8 Eigenvalue Problems

8.1 Motivation and Definition

Matrices can be used to represent linear transformations. Their effects can be: rotation, reflection, translation, scaling, permutation, etc., and combinations thereof. These transformations can be rather complicated, and therefore we often want to decompose a transformation into a few simple actions that we can better understand. Finding singular values and associated singular vectors is one such approach. In engineering, one often speaks of principal component analysis.

A more basic approach is to consider eigenvalues and eigenvectors.

Definition 8.1 *Let $A \in \mathbb{C}^{m\times m}$. If for some pair $(\lambda, \boldsymbol{x})$, $\lambda \in \mathbb{C}$, $\boldsymbol{x}(\neq \mathbf{0}) \in \mathbb{C}^m$ we have*

$$A\boldsymbol{x} = \lambda\boldsymbol{x},$$

then λ is called an eigenvalue *and $\boldsymbol{x}$ the associated* eigenvector *of A.*

Remark Eigenvectors specify the directions in which the matrix action is simple: any vector parallel to an eigenvector is changed only in length and/or orientation by the matrix A.

In practical applications, eigenvalues and eigenvectors are used to find *modes of vibrations* (e.g., in acoustics or mechanics), i.e., instabilities of structures can be investigated via an eigenanalysis.

In theoretical applications, eigenvalues often play an important role in the analysis of convergence of iterative algorithms (for solving linear systems), long-term behavior of dynamical systems, or stability of numerical solvers for differential equations.

8.2 Other Basic Facts

Some other terminology that will be used includes the *eigenspace* E_λ, i.e., the vector space of all eigenvectors corresponding to λ:

$$E_\lambda = \operatorname{span}\{\boldsymbol{x} : A\boldsymbol{x} = \lambda\boldsymbol{x}, \lambda \in \mathbb{C}\}.$$

Note that this vector space includes the zero vector — even though $\mathbf{0}$ is not an eigenvector.

The set of all eigenvalues of A is known as the *spectrum* of A, denoted by $\Lambda(A)$. The *spectral radius* of A is defined as

$$\rho(A) = \max\{|\lambda| : \lambda \in \Lambda(A)\}.$$

8.2.1 The Characteristic Polynomial

The definition of eigenpairs $A\boldsymbol{x} = \lambda\boldsymbol{x}$ is equivalent to

$$(A - \lambda I)\,\boldsymbol{x} = \mathbf{0}.$$

Thus, λ is an eigenvalue of A if and only if the linear system $(A - \lambda I)\boldsymbol{x} = \mathbf{0}$ has a *nontrivial* (i.e., $\boldsymbol{x} \neq \mathbf{0}$) solution.

This, in turn, is equivalent to $\det(A-\lambda I) = 0$. Therefore we define the *characteristic polynomial* of A as

$$p_A(z) = \det(zI - A).$$

Then we get

Theorem 8.2 *λ is an eigenvalue of A if and only if $p_A(\lambda) = 0$.*

Proof See above. ■

Remark This definition of p_A ensures that the coefficient of z^m is $+1$, i.e., p_A is a *monic* polynomial.

Example It is well known that even real matrices can have complex eigenvalues. For instance,

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

has a characteristic polynomial

$$p_A(z) = \begin{vmatrix} z & -1 \\ 1 & z \end{vmatrix} = z^2 + 1,$$

so that its eigenvalues are $\lambda_{1,2} = \pm i$ with associated eigenvectors

$$\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ i \end{bmatrix} \quad \text{and} \quad \boldsymbol{x}_2 = \begin{bmatrix} i \\ 1 \end{bmatrix}.$$

However, if A is symmetric (or Hermitian), then all its eigenvalues are real. Moreover, the eigenvectors to distinct eigenvalues are linearly independent, and eigenvectors to distinct eigenvalues of a symmetric/Hermitian matrix are orthogonal.

Remark Since the eigenvalues of an $m \times m$ matrix are given by the roots of a degree-m polynomial, it is clear that for problems with $m > 4$ we *will have to use iterative (i.e., numerical) methods* to find the eigenvalues.

8.2.2 Geometric and Algebraic Multiplicities

The number of linearly independent eigenvectors associated with a given eigenvalue λ, i.e., the dimension of E_λ is called the *geometric multiplicity* of λ.

The power of the factor $(z - \lambda)$ in the characteristic polynomial p_A is called the *algebraic multiplicity* of λ.

Theorem 8.3 *Any $A \in \mathbb{C}^{m\times m}$ has m eigenvalues provided we count the algebraic multiplicities. In particular, if the roots of p_A are simple, the A has m distinct eigenvalues.*

Example Take

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

so that $p_A(z) = (z-1)^3$. Thus, $\lambda = 1$ is an eigenvalue (in fact, the only one) of A with algebraic multiplicity 3. To determine its geometric multiplicity we need to find the associated eigenvectors.

To this end we solve $(A - \lambda I)\boldsymbol{x} = \mathbf{0}$ for the special case $\lambda = 1$. This yields the augmented matrix

$$\left[\begin{array}{ccc|c} 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

so that $x_1 = -x_3$ or

$$\boldsymbol{x} = \begin{bmatrix} \alpha \\ \beta \\ -\alpha \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + \beta \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix},$$

and therefore the geometric multiplicity of $\lambda = 1$ is only 2.

In general one can prove the following

Theorem 8.4 *The algebraic multiplicity of λ is always greater than or equal its geometric multiplicity.*

This prompts

Definition 8.5 *If the geometric multiplicity of λ is less than its algebraic multiplicity, then λ is called* defective.

Example As a continuation of the previous example we see that the matrix

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

has the same characteristic polynomial as before, i.e., $p_B(z) = (z-1)^3$, and $\lambda = 1$ is again an eigenvalue with algebraic multiplicity 3. To determine its geometric multiplicity we solve $(B - I)\boldsymbol{x} = \mathbf{0}$, i.e., look at

$$\left[\begin{array}{ccc|c} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Now there is no restriction on the components of $\boldsymbol{x}$ and we have

$$\boldsymbol{x} = \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \gamma \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

so that the geometric multiplicity of $\lambda = 1$ is 3 in this case.

At the other extreme, the matrix

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

also has the characteristic polynomial $p_C(z) = (z-1)^3$, so that $\lambda = 1$ has algebraic multiplicity 3. However, now the solution of $(C - I)\boldsymbol{x} = \mathbf{0}$, leads to

$$\left[\begin{array}{ccc|c} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right],$$

so that $x_2 = x_3 = 0$ and we get

$$\boldsymbol{x} = \alpha \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

This means that the geometric multiplicity of $\lambda = 1$ now is 1.

8.3 Determinant and Trace

The *trace* of a matrix A, $\mathrm{tr}(A)$, is given by the sum of its diagonal elements, i.e.,

$$\mathrm{tr}(A) = \sum_{j=1}^{m} a_j j.$$

Theorem 8.6 *If $A \in \mathbb{C}^{m\times m}$ with eigenvalues $\lambda_1, \ldots, \lambda_m$, then*

1. $\det(A) = \prod_{j=1}^{m} \lambda_j$,
2. $\mathrm{tr}(A) = \sum_{j=1}^{m} \lambda_j$.

Proof Recall that the definition of the characteristic polynomial, $p_A(z) = \det(zI - A)$, so that

$$p_A(0) = \det(-A) = (-1)^m \det(A).$$

On the other hand, we also know that we also have $p_A(z) = \prod_{j=1}^{m}(z - \lambda_j)$ which implies

$$p_A(0) = \prod_{j=1}^{m}(-\lambda_j) = (-1)^m \prod_{j=1}^{m} \lambda_j.$$

Comparing the two representations of $p_A(0)$ yields the first formula.

For the second one one can show that the coefficient of z^{m-1} in the representation $\det(zI - A)$ of the characteristic polynomial is $-\mathrm{tr}(A)$. On the other hand, the coefficient of z^{m-1} in the representation $\prod_{j=1}^{m}(z - \lambda_j)$ is $-\sum_{j=1}^{m} \lambda_j$. Together we get the desired formula. ■

8.4 Similarity and Diagonalization

Consider two matrices $A, B \in \mathbb{C}^{m\times m}$. A and B are called *similar* if

$$B = X^{-1}AX$$

for some nonsingular $X \in \mathbb{C}^{m\times m}$.

Theorem 8.7 *Similar matrices have the same characteristic polynomial, eigenvalues, algebraic and geometric multiplicities. The eigenvectors, however, are in general different.*

Theorem 8.8 *A matrix $A \in \mathbb{C}^{m\times m}$ is nondefective, i.e., has no defective eigenvalues, if and only if A is similar to a diagonal matrix, i.e.,*

$$A = X\Lambda X^{-1},$$

where $X = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m]$ is the matrix formed with the eigenvectors of A as its columns, and $\Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m)$ contains the eigenvalues.

Remark Due to this theorem, nondefective matrices are *diagonalizable*. Also note that nondefective matrices have linearly independent eigenvectors.

Remark The factorization

$$A = X\Lambda X^{-1}$$

is called the eigenvalue (or eigen-) decomposition of A. We can interpret this decomposition as a change of basis by which a coupled linear system is transformed to a decoupled diagonal system. This means

$$\begin{aligned} A\boldsymbol{x} = \boldsymbol{b} \quad &\Longleftrightarrow \quad X\Lambda X^{-1}\boldsymbol{x} = \boldsymbol{b} \\ &\Longleftrightarrow \quad \Lambda \underbrace{X^{-1}\boldsymbol{x}}_{=\hat{\boldsymbol{x}}} = \underbrace{X^{-1}\boldsymbol{b}}_{=\hat{\boldsymbol{b}}}. \end{aligned}$$

This shows that $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{b}}$ correspond to $\boldsymbol{x}$ and $\boldsymbol{b}$ as viewed in the basis of eigenvectors (i.e., columns of X).

If the eigenvectors of A are not only linearly independent, but also orthogonal, then we can factor A as

$$A = Q\Lambda Q^*$$

with a unitary matrix Q of eigenvectors. Thus, A is called *unitarily diagonalizable.*

Theorem 8.9 *If A is Hermitian, then A is unitarily diagonalizable. Moreover, Λ is real.*

More generally, A is called *normal* if

$$AA^* = A^*A,$$

and we have

Theorem 8.10 *A is unitarily diagonalizable if and only if A is normal.*

8.5 Schur Factorization

The most useful linear algebra fact summarized here for numerical analysis purposes is

Theorem 8.11 *Every square matrix $A \in \mathbb{C}^{m\times m}$ has a Schur factorization*

$$A = QTQ^*,$$

with unitary matrix Q and upper triangular matrix T such that $\mathrm{diag}(T)$ *contains the eigenvalues of A.*

Remark Note the similarity of this result to the singular value decomposition. The Schur factorization is quite general in that it exists for *every*, albeit only square, matrix. Also, the matrix T contains the eigenvalues (instead of the singular values), and it is upper triangular (i.e., not quite as nice as diagonal). On the other hand, only one unitary matrix is used.

Remark By using nonunitary matrices for the similarity transform one can obtain the Jordan normal form of a matrix in which T is bidiagonal.

Remark Both the Schur factorization and the Jordan form are considered not appropriate for numerical/practical computations because of the possibility of complex terms occurring in the matrix factors (even for real A). The SVD is preferred since all of its factors are real.

Proof We use induction on m. For $m = 1$ we have

$$A = (a_{11}), \quad Q = (1), \quad T = (a_{11}),$$

and the claim is clearly true.

For $m \geq 2$ we assume $\boldsymbol{x}$ is a normalized eigenvector of A, i.e., $\|\boldsymbol{x}\|_2 = 1$. Then we form

$$U = \left[\begin{array}{cc} \boldsymbol{x} & \hat{U} \end{array}\right] \in \mathbb{C}^{m\times m}$$

to be unitary by augmenting the first column, $\boldsymbol{x}$, by appropriate columns in $\hat{U}$. This gives us

$$U^*AU = \left[\begin{array}{c} \boldsymbol{x}^* \\ \hat{U}^* \end{array}\right] A \left[\begin{array}{cc} \boldsymbol{x} & \hat{U} \end{array}\right] = \left[\begin{array}{cc} \boldsymbol{x}^*A\boldsymbol{x} & \boldsymbol{x}^*A\hat{U} \\ \hat{U}A\boldsymbol{x} & \hat{U}^*A\hat{U} \end{array}\right].$$

Since $\boldsymbol{x}$ is an eigenvector of A we have $A\boldsymbol{x} = \lambda\boldsymbol{x}$, and after multiplication by $\boldsymbol{x}^*$

$$\boldsymbol{x}^*A\boldsymbol{x} = \lambda \underbrace{\boldsymbol{x}^*\boldsymbol{x}}_{=\|\boldsymbol{x}\|_2^2=1} = \lambda.$$

Similarly,

$$\hat{U}^*A\boldsymbol{x} = \hat{U}^*(\lambda\boldsymbol{x}) = \lambda\hat{U}^*\boldsymbol{x} = \boldsymbol{0}$$

since $\hat{U}^*\boldsymbol{x} = \boldsymbol{0}$ because U is unitary.

Therefore, U^*AU simplifies to

$$U^*AU = \left[\begin{array}{cc} \lambda & \boldsymbol{b}^* \\ \boldsymbol{0} & C \end{array}\right],$$

where we have used the abbreviations $\boldsymbol{b}^* = \boldsymbol{x}^* A\hat{U}$ and $C = \hat{U}^* A \hat{U} \in \mathbb{C}^{(m-1)\times(m-1)}$.

Now, by the induction hypothesis, C has a Schur factorization

$$C = V\hat{T}V^*$$

with unitary V and triangular $\hat{T}$.

To finish the proof we can define

$$Q = U \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & V \end{bmatrix}$$

and observe that

$$\begin{aligned} Q^*AQ &= \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & V^* \end{bmatrix} \underbrace{U^*AU}_{=\begin{bmatrix} \lambda & \boldsymbol{b}^* \\ \mathbf{0} & C \end{bmatrix}} \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & V \end{bmatrix} \\ &= \begin{bmatrix} \lambda & \boldsymbol{b}^*V \\ \mathbf{0} & V^*CV \end{bmatrix}. \end{aligned}$$

This last block matrix, however, is the desired upper triangular matrix T with the eigenvalues of A on its diagonal since $V^*CV = \hat{T}$ and the induction hypothesis ensures $\hat{T}$ already has the desired properties. ■

To summarize this section we can say

1. $A \in \mathbb{C}^{m\times m}$ is *diagonalizable*, i.e., $A = X\Lambda X^{-1}$, if and only if A is *nondefective.*
2. $A \in \mathbb{C}^{m\times m}$ is *unitarily diagonalizable*, i.e., $A = Q\Lambda Q^*$, if and only if A is *normal.*
3. $A \in \mathbb{C}^{m\times m}$ is *unitarily triangularizable*, i.e., $A = QTQ^*$ *for any square* A.