systems of differential equations

The non-stationary behavior of a multiparameter physical system is often described by a square system of linear differential equations with constant coefficients. The 3×3

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \dot{x} = Ax, \tag{6.183}$$

where overdot means differentiation with respect to time t, is such a system. Realistic systems can become considerably larger making their solution by hand impractical.

If matrix A in typical example (6.183) happens to have three linearly independent eigenvectors v_1, v_2, v_3 with three corresponding eigenvalues $\lambda_1, \lambda_2, \lambda_3$, then matrix $V = [v_1\ v_2\ v_3]$ diagonalizes A to the effect that $V^{-1}AV = D$ is such that $D_{11} = \lambda_1, D_{22} = \lambda_2,\ D_{33} = \lambda_3$. Linear transformation $x = Vy$ decouples system (6.183) into $\dot{y} = V^{-1}AVy$,

$$\begin{bmatrix} \dot{y}_1 \\ \dot{y}_2 \\ \dot{y}_3 \end{bmatrix} = \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \tag{6.184}$$

in which $\lambda_1, \lambda_2, \lambda_3$ may be real or complex, distinct or repeating. Each differential equation in system (6.184) is solved separately to yield $y_1 = c_1 e^{\lambda_1 t},\ y_2 = c_2 e^{\lambda_2 t},\ y_3 = c_3 e^{\lambda_3 t}$ for the three arbitrary constants of integration c_1, c_2, c_3. If λ_i is real, then y_i is exponential, but if λ_i is complex, then y_i turns trigonometric. In case real eigenvalue $\lambda_i > 0$, or complex eigenvalue $\lambda_i = \alpha_i + i\beta_i$ is with a positive real part, $\alpha_i > 0$, component $y_i = y_i(t)$ of solution vector y inexorably grows with the passage of time. When this happens to at least one eigenvalue of A, the system is said to be unstable, whereas if $\alpha_i < 0$ for all i, solution $y = y(t)$ subsides with time and the system is stable.

Returning to x we determine that

$$x = c_1 v_1 e^{\lambda_1 t} + c_2 v_2 e^{\lambda_2 t} + c_3 v_3 e^{\lambda_3 t} \tag{6.185}$$

and $x(0) = x_0 = c_1 v_1 + c_2 v_2 + c_3 v_3$. What we just did for the 3×3 system can be done for any $n \times n$ system as long as matrix A has n linearly independent eigenvectors.

Matters become more difficult when A fails to have n linearly independent eigenvectors. Consider the Jordan system

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} \lambda & 1 & \\ & \lambda & 1 \\ & & \lambda \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \dot{x} = Jx \tag{6.186}$$

with matrix J that has three equal eigenvalues but only one eigenvector, $v_1 = e_1$. Because the last equation is in x_3 only it may be immediately solved. Substitution of the computed x_3 into the second equation turns it into a nonhomogeneous linear equation that is immediately solved for x_2. Unknown function $x_1 = x_1(t)$ is likewise obtained from the first equation and we end up with

$$\begin{aligned} x_1 &= (c_1 + c_2 t + \frac{1}{2} c_3 t^2) e^{\lambda t} \\ x_2 &= (c_2 + c_3 t) e^{\lambda t} \\ x_3 &= c_3 e^{\lambda t} \end{aligned} \tag{6.187}$$

in which c_1, c_2, c_3 are arbitrary constants. Equation (6.187) can be written in vector fashion as

$$x = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} e^{\lambda t} + \begin{bmatrix} c_2 \\ c_3 \\ 0 \end{bmatrix} t e^{\lambda t} + \begin{bmatrix} \frac{1}{2} c_3 \\ 0 \\ 0 \end{bmatrix} t^2 e^{\lambda t} = w_1 e^{\lambda t} + w_2 t e^{\lambda t} + w_3 t^2 e^{\lambda t} \tag{6.188}$$

and $w_1 = x_0 = x(0)$.

But there is no real need to compute generalized eigenvectors, nor bring matrix A into Jordan form before proceeding to solve system (6.183). Repeated differentiation of $\dot{x} = Ax$ with back substitutions yields

$$x = Ix, \ \dot{x} = Ax, \ \ddot{x} = A^2 x, \ \dddot{x} = A^3 x \tag{6.189}$$

and according to the Cayley-Hamilton theorem, Theorem 6.41, numbers $\alpha_0, \alpha_1, \alpha_2$ exist so that

$$\dddot{x} + \alpha_2 \ddot{x} + \alpha_1 \dot{x} + \alpha_0 x = (A^3 + \alpha_2 A^2 + \alpha_1 A + \alpha_0 I) x = o. \tag{6.190}$$

Each of the unknown functions of system (6.183) satisfies by itself a third-order linear differential equation with the same coefficients as those of the characteristic equation

$$\lambda^3 + \alpha_2 \lambda^2 + \alpha_1 \lambda + \alpha_0 = 0 \tag{6.191}$$

of matrix A.

Suppose that all roots of eq.(6.191) are equal. Then

$$\begin{aligned} x &= w_1 e^{\lambda t} + w_2 t e^{\lambda t} + w_3 t^2 e^{\lambda t} \\ \dot{x} &= w_1 \lambda e^{\lambda t} + w_2 (e^{\lambda t} + \lambda t e^{\lambda t}) + w_3 (2te^{\lambda t} + \lambda t^2 e^{\lambda t}) \\ \ddot{x} &= w_1 \lambda^2 e^{\lambda t} + w_2 (2\lambda e^{\lambda t} + \lambda^2 t e^{\lambda t}) + w_3 (2e^{\lambda t} + 4\lambda t e^{\lambda t} + \lambda^2 t^2) e^{\lambda t} \end{aligned} \tag{6.192}$$

and

$$\begin{aligned} x(0) &= x_0 = w_1 \\ \dot{x}(0) &= Ax_0 = \lambda w_1 + w_2 \\ \ddot{x}(0) &= A^2 x_0 = \lambda^2 w_1 + 2\lambda w_2 + 2w_3. \end{aligned} \tag{6.193}$$

Constant vectors w_1, w_2, w_3 are readily expressed in terms of the initial condition vector x_0 as

$$w_1 = x_0,\ w_2 = (A - \lambda I)x_0,\ w_3 = \frac{1}{2}(A - \lambda I)^2 x_0 \tag{6.194}$$

and the solution of the 3×3 system is thereby completed.

Examples.

1. Consider the system

$$\dot{x} = \begin{bmatrix} 1 & 1 & & \\ & 1 & 0 & \\ & & 1 & 1 \\ & & & 1 \end{bmatrix} x, \quad \dot{x} = Ax \tag{6.195}$$

but ignore the fact that A is in Jordan form. All eigenvalues of A are equal to 1, the characteristic equation of A being $(\lambda - 1)^4 = 0$, and

$$x = w_1 e^t + w_2 t e^t + w_3 t^2 e^t + w_4 t^3 e^t. \tag{6.196}$$

Repeating the procedure that led to equations (6.192) and (6.193) but with the inclusion of $\dddot{x}$ we obtain

$$w_1 = x_0,\ w_2 = (A - \lambda I)x_0,\ w_3 = \frac{1}{2}(A - \lambda I)^2 x_0,\ w_4 = \frac{1}{6}(A - \lambda I)^3 x_0 \tag{6.197}$$

and if $x_0^T = [c_1\ c_2\ c_3\ c_4]$ for the arbitrary constants c_1, c_2, c_3, c_4, then

$$w_1 = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix}, w_2 = \begin{bmatrix} c_2 \\ \\ c_4 \\ \\ \end{bmatrix},\ w_3 = o,\ w_4 = o. \tag{6.198}$$

and

$$
\begin{aligned}
x_1 &= c_1 e^t + c_2 t e^t \\
x_2 &= c_2 e^t \\
x_3 &= c_3 e^t + c_4 t e^t \\
x_4 &= c_4 e^t
\end{aligned}
\tag{6.199}
$$

disclosing to us, in effect, the Jordan form of A.

2. Suppose that the solution $x = x(t)$ to $\dot{x} = Ax,\ A = A(5 \times 5)$, is

$$
\begin{aligned}
x = c_1 \begin{bmatrix} 1 \\ \\ \\ \\ \\ \end{bmatrix} e^t + c_2 \left(\begin{bmatrix} \\ 1 \\ \\ \\ \\ \end{bmatrix} e^t + \begin{bmatrix} 1 \\ \\ \\ \\ \\ \end{bmatrix} t e^t \right) + c_3 \begin{bmatrix} \\ \\ 1 \\ \\ \\ \end{bmatrix} e^t \\
+ c_4 \left(\begin{bmatrix} \\ \\ \\ 1 \\ \\ \end{bmatrix} e^t + \begin{bmatrix} \\ \\ 1 \\ \\ \\ \end{bmatrix} t e^t \right) + c_5 \left(\begin{bmatrix} \\ \\ \\ \\ 1 \end{bmatrix} e^t + \begin{bmatrix} \\ \\ \\ 1 \\ \\ \end{bmatrix} t e^t + \begin{bmatrix} \\ \\ 1 \\ \\ \\ \end{bmatrix} \frac{1}{2} t^2 e^t \right)
\end{aligned}
\tag{6.200}
$$

or

$$
\begin{aligned}
x = c_1 v_1 e^t + c_2 (v_2 e^t + v_1 t e^t) + c_3 v_3 e^t \\
+ c_4 (v_4 e^t + v_3 t e^t) + c_5 (v_5 e^t + v_4 t e^t + v_3 \frac{1}{2} t^2 e^t)
\end{aligned}
\tag{6.201}
$$

so that

$$
\begin{aligned}
x(0) &= c_1 v_1 + c_2 v_2 + c_3 v_3 + c_4 v_4 + c_5 v_5 \\
\dot{x}(0) &= c_1 v_2 + c_2 (v_1 + v_2) + c_3 v_3 + c_4 (v_3 + v_4) + c_5 (v_4 + v_5).
\end{aligned}
\tag{6.202}
$$

Writing $\dot{x}(0) = Ax(0)$ we obtain from eq.(6.2042) that

$$
\begin{aligned}
c_1 (Av_1 - v_1) + c_2 (Av_2 - v_2 - v_1) + c_3 (Av_3 - v_3) \\
+ c_4 (Av_4 - v_4 - v_3) + c_5 (Av_5 - v_5 - v_4) = o
\end{aligned}
\tag{6.203}
$$

and since c_1, c_2, c_3, c_4, c_5 are arbitrary it must so be that

$$
\begin{aligned}
&(A - I) v_1 = o,\ (A - I) v_2 = v_1 \\
&(A - I) v_3 = o,\ (A - I) v_4 = v_3,\ (A - I) v_5 = v_4.
\end{aligned}
\tag{6.204}
$$

Vectors v_1 and v_3 are two linearly independent eigenvectors of A, corresponding to $\lambda = 1$ of multiplicity five, while v_2, v_4, v_5 are generalized eigenvectors; v_2 emanating from v_1, and

v_4, v_5 emanating out of v_3. Hence the Jordan form of A consists of two blocks: a 2×2 block on top of a 3×3 block.

3. If it so happens that the minimum polynomial of $A = A(n \times n)$ is $A^2 - I = O$, then vector $x = x(t)$ of $\dot{x} = Ax$ satisfies the equation $\ddot{x} - x = o$, and $x = w_1 e^t + w_2 e^{-t}$, with $w_1 = 1/2(I + A)x_0, w_2 = 1/2(I - A)x_0$.

4. Consider the system

$$\dot{x} = \begin{bmatrix} 1 & -1 & \\ -1 & 2 & -1 \\ & -1 & 1 \end{bmatrix} x \tag{6.205}$$

written in full as

$$\begin{aligned} \dot{x}_1 &= x_1 - x_2 \\ \dot{x}_2 &= -x_1 + 2x_2 - x_3 \,. \\ \dot{x}_3 &= -x_2 + x_3 \end{aligned} \tag{6.206}$$

Repeated differentiation of the first equation with substitution of $\dot{x}_2, \dot{x}_3$ from the other two equations produces

$$\begin{aligned} x_1 &= x_1 \\ \dot{x}_1 &= x_1 - x_2 \\ \ddot{x}_1 &= 2x_1 - 3x_2 + x_3 \\ \dddot{x}_1 &= 5x_1 - 9x_2 + 4x_3 \end{aligned} \tag{6.207}$$

that we linearly combine as

$$\dddot{x}_1 + \alpha_2 \ddot{x}_1 + \alpha_1 \dot{x}_1 + \alpha_0 x_1 = x_1(5 + 2\alpha_2 + \alpha_1 + \alpha_0) + x_2(-9 - 3\alpha_2 - \alpha_1) + x_3(4 + \alpha_2). \tag{6.208}$$

Equating the coefficients of x_1, x_2, x_3 on the right-hand side of the above equation to zero we obtain the third-order differential equation $\dddot{x}_1 - 4\ddot{x}_1 + 3\dot{x}_1 = 0$ for x_1 without explicit recourse to the Cayley-Hamilton theorem. It results that $x_1 = c_1 + c_2 e^t + c_3 e^{3t}$ which we may now put into $x_2 = -\dot{x}_1 + x_1, x_3 = \ddot{x}_1 - 3\dot{x}_1 + x_1$ to have the two other solutions.

exercises

6.14.1. Solve the linear differential systems

$$\dot{x} = \begin{bmatrix} & 1 \\ 1 & \end{bmatrix} x, \ \dot{x} = \begin{bmatrix} & 1 \\ -1 & \end{bmatrix} x, \ \ddot{x} = \begin{bmatrix} & 1 \\ 1 & \end{bmatrix} x, \ \dot{x} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} x, \ \dot{x} = \begin{bmatrix} -1 & 2 \\ 2 & -1 \end{bmatrix} x$$

and examine the behavior of each $x = x(t)$ as $t \rightarrow \infty$.

6.14.2. The solution of the initial value problem

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad x(0) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

is $x_1 = (1+t)e^t, x_2 = e^t$.

Solve

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ & 1+\epsilon \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad x(0) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

and examine the solution as $\epsilon \rightarrow 0$.

6.15 Difference and tridiagonal matrices

The highly structured tridiagonal finite difference matrices of Chapter 3 allow the explicit computation of their eigenvalues and eigenvectors. Consider the $n \times n$ stiffness matrix

$$A = \frac{1}{h} \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & \ddots & -1 & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix} \qquad h = \frac{1}{n+1} \tag{6.209}$$

for the string fixed at both ends.

Writing $(A - \lambda I)x = o$ equation by equation,

$$-x_{k-1} + (2 - \lambda h)x_k - x_{k+1} = 0 \quad k = 1, 2, \ldots, n \tag{6.210}$$

$$x_0 = x_{n+1} = 0$$

we observe that the interior difference equations are solved by $x_k = e^{ik\theta}$, $i = \sqrt{-1}$, provided that

$$\lambda h = 2(1 - \cos\theta). \tag{6.211}$$

Because the finite difference equations are solved by both $\cos k\theta$ and $\sin k\theta$, and since the equations are linear they are also solved by the linear combination

$$x_k = \alpha_1 \cos k\theta + \alpha_2 \sin k\theta. \tag{6.212}$$

Boundary condition $x_0 = 0$ is satisfied with $\alpha_1 = 0$. To satisfy the second boundary condition $x_{n+1} = 0$ we must have

$$\alpha_2 \sin(n+1)\theta = 0 \tag{6.213}$$

and we avoid the trivial solution by taking

$$(n+1)\theta = \pi,\ 2\pi, \ldots, n\pi \tag{6.214}$$

so that

$$\lambda_j h = 2(1 - \cos \pi j h) \quad h = 1/(n+1) \tag{6.215}$$

or

$$\lambda_j = 4h^{-1} \sin^2 \frac{\pi j h}{2} \quad j = 1, 2, \ldots, n \tag{6.216}$$

which are the n eigenvalues of A. No new ones appear with $j > n$. The corresponding eigenvectors are the columns of the $n \times n$ matrix X

$$X_{ij} = \sin \frac{\pi i j}{(n+1)} \tag{6.217}$$

that we verify to be orthogonal.

As the string is divided into smaller and smaller segments to improve the approximation accuracy, matrix A increases in size. When $h << 1$, $\sin(\pi h/2) = \pi h/2$, $\lambda_1 = \pi^2 h$, $\lambda_n = 4h^{-1}$, and

$$\kappa_2(A) = \frac{\lambda_n}{\lambda_1} = \frac{4}{\pi^2} h^{-2}. \tag{6.218}$$

Matrix A becomes ill-conditioned as n increases. This is a basic computational fact of life for finite difference and finite element matrices.

Matrix A we dealt with above is for a fixed-fixed string and we expect no zero eigenvalues. Releasing one end point of the string still leaves the matrix nonsingular, but release of also the second end point gives rise to a zero eigenvalue corresponding to the up and down rigid body motion of the string. In all three cases matrix A has n distinct eigenvalues with possibly one zero eigenvalue. We shall presently show that this property is shared by all symmetric tridiagonal matrices.

We refer to the $n \times n$ tridiagonal matrix

$$T = \begin{bmatrix} \alpha_1 & \gamma_2 & & & \\ \beta_2 & \alpha_2 & \gamma_3 & & \\ & \beta_3 & \alpha_3 & \gamma_4 & \\ & & \beta_4 & \ddots & \ddots \\ & & & \ddots & \alpha_n \end{bmatrix} \tag{6.219}$$

and enter the following

Definition. *Tridiagonal matrix T is irreducible if $\beta_2, \beta_3, \ldots, \beta_n$ and $\gamma_2, \gamma_3, \ldots, \gamma_n$ are all nonzero, otherwise it is reducible.*

Theorem 6.47.

1. For any irreducible tridiagonal matrix T there exists a nonsingular diagonal matrix D so that $DT = T'$ is an irreducible symmetric tridiagonal matrix.

2. For any tridiagonal matrix T with $\beta_i \gamma_i > 0 \;\; i = 2, 3, \ldots, n$ there exists a nonsingular diagonal matrix D so that T' in the similarity transformation $T' = DTD^{-1}$ is a symmetric tridiagonal matrix.

3. For any irreducible symmetric tridiagonal matrix T there exists a nonsingular diagonal matrix D so that $DTD = T'$ is of the form

$$T' = \begin{bmatrix} \alpha_1 & 1 & & \\ 1 & \alpha_2 & 1 & \\ & 1 & \ddots & 1 \\ & & 1 & \alpha_n \end{bmatrix}. \tag{6.220}$$

Proof.

1. $D_{ii} = d_i, d_1 = 1, d_{i+1} = (\gamma_{i+1}/\beta_{i+1})d_i$.

2. $D_{ii} = d_i, d_1 = 1, d_i = (\gamma_2\gamma_3 \ldots \gamma_i/\beta_2\beta_3 \ldots \beta_i)^{1/2}$.

3. $D_{ii} = d_i, d_1 = 1, \; d_i d_{i+1} = 1/\beta_{i+1}$.

End of proof.

Theorem 6.48. *If tridiagonal matrix T is irreducible, then nullity (T) is either 0 or 1, and consequently rank $(T) \geq n - 1$.*

Proof. Elementary eliminations that use the (nonzero) γ's on the upper diagonal as pivots, followed by row interchanges produce

$$T' = \begin{bmatrix} \times & & & & \\ \times & 1 & & & \\ \times & & 1 & & \\ \times & & & 1 & \\ \times & & & & 1 \end{bmatrix} \tag{6.221}$$

that is equivalent to T. If $T'_{11} = 0$, then rank (T') = rank $(T) = n-1$, nullity (T') = nullity $(T) = 1$, and if $T'_{11} \neq 0$, then rank (T') = rank $(T) = n$, nullity (T') = nullity $(T) = 0$. End of proof.

Theorem 6.49. *Irreducible symmetric tridiagonal matrix $T = T(n \times n)$ has n distinct real eigenvalues.*

Proof. The nullity of $T - \lambda I$ is zero when λ is not an eigenvalue of T, and is 1 when $\lambda = \lambda_i$ is an eigenvalue of T. This means that there is only one eigenvector corresponding to λ_i. By the assumption that T is symmetric there is an orthogonal Q so that $Q^T(T - \lambda I)Q = D$, where $D_{ii} = \lambda_i - \lambda$. Hence, rank $(T - \lambda I)$ = rank $(D) = n - 1$ if $\lambda = \lambda_i$, and the eigenvalues of T are distinct. End of proof.

Theorem 6.49 does not say *how* distinct the eigenvalues of symmetric irreducible T are, depending on the relative size of the off-diagonal entries $\beta_2, \beta_3, \ldots, \beta_n$. Matrix

$$T = \begin{bmatrix} 3 & 1 & & & & & \\ 1 & 2 & 1 & & & & \\ & 1 & 1 & 1 & & & \\ & & 1 & 0 & 1 & & \\ & & & 1 & 1 & 1 & \\ & & & & 1 & 2 & 1 \\ & & & & & 1 & 3 \end{bmatrix}, \tag{6.222}$$

in which $\alpha_i = m + 1 - i \;\; i = 1, 2, \ldots, m+1, \alpha_{n-i+1} = \alpha_i \;\; i = 1, 2, \ldots, m$ of order $n = 2m + 1$ looks innocent, but it is known to have eigenvalues that may differ by a mere $(m!)^{-2}$.

exercises

6.15.1. Show that the eigenvalues of the $n \times n$

$$A = \begin{bmatrix} & \alpha & & \\ \beta & & \alpha & \\ & \beta & & \alpha \\ & & \beta & \end{bmatrix} \quad \alpha, \beta > 0$$

are

$$\lambda_j = 2\sqrt{\alpha\beta}\cos\frac{j\pi}{n+1}, \quad j = 1, 2, \ldots, n.$$

6.15.2. Show that the eigenvalues of

$$T = \begin{bmatrix} \alpha_1 & \beta_2 & & \\ \beta_2 & \alpha_2 & \beta_3 & \\ & \beta_3 & \alpha_3 & \beta_4 \\ & & \beta_4 & \alpha_4 \end{bmatrix} \text{ and } T' = \begin{bmatrix} \alpha_1 & |\beta_2| & & \\ |\beta_2| & \alpha_2 & |\beta_3| & \\ & |\beta_3| & \alpha_3 & |\beta_4| \\ & & |\beta_4| & \alpha_4 \end{bmatrix}$$

are the same.

6.15.3. Show that $x = [0 \ \times \ \times \ \ldots \ \times]^T$ cannot be an eigenvector of a symmetric, irreducible tridiagonal matrix, nor $x = [\times \ \times \ \times \ \ldots \ \times \ 0]$.

6.15.4. Show that if $x = [x_1 \ 0 \ x_2 \ldots x_n]^T$ is an eigenvector of irreducible tridiagonal T, $Tx = \lambda x$, then $\lambda = T_{ii} = \alpha_1$. Also, that if $x = [x_1 \ x_2 \ 0 \ x_3 \ldots x_n]^T$ is an eigenvector of T, then λ is an eigenvalue of

$$T_2 = \begin{bmatrix} \alpha_1 & \beta_2 \\ \beta_2 & \alpha_2 \end{bmatrix}.$$

Continue in this manner and prove that the eigenvector corresponding to an extreme eigenvalue of T has no zero components.

Notice that this does not preclude the possibility of $x_j \to 0$ as $n \to \infty$ for some entry of normalized eigenvector x.

6.16 Variational principles

We return to matters considered in the opening section of this chapter.

When 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 $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$, with orthogonal eigenvectors $x_1, x_2, \ldots, x_n$. Then*

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

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

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

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

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

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*

$$\begin{matrix} \min\limits_{x \neq o} \dfrac{x^T A x}{x^T x} \le \lambda_{k+1} \\ \\ \max\limits_{x \neq o} \dfrac{x^T A x}{x^T x} \ge \lambda_{n-k} \end{matrix} \qquad x^T p_1 = x^T p_2 = \cdots = x^T p_k = 0. \tag{6.237}$$

Proof. We order the eigenvalues of A in the ascending order $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$, 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 $\lambda(x) \le \lambda_2$, and obviously $\min \lambda(x) \le \lambda_2$.

On the other hand if we choose $\alpha_1 = \alpha_2 = \cdots = \alpha_{n-2} = 0$, then we are left with the constraint equation $\alpha_{n-1} x_{n-1}^T p_1 + \alpha_n x_n^T p_1 = 0$, 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 $\lambda(x) \ge \lambda_{n-1}$, and obviously $\max \lambda(x) \ge \lambda_{n-1}$.

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 $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$ with corresponding eigenvectors $x_1, x_2, \ldots, x_n$. If*

$$\lambda'_k = \min_{x \neq o} \lambda(x), \begin{cases} x^T x_1 = x^T x_2 = \cdots = x^T x_{k-1} = 0 \\ x^T p_1 = \cdots = x^T p_m = 0 \end{cases}, \quad 1 \le k \le n - m \tag{6.239}$$

then

$$\lambda_k \le \lambda'_k \le \lambda_{k+m} \tag{6.240}$$

In particular, for $m = 1$

$$\lambda_1 \le \lambda'_1 \le \lambda_2, \quad \lambda_2 \le \lambda'_2 \le \lambda_3, \ \cdots, \ \lambda_{n-1} \le \lambda'_n \le \lambda_n. \tag{6.241}$$

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* $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$, *be partitioned as*

$$A = \begin{array}{c} \begin{array}{cc} n-m & m \end{array} \\ \left[\begin{array}{cc} A' & C \\ C^T & B \end{array}\right] \end{array} \begin{array}{l} \\ n-m \\ m. \end{array} \tag{6.242}$$

If $\lambda'_1 \le \lambda'_2 \le \cdots \le \lambda'_{n-m}$ *are the eigenvalues of* A' *then*

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

Proof. $\min x'^T A' x' / x'^T x'$, $x' \in R^{n-m}$ can be interpreted as minimization of $x^T Ax/x^T x$, $x \in R^n$ under the m constraints $x^T e_{n-m+1} = \cdots = x^T e_n = 0$. 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 $\min\limits_{\lambda} r^T r = x^T A^2 x - (x^T Ax)^2$. 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.

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

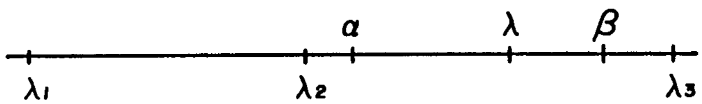

Fig. 6.3

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

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

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

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

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

$$\epsilon_1 = \frac{\|r_1\|}{\|x'_1\|} = \frac{\sqrt{5}}{14} = 0.1597,\ \ \epsilon_2 = \frac{\|r_2\|}{\|x'_2\|} = \frac{\sqrt{2}}{9} = 0.1571,\ \ \epsilon_3 = \frac{\|r_3\|}{\|x'_3\|} = \frac{\sqrt{5}}{9} = 0.2485 \tag{6.267}$$

and have from Theorem 6.57 that

$$0.0546 \le \lambda_1 \le 0.374,\ \ 1.398 \le \lambda_2 \le 1.713,\ \ 2.974 \le \lambda_3 \le 3.471. \tag{6.268}$$

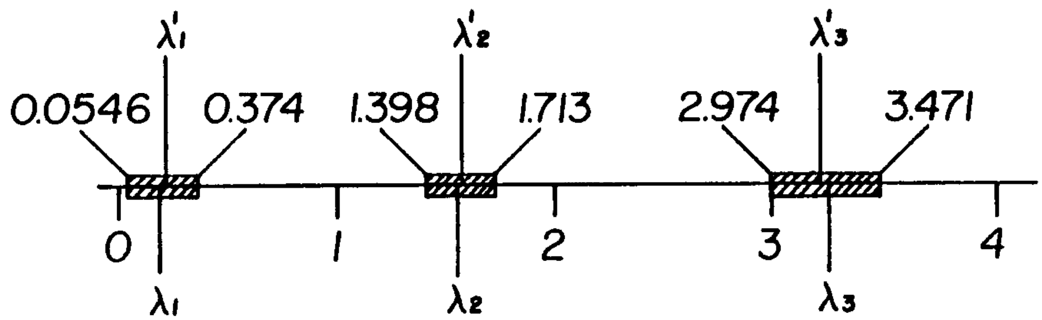

Fig. 6.4

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.

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 $\lambda_1 \leq \lambda_1'$, and hence we select $\alpha = \lambda_1,\ \beta = 1.398$ in Kato's inequality so as to have

$$(\lambda_1' - \lambda_1)(1.398 - \lambda_1') \leq \epsilon_1^2 \tag{6.269}$$

and

$$0.1927 \leq \lambda_1 \leq 0.2143. \tag{6.270}$$

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

$$\lambda_2' \leq \lambda_2 \leq \lambda_2' + \frac{\epsilon_2^2}{\lambda_2' - \lambda_1'} \tag{6.271}$$

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 $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$, 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 $\lambda = x^T Ax$, then*

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

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*

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

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

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

With the proper sign choice for $x, \frac{1}{2}e_j^T e_j < 1$, 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*

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

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

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

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*

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

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

$$\lambda = x^HAx = A_{11}x_1\overline{x}_1 + A_{12}\overline{x}_1x_2 + A_{21}x_1\overline{x}_2 + \cdots + A_{nn}x_n\overline{x}_n \tag{6.295}$$

and

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

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

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

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

$$\begin{array}{lclcl} \min\limits_x \dfrac{x^TCx}{x^Tx} & \ge & \min\limits_x \dfrac{x^TAx}{x^Tx} & + & \min\limits_x \dfrac{x^TBx}{x^Tx} \\ x^Ta_1 = \cdots = x^Ta_{i-1} = 0 & & x^Ta_1 = \cdots = x^Ta_{i-1} = 0 & & x^Tb_1 = \cdots = x^Tb_{j-1} = 0 \\ x^Tb_1 = \cdots = x^Tb_{j-1} = 0. & & & & \end{array} \tag{6.304}$$

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

$$\begin{array}{lclcl} \max\limits_x \dfrac{x^TCx}{x^Tx} & \le & \max\limits_x \dfrac{x^TAx}{x^Tx} & + & \max\limits_x \dfrac{x^TBx}{x^Tx} \\ x^Ta_{i+1} = \cdots = x^Ta_n = 0 & & x^Ta_{i+1} = \cdots = x^Ta_n = 0 & & x^Tb_{j+1} = \cdots = x^Tb_n = 0 \\ x^Tb_{j+1} = \cdots = x^Tb_n = 0. & & & & \end{array} \tag{6.305}$$

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 $-\epsilon \le \beta_1, \ \beta_n \le \epsilon$. Then

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

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 $Ax - \lambda x = r,\ \lambda = x^TAx/x^Tx$, then $x^Tr = 0$. Also, that $Bx = \lambda x$ for

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

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 $\lambda(x) = x^TAx/x^Tx$.

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 $\sigma_i = (\lambda_i(A^TA))^{1/2}$ 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

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

Generalize to more deletions.

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

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

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

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.

6.17 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*

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

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 $x = [x_1\ x_2\ \dots\ x_n]^T$ 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 $A'_{ii} = 0$.

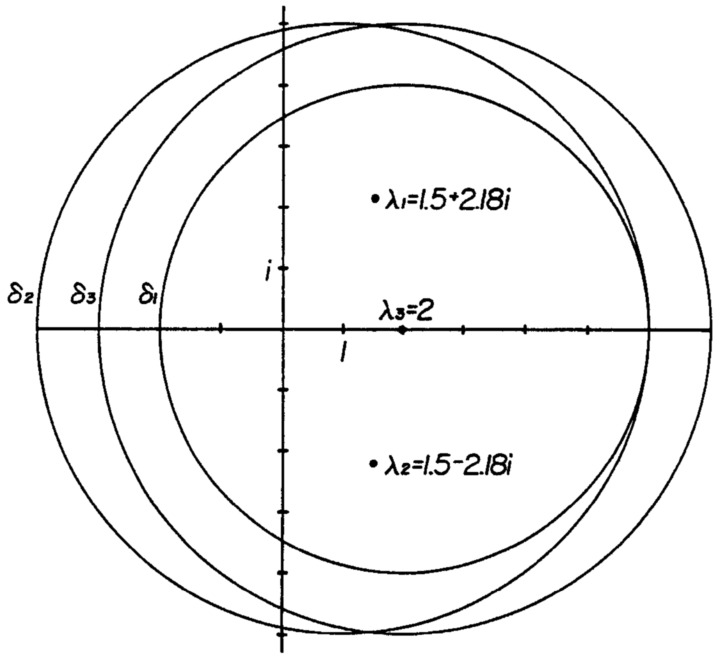

Fig. 6.5

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

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 $0 \leq \lambda \leq 4$ 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 $-4 \leq \lambda \leq 16$, 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, $A \geq O$, if $A_{ij} \geq 0$ 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 $\alpha_1 \geq 0,\ \alpha_2 > 0,\ \alpha_3 > 0, \ldots, \alpha_n > 0,\ \alpha_{n+1} \geq 0$. Then eigenvector x corresponding to its minimal eigenvalue is positive, $x > o$.

Proof. If $x = [x_1\ x_2\ \ldots\ x_n]^T$ 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 $x = [1\ 1\ \ldots\ 1]^T$. 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 $e = [1\ 1\ 1\ \ldots, 1]^T$. 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

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

and

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

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

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

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

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

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 $-n\epsilon \leq \beta \leq n\epsilon$. If $C = A + B$, then according to Theorem 6.62 $|\gamma_i - \alpha_i| \leq n\epsilon$ where $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_n$ are the eigenvalues of A and $\gamma_1 \leq \gamma_2 \leq \cdots \leq \gamma_n$ 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 $A(\tau) = D + \tau A'$ $0 \leq \tau \leq 1$. 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 $0 \leq \tau \leq 1$. 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 $\delta_1 \cup \delta_2$

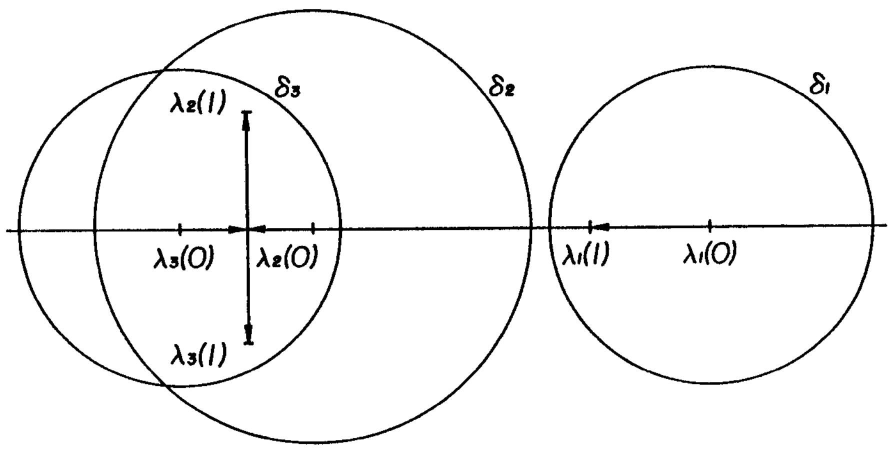

Fig. 6.6

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

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

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 $x = u + iv, u^T u = v^T v = 1$, 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

$v^T v = 1$. Presently,

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

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

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

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

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

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 $i = 1, 2, \dots, n$ 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

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

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

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

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 $\text{grad}\rho(y) = o$ 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 $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. If $\rho_1 \leq \rho_2 \leq \cdots \leq \rho_m$ 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 $\rho_k \geq \lambda_k$ 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 $|\rho_j - \lambda| \leq \|r_j\| / \|{x_j}'\|$ 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 $V^TAVy = \rho V^TVy$ 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^TV itself revealed a spectral condition number $\kappa(V^TV) = (1.5m)!$ which means $\kappa = 6 \cdot 5\ 10^{15}$ for $m = 12$, and all the high accuracy used could not save V^TV 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^TAx/x^Tx is minimal.

6.19 Round-off errors

In this section we consider the basic round-off perturbation effects on eigenvalues, mainly for difference matrices that have large eigenvalue spreads.

Even if a matrix is exactly given the mere act of writing it into the computer perturbs it slightly, and the same happens to any symbolically given eigenvector. Formation of Ax_j is also done in finite arithmetic and is not exact. The perturbations may be minute but their effect is easily magnified to disastrous proportions.

Suppose first that matrix $A = A^T$ is written exactly, that the arithmetic is exact, and that the only change in eigenvector x_j is due to round-off: $x'_j = x_j + \epsilon w,\ \|w\| = 1,\ |\epsilon| << 1.$

Then

$$\lambda_j' - \lambda_j = \epsilon^2(w^TAw - \lambda_j),\ \lambda_j' = x_j'^T Ax_j'/x_j'^T x_j' \tag{6.360}$$

is very small if ϵ is small. Round-off error damage does not come from eigenvector perturbations but rather from the perturbation of A and the inaccurate formation of the product Ax_j.

Both effects are accounted for by assuming exact arithmetic on the perturbed $A + A'$. Now

$$\lambda_j' = x_j^T(A + A')x_j = \lambda_j - x_j^T A' x_j \tag{6.361}$$

and it is reasonable to assume here that $x_j^T A' x_j = \epsilon\lambda_n$ for any x_j so that

$$\frac{|\lambda_j' - \lambda_j|}{\lambda_j} = \epsilon\left(\frac{\lambda_n}{\lambda_1}\right)\frac{\lambda_1}{\lambda_j}. \tag{6.362}$$

This is the ultimate accuracy barrier of the round-off errors, and it can be serious in finite difference matrices that have large λ_n/λ_1 ratios.

The finite difference matrices

$$A = \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix} \text{ and } A^2 = \begin{bmatrix} 5 & -4 & 1 & & \\ -4 & 6 & -4 & 1 & \\ 1 & -4 & 6 & -4 & 1 \\ & 1 & -4 & 6 & -4 \\ & & 1 & -4 & 5 \end{bmatrix} \tag{6.363}$$

are with $\lambda_n/\lambda_1 = n^2$ and $\lambda_n/\lambda_1 = n^4$, respectively. Figures 6.7 and 6.8 show the round-off error in the eigenvalues of A and A^2, respectively, computed from Rayleigh's quotient $\lambda_j' = x_j^T Ax_j/x_j^T x_j$ with analytically correct eigenvectors and with machine accuracy $\epsilon = 10^{-7}$. The most serious relative round-off error is in the lower eigenvalues, and is nearly proportional to n^2 for A and nearly proportional to n^4 for A^2.

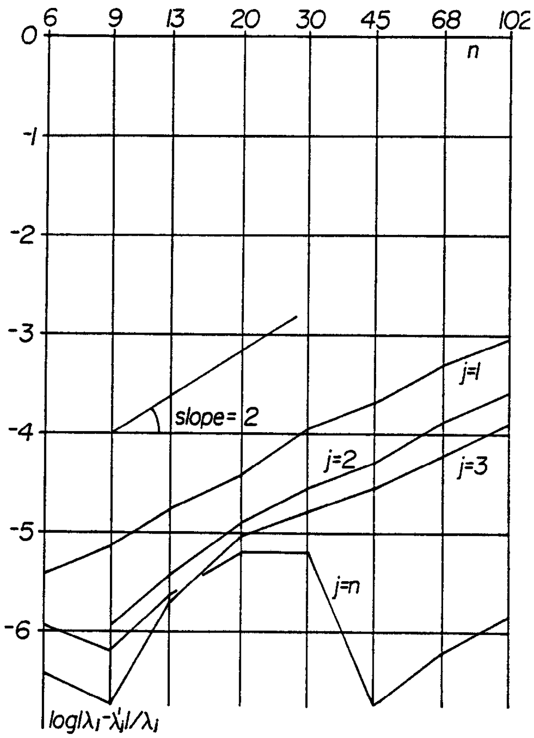

Fig. 6.7

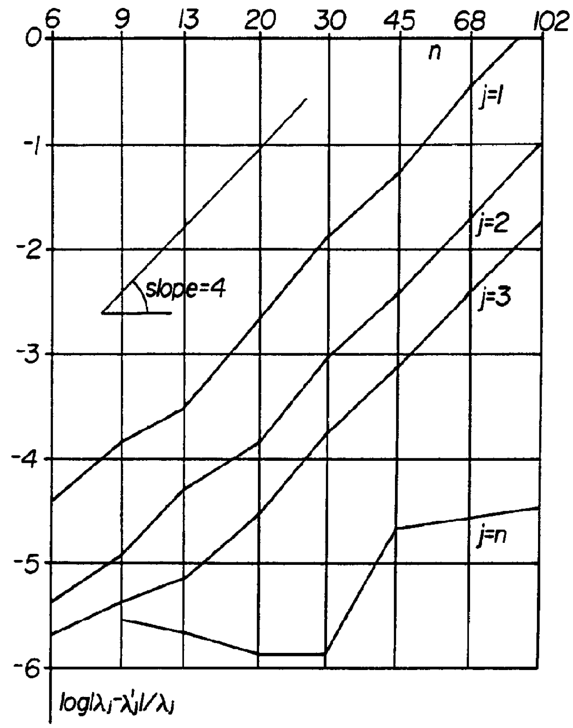

Fig. 6.8

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