Matrices and Transformations

Anthony J. Pettofrezzo[1]

No Institute Given

1 Eigenvalues and Eigenvectors

1.1 Characteristic Functions

Associated with each square matrix $A = ((a_{ij}))$ of order n is a function

$$f(\lambda) = |A - \lambda I| = \begin{vmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - \lambda & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} - \lambda \end{vmatrix} \tag{1}$$

called the **characteristic function** of A. The equations

$$f(\lambda) = |A - \lambda I| = 0 \tag{2}$$

can be expressed in the polynomial form

$$c_0\lambda^n + c_1\lambda^{n-1} + \dots + c_{n-1}\lambda + c_n = 0 \tag{3}$$

and is called the **characteristic equation** of matrix A.

Example 1. Find the characteristic equation of matrix A where

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 2 & 2 \\ 0 & 2 & 3 \end{pmatrix}$$

The characteristic equation of A is

$$\begin{vmatrix} 1 - \lambda & 2 & 0 \\ 2 & 2 - \lambda & 2 \\ 0 & 2 & 3 - \lambda \end{vmatrix} = 0$$

that is,

$$(1 - \lambda) \begin{vmatrix} 1 - \lambda & 2 \\ 2 & 2 - \lambda \end{vmatrix} - 2 \begin{vmatrix} 2 & 2 \\ 0 & 3 - \lambda \end{vmatrix} = 0$$

$$(1 - \lambda)(\lambda^2 - 5\lambda + 2) - 2(6 - 2\lambda) = 0$$

$$\lambda^3 - 6\lambda^2 + 3\lambda + 10 = 0$$

In some instances the task of expressing the characteristic equation of a matrix in polynomial form may be simplified considerably by introducing the concept of the trace of a matrix. The sum of the diagonal elements of a matrix A is denoted by $tr(A)$ For example, the trace of matrix A in Example 1 is 1+2+3; that is, 6. Let $t_1 = tr(A)$, $t_2 = tr(A^2), \dots, t_n = tr(A^n)$. It can be shown that the coefficients of the characteristic equation are given by the equations:

$$\begin{cases} c_0 = & 1, \\ c_0 = & -t_1, \\ c_0 = & -\frac{1}{2}(c_1 t_1 + t_2), \\ c_0 = & -\frac{1}{3}(c_2 t_1 + c_1 t_2 + t_3), \\ \cdots & \cdots \\ c_0 = & -\frac{1}{n}(c_{n-1} t_1 + c_{n-2} t_2 + \cdots + c_1 t_{n-1} + t_n). \end{cases} \tag{4}$$

Equation 4 make it possible to calculate the coefficients of the characteristic equation of a matrix A by assuming the diagonal elements of the matrices of the form A^n. This numerical process is easily programmed on a large-scale digital computer, or for small values of n may be computed manually without difficulty.

The n roots $\lambda_1, \lambda_2, \dots, \lambda_n$ of the characteristic equation 3 of a matrix A are called the **eigenvalues** of A.

The trace of a matrix A of order n is equal to the sum of the n eigenvalues of A.

Many applications of matrix algebra in mathematics, physics, and engineering involve the concept of a set of nonzero vectors being mapped onto the zero vector by means of the matrix $A - \lambda_i I$, where λ_i, is an eigenvalue of matrix A. Any nonzero column vector, denoted by X_i, such that

$$(A - \lambda_i I)X_i = 0 \tag{5}$$

is called **eigenvector** of matrix A. It is guaranteed that at least one eigenvector exists for each λ_i since equation 5 represents a system of n linear homogeneous equations which has a nontrivial solution $X_i \neq 0$ if and only if $|A - \lambda_i I| = 0$; that is, if and only if λ_i is an eigenvalue of A. Furthermore, note that any nonzero scalar multiple of an eigenvector associated with an eigenvalue is also an eigenvector associated with that eigenvalue.

The eigenvalues of a matrix are also called the **proper values**, the **latent values**, and the **characteristic values** of the matrix. The eigenvectors of a matrix are also called the **proper vectors**, the **latent vectors**, and the **characteristic vectors** of the matrix.

Example 2. Determine a set of eigenvectors of the matrix $A = \begin{pmatrix} 3 & 1 \\ 2 & 2 \end{pmatrix}$.

Associated with $\lambda_1 = 1$ are the eigenvectors $\begin{pmatrix} x_1 & x_2 \end{pmatrix}^T$ for which

$$(A - I)\begin{pmatrix} x_1 & x_2 \end{pmatrix}^T = 0$$

that is,

$$\begin{pmatrix} 2 & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

It follows that $-2x_1 = x_2$. If x_1 is chosen as some convenient arbitrary scalar, say 1, x_2 becomes -2. Hence, $\begin{pmatrix} 1 & -2 \end{pmatrix}^T$ is an eigenvector associated with the eigenvalue 1.

Similarly, associated with $\lambda_2 = 4$ are the eigenvectors $\begin{pmatrix} x_1 & x_2 \end{pmatrix}^T$ for which

$$(A - 4I) \begin{pmatrix} x_1 & x_2 \end{pmatrix}^T = 0$$

that is,

$$\begin{pmatrix} -1 & 1 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Hence, $x_1 = x_2$, and $\begin{pmatrix} 1 & 1 \end{pmatrix}^T$ is an eigenvector associated with the eigenvalue 4. Therefore, one set of eigenvectors of the matrix A is $\{\begin{pmatrix} 1 & -2 \end{pmatrix}^T, \begin{pmatrix} 1 & 1 \end{pmatrix}^T\}$.

It should be noted that $\begin{pmatrix} k & -2k \end{pmatrix}^T$ and $\begin{pmatrix} k & k \end{pmatrix}^T$, where k is any nonzero scalar, represent the general forms of the eigenvectors of A.

1.2 A Geometric Interpretation of Eigenvectors

Consider a magnification of the plane represented by the matrix A where

$$A = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}$$

The eigenvalues of A are $\lambda_1 = 3$ and $\lambda_2 = 2$. Every eigenvector associated with λ_1 is of the form $\begin{pmatrix} k & 0 \end{pmatrix}^T$, where k is any nonzero scalar, since

$$\begin{pmatrix} 3-3 & 0 \\ 0 & 2-3 \end{pmatrix} \begin{pmatrix} k \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Furthermore, the set of vectors of the form $\begin{pmatrix} k & 0 \end{pmatrix}^T$ is such that

$$A \begin{pmatrix} k & 0 \end{pmatrix}^T = \lambda_1 \begin{pmatrix} k & 0 \end{pmatrix}^T$$

that is

$$\begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} k \\ 0 \end{pmatrix} = 3 \begin{pmatrix} k \\ 0 \end{pmatrix}$$

Hence, the set of eigenvectors associated with $\lambda_1 = 3$ is mapped onto itself under the transformation represented by A, and the image of each eigenvector is a fixed scalar multiple of hat eigenvector. The fixed scalar multiple is equal to the eigenvalue with which the set of eigenvectors is associated.

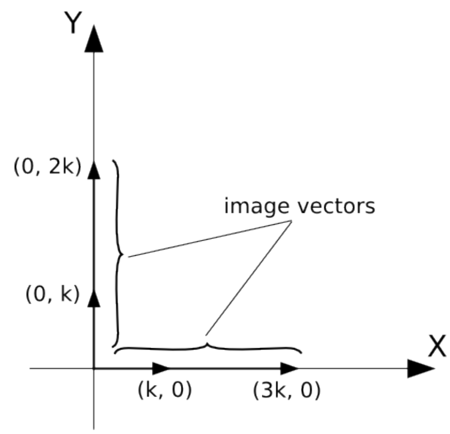

Fig. 1.

Similarly, every eigenvector associated with λ_2 is of the form $\left(0\ k\right)^T$, where k is any nonzero scalar. The set of vectors of the form $\left(0\ k\right)^T$ is such that

$$A\left(0\ k\right)^T = \lambda_1 \left(0\ k\right)^T$$

that is,

$$\begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ k \end{pmatrix} = 2 \begin{pmatrix} 0 \\ k \end{pmatrix}$$

Hence the set of eigenvectors associated with $\lambda_2 = 2$ is mapped onto itself under the transformation represented by A, and the image of each eigenvector is a fixed scalar multiple of the eigenvalue. The fixed scalar multiple is λ_2; that is, 2.

Note that the sets of vectors of the forms $\left(k\ 0\right)^T$ and $\left(0\ k\right)^T$ lie along the x-axis and y-axis, respectively 1.2. Under the magnification of the plane represented by the matrix

$$A = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}$$

the *one-dimensional vector spaces* containing the sets of vectors of the forms $\left(k\ 0\right)^T$ and $\left(0\ k\right)^T$ are mapped *onto* themselves, respectively, and are called **invariant vector spaces**. The invariant vector spaces help characterize or describe a particular transformation of the plane.

1.3 Some Theorems

In this section several theorems concerning the eigenvalues and eigenvectors of matrices in general and of symmetric matrices in particular will be proved.

These theorems are important for an understanding of the remaining sections of this text.

Notice that in Example 2 the eigenvector associated with the distinct eigenvalues of matrix A are linearly independent; that is,

$$k_1 \begin{pmatrix} 1 \\ -2 \end{pmatrix} + k_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

implies $k_1 = k_2 = 0$. This is not a coincidence. The following theorem states a sufficient condition for eigenvector associated with the eigenvalues of a matrix to be linearly independent.

Theorem 1. *If the eigenvalues of a matrix are distinct, then the associated eigenvectors are linearly independent.*

Proof. Let A be a square matrix of order n with distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ and associated eigenvectors $X_1, X_2, \ldots, X_n$, respectively. Assume that the set of eigenvectors are linearly dependent. Then there exists scalars $k_1, k_2, \ldots, k_n$, not all zero, such that

$$k_1 X_1 + k_2 X_2 + \cdots + k_n X_n = 0 \tag{6}$$

Consider premultiplying both sides of 6 by

$$(A - \lambda_2 I)(A - \lambda_3 I) \cdots (A - k_n I)$$

By use of equation 5, obtain

$$k_1 (A - \lambda_2 I)(A - \lambda_3 I) \cdots (A - k_n I) X_1 = 0 \tag{7}$$

Since $(A - \lambda_i I) X_i = 0$, then $A X_1 = \lambda_1 X_1$. Hence, equation 7 may be written as

$$k_1 (\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3) \cdots (\lambda_1 - \lambda_n) X_1 = 0$$

which implies $k_1 = 0$. Similarly, it can be shown that $k_1 = k_2 = \cdots = k_n = 0$, which is contrary to the hypothesis. Therefore, the set of eigenvectors are linearly independent.

It should be noted that if the eigenvalues of a matrix are not distinct, the associated eigenvectors may or may not be linearly independent. For example, consider the matrices

$$A = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix} \quad \textit{and} \quad B = \begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix}$$

Both matrices have $\lambda_1 = \lambda_2 = 3$; that is, an eigenvalue of multiplicity two. Any nonzero vector of the form $\begin{pmatrix} x_1 & x_2 \end{pmatrix}^T$ is an eigenvector of A for λ_1 and λ_2. Hence, it is possible to choose any two linearly independent vectors such as $\begin{pmatrix} 1 & 0 \end{pmatrix}^T$ and $\begin{pmatrix} 0 & 1 \end{pmatrix}^T$ as eigenvectors of A that are associated with λ_1 and λ_2, respectively. Only a vector of the form $\begin{pmatrix} x_1 & 0 \end{pmatrix}^T$ is, however, an eigenvector of B for λ_1 and λ_2. Any two vectors of this form are linearly dependent; that is, one is a linear function of the other.

Hermitian Matrices A **complex matrix** is a matrix whose elements are complex numbers. Since every real number is a complex number, every real matrix is a complex matrix, but not every complex matrix is a real matrix.

If A is a complex matrix, then $\bar{A}$ denotes the matrix obtained from A by replacing each element $z = a + bi$ with its conjugate $\bar{z} = a - bi$. The matrix $\bar{A}$ is called the **conjugate** of matrix A.

Note that matrix A is a real matrix if and only if $A = \bar{A}$. The transpose of the conjugate of a matrix A will be denoted by A^*; that is, $A^* = \left(\bar{A}\right)^T$. If $A = ((a_{ij}))$, then $A^T = ((a_{ji}))$, $\bar{A} = ((\bar{a_{ij}}))$ and $\left(\bar{A}\right)^T = ((\bar{a_{ji}})) = \left(\overline{A^T}\right)$. Hence, the transpose of the conjugate of a matrix is equal to the conjugate of the transpose of the matrix.

A matrix A such that $A = A^*$ is called a **Hermitian matrix**; that is, a matrix $A = ((a_{ij}))$ is a Hermitian matrix if and only if $a_{ij} = \overline{a_{ij}}$ for all pairs (i, j). Since $a_{ii} = \overline{a_{ii}}$ only if a_{ii} is a real number, the diagonal elements of a Hermitian matrix are real numbers. If A is a real symmetric matrix, then $a_{ij} = a_{ji}$, and $a_{ij} = \overline{a_{ji}}$ for all pairs (i, j). Every real skew-symmetric matrix is a skew-Hermitian matrix.

A matrix A such that $A = -A^*$ is called a **skew-Hermitian matrix**; that is, a matrix $A = ((a_{ij}))$ is a skew-Hermitian matrix if and only if $a_{ij} = -\overline{a_{ij}}$ for all pairs (i, j). Every real skew-symmetric matrix is a skew-Hermitian matrix.

Theorem 2. *If A is a Hermitian matrix, then the eigenvalues of A are real.*

Proof. Let A be a Hermitian matrix, λ_i be any eigenvalue of A, and X_i be an eigenvector associated with λ_i. Then

$$\begin{aligned}(A - \lambda_i I)X_i &= 0\\ AX_i - \lambda_i X_i &= 0\\ X_i^* A X_i - \lambda_i X_i^* X_i &= 0\end{aligned}$$

Since every eigenvector is a nonzero vector, $X_i^* X_i$ is a nonzero real number and

$$\lambda_i = \frac{X_i^* A X_i}{X_i^* X_i}$$

Furthermore,
$X_i^* A X_i = X_i^* A^* X_i$ since $A^* = A$
$X_i^* A X_i = (X_i^* A X_i)^*$
$X_i^* A X_i = \overline{X_i^* A X_i}$ since $X_i^* A X_i$ is a matrix of one element;
that is, $X_i^* A X_i$ equals its own conjugate, and hence is real. Therefore, λ_i is equal to the quotient of two real numbers, and is real.

Theorem 3. *If A is a real symmetric matrix, then the eigenvalues of A are real.*

Proof. Since every real symmetric matrix is a Hermitian matrix, the proof follows from Theorem 2.

Before presenting the next theorem it is necessary to consider the following definition: two complex eigenvectors X_1 and X_2 are defined as orthogonal if $X_1^* X_2 = 0$. For example, if $X_1 = \begin{pmatrix} -i & 2 \end{pmatrix}^T$ and $X_2 = \begin{pmatrix} 2i & 1 \end{pmatrix}^T$, then $X_1^* X_2 = \begin{pmatrix} i & 2 \end{pmatrix} \begin{pmatrix} 2i & 1 \end{pmatrix}^T = 0$. Hence, X_1 and X_2 are orthogonal.

Theorem 4. *If A is a Hermitian matrix, then the eigenvectors of A associated with distinct eigenvalues are mutually orthogonal vectors.*

Proof. Let A be a Hermitian matrix, and let X_1 and X_2 be eigenvectors associated with any two distinct eigenvalues λ_1 and λ_2, respectively. Then

$$(A - \lambda_1 I)X_1 = 0 \quad \textit{and} \quad (A - \lambda_2 I)X_2 = 0$$

...

Theorem 5. *If A is a real symmetric matrix, then the eigenvectors of A associated with distinct eigenvalues are mutually orthogonal vectors.*

Inverse of a Matrix This section will be concerned with the problem of finding a multiplicative inverse, if it exists, for any given square matrix. A **left multiplicative inverse** of a matrix A is a matrix B such that $BA = I$; a **right multiplicative inverse** of a matrix A is a matrix C such that $AC = I$. If a left and a right multiplicative inverse of a matrix A are equal, the the left (right) inverse is called, simply, a **multiplicative inverse** of A and is denoted by A^{-1}.

Theorem 6. *A left multiplicative inverse of a square matrix A is a multiplicative inverse of A.*

Proof. Suppose $BA = I$, then

....

Theorem 7. *A right multiplicative inverse of a square matrix A is a multiplicative inverse of A.*

Theorem 8. *The multiplicative inverse, if it exists, of a square matrix A is unique*

Proof. Let A^{-1} and B be any of two multiplicative inverses of the square matrix A. Since $A^{-1}A = I$ and $BA = I$, then

....

Not every square matrix has a multiplicative inverse. In fact, the necessary and sufficient condition for the multiplicative inverse of a matrix A to exist is that $\det A \neq 0$. A square matrix A is said to be **nonsingular** if $\det A \neq 0$, and **singular** if $\det A = 0$.

It should be mentioned that if A is not a square matrix, then it is possible for A to have a left or a right multiplicative inverse, but not both.

1.4 Diagonalization of Matrices

It has been noted that an eigenvector X_i such that $(A - \lambda_i I)X_i = 0$, for $i = 1, 2, \ldots, n$, may be associated with each eigenvalue λ_i. This relationship may be expressed in the alternate form

$$AX_i = \lambda_i X_i \quad for \quad i = 1, 2, \ldots, n \tag{8}$$

If a square matrix of order n whose columns are eigenvectors X_i of A is constructed and denoted by X, then the equations of 8 may be expressed in the form

$$AX = X\Lambda \tag{9}$$

where Λ is a diagonal matrix whose diagonal elements are the eigenvalues of A; that is

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \tag{10}$$

It has been proved that the eigenvectors associated with distinct eigenvalues are linearly independent (Theorem 1. Hence, the matrix X will be nonsingular if the λ_i's are distinct. If both sides of equation 9 are multiplied by X_i^{-1}, the result is

$$X^{-1}AX = \Lambda \tag{11}$$

Thus, by use of a matrix of eigenvectors and its inverse, it is possible to transform any matrix A with distinct eigenvalues to a diagonal matrix whose diagonal elements are the eigenvalues of A. The transformation expressed by 11 is referred to as the diagonalization of matrix A. If the eigenvalues are not distinct, the diagonalization of matrix A may not be possible. For example, the matrix

$$A = \begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix}$$

cannot be diagonalized as in 11.

A matrix such as matrix A in equation 11 sometimes is spoken of as being **similar** to the diagonal matrix. In general, if there exists a nonsingular matrix C such that $C^{-1}AC = B$ for any two square matrices A and B of the same order, the A and B are called **similar matrices**, and the transformation of A to B is called a **similarity transformation**. Furthermore, if B is a diagonal matrix whose diagonal elements are the eigenvalues of A, the B is called the **classical canonical form** of matrix A. It is a unique matrix except for the order in which the eigenvalues appear along the principal diagonal.

The matrix X of 11 whose columns are eigenvectors of matrix A often is called a **modal matrix** of A. Recall that each eigenvector may be multiplied by any nonzer scalar.

Theorem 9. *Every real symmetric matrix can be orthogonally transformed to the classical canonical form.*

Theorem 9 is sometimes called the **Principal Axes Theorem**.

1.5 The Hamilton-Cayley Theorem

An important and interesting theorem of the theory of matrices is the **Hamilton-Cayley Theorem**:

Theorem 10. *Every square matrix A satisfies its own characteristic equation $|A - \lambda I| = 0$.*

More precisely, if λ is replaced by the matrix A of order n and each real number c_n is replaced by the scalar multiple $c_n I$ where I is the identity matrix of order n, then the characteristic equation of matrix A becomes a valid matrix equation; that is,

$$c_0 A^n + c_1 A^{n-1} + \cdots + c_{n-1} A + c_n I = 0 \tag{12}$$

A heuristic argument may be used to prove the Hamilton-Cayley Theorem for a matrix A with distinct eigenvalues. Replace the variable λ by the square matrix A and c_n by $c_n I$ in expression for the characteristic function of A and obtain

$$f(A) = c_0 A^n + c_1 A^{n-1} + \cdots + c_{n-1} A + c_n I \tag{13}$$

Postmultiply both sides of equation 15 by an eigenvector X_i of A associated with λ_i and obtain

$$f(A) X_i = (c_0 \lambda_i^n + c_1 \lambda_i^{n-1} + \cdots + c_{n-1} \lambda_i + c_n) X_i \tag{14}$$

since $A^k X_i = \lambda_i^k X_i$. Since

$$c_0 \lambda_i^n + c_1 \lambda_i^{n-1} + \cdots + c_{n-1} \lambda_i + c_n = 0 \quad for \quad i = 1, 2, \ldots, n \tag{15}$$

then

$$f(A) X_i = 0 \quad for \quad i = 1, 2, \ldots, n$$

Hence,

$$f(A) X = 0 \tag{16}$$

where X is a matrix of eiegnvectors. Since the eigenvectors are linearly independent by Theorem 1, the matrix of eigenvectors has a unique inverse X^{-1}.

If both sides of equation 16 are postmultiplied by X^{-1}, the result is $f(A) = 0$, and the theorem is proved.

Proofs of the Hamilton-Cayley Theorem for the general case without restrictions on the eigenvalues of A may be found in most advanced texts on linear algebra.

The Hamilton-Cayley Theorem may be applied to the problem of determining the inverse of a nonsingular matrix A. Let

$$c_0\lambda^n + c_1\lambda^{n-1} + \cdots + c_{n-1}\lambda + c_n = 0$$

be the characteristic equation of A. Note that since A is a nonsingular matrix, $\lambda_i \neq 0$; that is, every eigenvalue is nonzero, and $c_n \neq 0$. By the Hamilton-Cayley Theorem,

$$c_0A^n + c_1A^{n-1} + \cdots + c_{n-1}A + c_nI = 0$$

and

$$I = -\frac{1}{c_n}(c_0A^n + c_1A^{n-1} + \cdots + c_{n-1}A) \qquad (17)$$

If both sides of 18 are multiplied by A^{-1}, the result is

$$A^{-1} = -\frac{1}{c_n}(c_0A^{n-1} + c_1A^{n-2} + \cdots + c_{n-1}I) \qquad (18)$$

Note that the calculation of an inverse by use of equation 18 is quite adaptable to high-speed digital computers and is not difficult to compute manually for small values of n. In calculating the powers of matrix A necessary in equation 18, the necessary information concerning $tr(A^k)$ for calculating the c's is also obtained.

A^{-k}, where k is positive integer, is defined to be equal to $\left(A^{-1}\right)^k$. By use of equation 18, it is now possible to express any negative integral power of a nonsingular matrix A of order n in terms of a linear function of the first $(n-1)$ powers of A.

References

1. Anthony J. Pettofrezzo. *Matrices and Transformations.* Dover Publications, June 1978.