13. Eigenvalues and Eigenvectors

Pre-multiplying a non-zero $(n \times 1)$ vector $\mathbf{x}$ by an $(n \times n)$ matrix $\mathbf{A}$ generally results in a very different vector (different length, different direction). However, in some cases, the new vector is just a multiple of the original vector (same or opposite direction, only length may have changed).

<u>Example 1</u> Let $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix}$. We have

$$\begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix} = 3\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$ same vector, only stretched by multiple of 3.

The vector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is called an eigenvector of $\mathbf{A}$, and 3 is called an eigenvalue of $\mathbf{A}$.

Another eigenvector of $\mathbf{A}$ is $\begin{bmatrix} -\frac{2}{3} \\ 1 \end{bmatrix}$, since

$$\begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix}\begin{bmatrix} -\frac{2}{3} \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{4}{3} \\ -2 \end{bmatrix} = -2\begin{bmatrix} -\frac{2}{3} \\ 1 \end{bmatrix}$$ opposite direction, $-2\times$ the original vector.

This eigenvector is associated with the eigenvalue -2.

<u>Example 2</u> Let $\mathbf{B} = \begin{bmatrix} 0.9 & 0.3 \\ 0.1 & 0.7 \end{bmatrix}$. We have

$$\begin{bmatrix} 0.9 & 0.3 \\ 0.1 & 0.7 \end{bmatrix}\begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -0.6 \\ 0.6 \end{bmatrix} = (0.6)\begin{bmatrix} -1 \\ 1 \end{bmatrix}$$ same direction, reduced by factor of 0.6.

$$\begin{bmatrix} 0.9 & 0.3 \\ 0.1 & 0.7 \end{bmatrix}\begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} = (1)\begin{bmatrix} 3 \\ 1 \end{bmatrix}$$ same vector is returned.

The vectors $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$ are eigenvectors of the matrix $\mathbf{B}$, and with corresponding eigenvalues 0.6 and 1 respectively.

Eigenvalues and eigenvectors are useful in many applications, including dynamic problems involving differential or difference equations. Eigenvalues (and eigenvectors) are also intimately connected to other matrix concepts such as the determinant, rank, and definiteness of a matrix.

Example 3 Certain matrices can be decomposed into the following form

$$\mathbf{B} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1}$$

where the columns of **S** are the eigenvectors of **B**, and $\boldsymbol{\Lambda}$ is a diagonal matrix with the eigenvalues making up the diagonal. For instance, you can verify that

$$\mathbf{B} = \begin{bmatrix} 0.9 & 0.3 \\ 0.1 & 0.7 \end{bmatrix} = \underbrace{\begin{bmatrix} 3 & -1 \\ 1 & 1 \end{bmatrix}}_{\mathbf{S}} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 0.6 \end{bmatrix}}_{\boldsymbol{\Lambda}} \underbrace{\begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ -\frac{1}{4} & \frac{3}{4} \end{bmatrix}}_{\mathbf{S}^{-1}}.$$

Suppose $\mathbf{x}_{t+1} = \mathbf{B}\mathbf{x}_t$ for all $t = 0,1,2,...$. For a given $\mathbf{x}_0$, what is $\mathbf{x}_{100}$? What happens as $t \to \infty$? Obviously $\mathbf{x}_{100} = \mathbf{B}^{100}\mathbf{x}_0$, and in general $\mathbf{x}_t = \mathbf{B}^t\mathbf{x}_0$. In order to calculate **B** to the power of 100, we can exploit the decomposition above to get

$$\mathbf{B}^t = \underbrace{\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1}\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1}\cdots\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1}}_{t \text{ instances of } \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1}} = \mathbf{P}\boldsymbol{\Lambda}^t\mathbf{P}^{-1}.$$

The power $\boldsymbol{\Lambda}^t$ is straightforward to compute, as it is a diagonal matrix. In this example, it is also obvious that as $t \to \infty$, we have

$$\boldsymbol{\Lambda}^t = \begin{bmatrix} 1^t & 0 \\ 0 & 0.6^t \end{bmatrix} \to \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \text{ as } t \to \infty,$$

so we can compute that

$$\mathbf{x}_t = \mathbf{B}^t\mathbf{x}_0 \to \begin{bmatrix} 3 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ -\frac{1}{4} & \frac{3}{4} \end{bmatrix} \mathbf{x}_0 = \begin{bmatrix} \frac{3}{4} & \frac{3}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix} \mathbf{x}_0 \text{ as } t \to \infty.$$

Example 4 In Examples 1 and 2, one may have noticed that

$$\det(\mathbf{A}) = -6 = (3)(-2) \text{ and } tr(\mathbf{A}) = 1 = 3 + (-2)$$

and

$$\det(\mathbf{B}) = 0.6 = (1)(0.6) \text{ and } tr(\mathbf{B}) = 1.6 = 1 + 0.6.$$

In general, the product of the eigenvalues of a matrix gives its determinant, and the sum of the eigenvalues gives its trace.

Definition For any $n \times n$ matrix $\mathbf{A}$, a non-zero vector **x** is said to be an eigenvector of **A** if

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \text{ for some scalar } \lambda$$

The scalar λ is said to be an eigenvalue of $\mathbf{A}$ associated with the eigenvector $\mathbf{x}$.

To find the eigenvalues and eigenvectors of a square matrix **A**, we note that

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

This system of equation has a non-trivial solution $\mathbf{x} \neq \mathbf{0}$ only if

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0$$

(this is called the characteristic equation of the matrix $\mathbf{A}$). Therefore, we can find the eigenvalues of $\mathbf{A}$ by finding all λ that satisfy $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$. Then, for each λ, we find the associated eigenvector $\mathbf{x}$ from the equation $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$.

Example 5 Find the eigenvalues of the matrix $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix}$.

The eigenvalues satisfy $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$. Since

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \det\left(\begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix} - \lambda\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) = \det\left(\begin{bmatrix} 1-\lambda & 2 \\ 3 & -\lambda \end{bmatrix}\right) = \lambda^2 - \lambda - 6 = (\lambda - 3)(\lambda + 2),$$

the eigenvalues are $\lambda = 3$ and $\lambda = -2$.

For $\lambda = 3$, the system $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$ is

$$\begin{bmatrix} 1-\lambda & 2 \\ 3 & -\lambda \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -2 & 2 \\ 3 & -3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

which gives $x_1 = x_2$, i.e. any vector of the form $\mathbf{x} = s\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is an eigenvector of $\mathbf{A}$ associated with $\lambda = 3$. Similarly, for $\lambda = -2$, the system $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$ is

$$\begin{bmatrix} 1-\lambda & 2 \\ 3 & -\lambda \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 3 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

which gives $x_1 = -\frac{2}{3}x_2$, i.e., any vector of the form $\mathbf{x} = s\begin{bmatrix} -\frac{2}{3} \\ 1 \end{bmatrix}$ is an eigenvector of $\mathbf{A}$ associated with $\lambda = -2$.

Note that if $\mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^{\mathrm{T}}$ is an eigenvector associated with an eigenvalue λ, then the vector $s\mathbf{x} = \begin{bmatrix} sx_1 & sx_2 \end{bmatrix}^{\mathrm{T}}$ for any s is also an eigenvector of $\mathbf{A}$ associated with λ, since

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \Leftrightarrow \mathbf{A}(s\mathbf{x}) = \lambda(s\mathbf{x}).$$

For example, for the matrix **A** in example 1, the vectors $\mathbf{x} = \begin{bmatrix} 1 & 1 \end{bmatrix}^{\mathrm{T}}$, $\begin{bmatrix} 2 & 2 \end{bmatrix}^{\mathrm{T}}$, $\begin{bmatrix} 3 & 3 \end{bmatrix}^{\mathrm{T}}$, and so on, are all eigenvectors associated with the eigenvalue $\lambda = 3$; there is an entire line of eigenvectors associated with this eigenvalue. In some applications, we restrict eigenvectors to unit length, so the eigenvector for **A** associated with the eigenvalue $\lambda = 3$ is $\mathbf{x} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}^{\mathrm{T}}$.

Exercises

1. Find the eigenvalues and associated eigenvectors of the matrices

(a) $\begin{bmatrix} 5 & -2 \\ 4 & -1 \end{bmatrix}$; *Ans: Eigenvalues are* $\lambda = 3$ *and* $\lambda = 1$ *with eigenvectors* $s\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ *and* $t\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ *resp.*

(b) $\begin{bmatrix} 1 & 1 \\ 4 & 1 \end{bmatrix}$ (c) $\begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$

(e) $\begin{bmatrix} 2 & 1 \\ -1 & 4 \end{bmatrix}$ (f) $\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 5 & 3 & 2 \end{bmatrix}$

Exercise 1(d) shows that eigenvalues may take value zero. There is nothing unusual about this, although it does say something important regarding the matrix (more on that later).

In general, an $(n \times n)$ matrix will have n eigenvalues but exercises 1(e) and (f) shows that these eigenvalues need not be distinct. The matrix in exercise 1(e) has eigenvalue $\lambda = 3$, occurring twice. In exercise 1(f), the eigenvalues are $\lambda = 1$, 1, and 2. We say that the eigenvalue $\lambda = 1$ occurs with multiplicity 2.

In the examples we have seen so far, there was a single line of eigenvectors associated with a given eigenvalue. In some cases, there are multiple lines of eigenvectors associated with a single eigenvalue:

2. Find the eigenvalues of the matrix

$$\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}.$$

Show that any vector $\mathbf{x} \neq \mathbf{0}$ is an eigenvector associated with the eigenvalues. Does this make any geometric sense?

Note that eigenvalues can take complex values.

3. Find the eigenvalues and associated eigenvectors of the matrix $\mathbf{A} = \begin{bmatrix} 2 & 4 \\ -2 & 6 \end{bmatrix}$.

You should find $\lambda = 4 \pm 2i$. *The eigenvectors can be expressed in many ways. One possibility is* $s\begin{bmatrix} 1-i \\ 1 \end{bmatrix}$ *and* $t\begin{bmatrix} 1 \\ \frac{1-i}{2} \end{bmatrix}$ *respectively.*

~~~

A detailed look at the (2×2) *case*

We study the (2×2) case in detail. A few of the results obtained here apply only to the (2×2) case, but many generalize to the $(n \times n)$ case (albeit with more complicated proofs). Studying the (2×2) case will help us develop intuition for these results with a minimum of algebra.

Let $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$. The characteristic polynomial is

$$\begin{aligned} \rho(\lambda) &= \det\begin{pmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{pmatrix} \\ &= (a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21} = \lambda^2 - (a_{11} + a_{22})\lambda + (a_{11}a_{22} - a_{12}a_{21}) \end{aligned}$$

i.e.,

$$\rho(\lambda) = \lambda^2 - tr(\mathbf{A})\lambda + \det(\mathbf{A})$$

Therefore, the eigenvalues are

$$\boxed{\lambda_{1,2} = \frac{tr(\mathbf{A}) \pm \sqrt{tr(\mathbf{A})^2 - 4\det(\mathbf{A})}}{2}}$$
~~~

Useful results

1. The two eigenvalues of $\mathbf{A}$ are $\left\{\begin{array}{c}\text{real}\\ \text{identical}\\ \text{complex}\end{array}\right\}$ if $\left\{\begin{array}{c}tr(\mathbf{A})^2 \geq 4\det(\mathbf{A})\\ tr(\mathbf{A})^2 = 4\det(\mathbf{A})\\ tr(\mathbf{A})^2 < 4\det(\mathbf{A})\end{array}\right\}$.

The eigenvalues of a (2×2) matrix can be identical only if they are real since complex roots of a polynomial always appear in conjugate pairs. However, matrices that are (4×4) or larger can have repeated pairs of complex roots.

It is clear from the characteristic polynomial that if $\mathbf{A}$ is triangular (or diagonal), i.e. if $a_{12}=0$ or $a_{21}=0$, then the eigenvalues of $\mathbf{A}$ are simply its diagonal elements:

$$\lambda_1 = a_{11} \text{ and } \lambda_2 = a_{22}.$$

2. The eigenvalues of a symmetric matrix are guaranteed to be real. In the (2×2) case, if $a_{12}=a_{21}$ then

$$tr(\mathbf{A})^2 - 4\det(\mathbf{A}) = (a_{11}+a_{22})^2 - 4(a_{11}a_{22}-a_{12}{}^2) = (a_{11}-a_{22})^2 + 4a_{12}{}^2 \geq 0$$

Note that the eigenvalues will also be distinct, unless $a_{11}=a_{22}$ and $a_{12}(=a_{21})=0$.

3. The product of the eigenvalues of a (2×2) matrix gives its determinant. The sum of the eigenvalues gives the trace:

$$\det(\mathbf{A}) = \lambda_1\lambda_2 \text{ and } tr(\mathbf{A}) = \lambda_1 + \lambda_2$$

In detail:

$$\begin{aligned}\lambda_1\lambda_2 &= \left(\frac{tr(\mathbf{A})+\sqrt{tr(\mathbf{A})^2-4\det(\mathbf{A})}}{2}\right)\left(\frac{tr(\mathbf{A})-\sqrt{tr(\mathbf{A})^2-4\det(\mathbf{A})}}{2}\right)\\ &= \frac{1}{4}\left(tr(\mathbf{A})^2 - tr(\mathbf{A})^2 + 4\det(\mathbf{A})\right)\\ &= \det(\mathbf{A})\end{aligned}$$

$$\lambda_1 + \lambda_2 = \frac{2tr(\mathbf{A})}{2} = tr(\mathbf{A}).$$

We will see that similar results hold for larger matrices. For the (2×2) case, one consequence of this result is that (when the eigenvalues are real):

both eigenvalues are positive $\Leftrightarrow$ $\det(\mathbf{A})>0$ and $tr(\mathbf{A})>0$

both eigenvalues are negative $\Leftrightarrow$ $\det(\mathbf{A})>0$ and $tr(\mathbf{A})<0$

the two eigenvalues have opposite signs $\Leftrightarrow$ $\det(\mathbf{A})<0$

4. If $\mathbf{A}$ is singular, i.e., $\det(\mathbf{A})=0$, then at least one of the eigenvalues are zero:

if $\det(\mathbf{A})=0$, then $\lambda_1 = tr(\mathbf{A})$, $\lambda_2=0$;

if $tr(\mathbf{A})=0$ and $\det(\mathbf{A})=0$, then $\lambda_1=\lambda_2=0$.

5. If λ_1 and λ_2 are the eigenvalues of $\mathbf{A}$, then the eigenvalues of $\mathbf{A}^2 = \mathbf{AA}$ are $\lambda_1{}^2$ and $\lambda_2{}^2$. The associated eigenvectors remain the same.

We can deduce result 5 without referring to the formula for the eigenvalues. An eigenvalue λ_1 and the associated eigenvector $\mathbf{x}_1$ satisfies

$$\mathbf{Ax}_1 = \lambda_1 \mathbf{x}_1$$

Pre-multiplying by $\mathbf{A}$ on both sides, we have

$$\mathbf{A}^2\mathbf{x}_1 = \lambda_1 \mathbf{Ax}_1 = \lambda_1(\lambda_1\mathbf{x}_1) = \lambda_1{}^2\mathbf{x}_1 .$$

This says that $\lambda_1{}^2$ is an eigenvalue of $\mathbf{A}^2$, with associated eigenvector $\mathbf{x}_1$. This result obviously generalizes to others powers of $\mathbf{A}$, and to general $(n \times n)$ matrices.

An important special case when the matrix $\mathbf{A}$ is idempotent (that is, when $\mathbf{AA} = \mathbf{A}$). Since $\mathbf{AA} = \mathbf{A}$, both $\mathbf{AA}$ and $\mathbf{A}$ have identical eigenvalues, i.e., $\lambda_i{}^2 = \lambda_i$. This means that 1 and 0 are the only possible values for eigenvalues of idempotent matrices.

6. If λ_1 and λ_2 are the eigenvalues of $\mathbf{A}$, and $\mathbf{A}$ is invertible, then the eigenvalues of $\mathbf{A}^{-1}$ are $1/\lambda_1$ and $1/\lambda_2$, with the same associated eigenvectors.

Starting with $\mathbf{Ax}_1 = \lambda_1\mathbf{x}_1$ and pre-multiplying by $\mathbf{A}^{-1}$, we have

$$\mathbf{A}^{-1}\mathbf{Ax}_1 = \lambda_1\mathbf{A}^{-1}\mathbf{x}_1$$

But the LHS is $\mathbf{x}_1$, therefore $\lambda_1\mathbf{A}^{-1}\mathbf{x}_1 = \mathbf{x}_1$, from which we get

$$\mathbf{A}^{-1}\mathbf{x}_1 = (1/\lambda_1)\mathbf{x}_1 .$$

7. The eigenvalues of $\mathbf{A}^{\mathrm{T}}$ are the same as those of $\mathbf{A}$.

That the eigenvalues of the transpose are the same as those of the original matrix is easy to see – a matrix and its transpose share the same characteristic polynomial.

8. Suppose that the eigenvalues of $\mathbf{A}$ are real and distinct, i.e. $\lambda_1 \neq \lambda_2$. Then the eigenvectors $\mathbf{x}_1$ and $\mathbf{x}_2$ are linearly independent, i.e.

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 = \mathbf{0} \;\; \Rightarrow \;\; c_1 = c_2 = 0 .$$

Proof Starting with $c_1\mathbf{x}_1 + c_2\mathbf{x}_2 = \mathbf{0}$, pre-multiply by $\mathbf{A}$ to get

$$c_1\mathbf{Ax}_1 + c_2\mathbf{Ax}_2 = \mathbf{A0} = \mathbf{0} .$$

Substituting $\mathbf{Ax}_1 = \lambda_1\mathbf{x}_1$ and $\mathbf{Ax}_2 = \lambda_2\mathbf{x}_2$ gives

$$c_1\lambda_1\mathbf{x}_1 + c_2\lambda_2\mathbf{x}_2 = \mathbf{0} .$$

Multiplying $c_1\mathbf{x}_1 + c_2\mathbf{x}_2 = \mathbf{0}$ throughout by λ_1 gives

$$c_1\lambda_1\mathbf{x}_1 + c_2\lambda_1\mathbf{x}_2 = \mathbf{0} .$$

Substituting $c_1\lambda_1\mathbf{x}_1 = -c_2\lambda_1\mathbf{x}_2$ into $c_1\lambda_1\mathbf{x}_1 + c_2\lambda_2\mathbf{x}_2 = \mathbf{0}$ gives

$$-c_2\lambda_1\mathbf{x}_2 + c_2\lambda_1\mathbf{x}_2 = \mathbf{0}\text{, or}$$

$$c_2(\lambda_2 - \lambda_1)\mathbf{x}_2 = \mathbf{0}.$$

Since $\mathbf{x}_2 \neq \mathbf{0}$ and $\lambda_1 \neq \lambda_2$, we have $c_2 = 0$. A similar argument shows that $c_1 = 0$.

9. (Diagonalization) Suppose that the eigenvalues λ_1 and λ_2 of $\mathbf{A}$ are real and distinct, with eigenvectors $\mathbf{x}_1$ and $\mathbf{x}_2$. Let

$$\mathbf{P} = \begin{bmatrix}\mathbf{x}_1 & \mathbf{x}_2\end{bmatrix} \text{ and } \mathbf{\Lambda} = \begin{bmatrix}\lambda_1 & 0 \\ 0 & \lambda_2\end{bmatrix}.$$

Then

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{\Lambda}\text{, or equivalently } \mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}.$$

We say that $\mathbf{A}$ is diagonalizable.

We note first that $\mathbf{P}^{-1}$ exists, since $\mathbf{x}_1$ and $\mathbf{x}_2$ are linearly independent. The equations $\mathbf{A}\mathbf{x}_1 = \lambda_1\mathbf{x}_1$ and $\mathbf{A}\mathbf{x}_2 = \lambda_2\mathbf{x}_2$ can be put together into a single equation $\mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{\Lambda}$. The result follows.

10. Suppose $\mathbf{A}$ is symmetric (so we know its eigenvalues are real) with distinct eigenvalues $\lambda_1 \neq \lambda_2$. Then the eigenvectors $\mathbf{x}_1$ and $\mathbf{x}_2$ are orthogonal, i.e. $\mathbf{x}_1{}'\mathbf{x}_2 = 0$.

Proof

We have $\mathbf{A}\mathbf{x}_1 = \lambda_1\mathbf{x}_1$. Pre-multiplying the first by $\mathbf{x}_2^{\mathrm{T}}$ gives $\mathbf{x}_2^{\mathrm{T}}\mathbf{A}\mathbf{x}_1 = \lambda_1\mathbf{x}_2^{\mathrm{T}}\mathbf{x}_1$. Taking transpose gives $(\mathbf{x}_2^{\mathrm{T}}\mathbf{A}\mathbf{x}_1)^{\mathrm{T}} = \mathbf{x}_1^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{x}_2 = \mathbf{x}_1^{\mathrm{T}}\mathbf{A}\mathbf{x}_2 = \lambda_1\mathbf{x}_1^{\mathrm{T}}\mathbf{x}_2$, where we have used the fact that $\mathbf{A}$ is symmetric. We also have $\mathbf{A}\mathbf{x}_2 = \lambda_2\mathbf{x}_2$. Pre-multiplying by $\mathbf{x}_1^{\mathrm{T}}$ gives $\mathbf{x}_1^{\mathrm{T}}\mathbf{A}\mathbf{x}_2 = \lambda_2\mathbf{x}_1^{\mathrm{T}}\mathbf{x}_2$. Therefore

$$\lambda_1\mathbf{x}_1^{\mathrm{T}}\mathbf{x}_2 - \lambda_2\mathbf{x}_1^{\mathrm{T}}\mathbf{x}_2 = (\lambda_1 - \lambda_2)\mathbf{x}_1^{\mathrm{T}}\mathbf{x}_2 = 0.$$

Since $\lambda_1 \neq \lambda_2$, we have $\mathbf{x}_1^{\mathrm{T}}\mathbf{x}_2 = 0$.

If the eigenvectors were chosen to have unit length, then the eigenvectors are *orthonormal*:

$$\mathbf{x}_1^{\mathrm{T}}\mathbf{x}_2 = 0 \text{ and } \mathbf{x}_1^{\mathrm{T}}\mathbf{x}_1 = 1.$$

If we pick the unit eigenvectors when constructing the matrix $\mathbf{P}$ in result (9), we would have $\mathbf{P}^{\mathrm{T}}\mathbf{P} = \mathbf{I}$. In other words, we have $\mathbf{P}^{-1} = \mathbf{P}^{\mathrm{T}}$. We say that the matrix $\mathbf{P}$ is orthonormal.

What about the case where the two eigenvalues are not distinct? For a (2×2) symmetric matrix $\mathbf{A}$, $\lambda_1 = \lambda_2 (= \lambda)$ iff

$$\mathbf{A} = \begin{bmatrix}\lambda & 0 \\ 0 & \lambda\end{bmatrix}$$

which is already diagonal. (Or put differently, every vector $\mathbf{x}$ is an eigenvector, and we are at liberty to pick a pair of orthonormal eigenvectors to construct the matrix $\mathbf{P}$, so we pick $\mathbf{x}_1 = \begin{bmatrix}1 & 0\end{bmatrix}^{\mathrm{T}}$ and $\mathbf{x}_2 = \begin{bmatrix}0 & 1\end{bmatrix}^{\mathrm{T}}$, i.e., we pick $\mathbf{P}$ to be the identity matrix.)

We can therefore re-state result (10) as the next result:

11. Suppose $\mathbf{A}$ is symmetric (in which case we know its eigenvalues are real, though perhaps not distinct). Then there exists an orthonormal matrix $\mathbf{P}$ such that

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{\Lambda}, \text{ or equivalently } \mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$$

where

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \text{ and } \mathbf{P} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{bmatrix},$$

and where $\mathbf{x}_1$ and $\mathbf{x}_2$ are eigenvectors with unit length associated with the eigenvalues λ_1 and λ_2 respectively. This is the "Spectral Theorem for Symmetric Matrices", stated and proved here for (2×2) matrices. (The result applies to general symmetric matrices).

Exercises

4. The (non-symmetric) matrix $\mathbf{A} = \begin{bmatrix} 5 & -2 \\ 4 & -1 \end{bmatrix}$ has (real, distinct) eigenvalues $\lambda_1 = 3$ and $\lambda_2 = -1$.

(a) Verify that $tr(\mathbf{A}) = \lambda_1 + \lambda_2$ and $\det(\mathbf{A}) = \lambda_1\lambda_2$;

(b) Verify that the eigenvalues of $\mathbf{A}^2$ are λ_1^2 and λ_2^2;

(c) Verify that the eigenvalues of $\mathbf{A}^{-1}$ are $1/\lambda_1$ and $1/\lambda_2$;

(d) Verify that the eigenvalues of $\mathbf{A}'$ are the same as those of $\mathbf{A}$. What are the associated eigenvectors?

(e) Verify that the eigenvectors of $\mathbf{A}$ are linearly independent;

(f) Verify the diagonalization formula in result (10);

(g) Show that the eigenvectors are **not** orthogonal.

5. Find the eigenvalues and eigenvectors of $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$. Verify result (11).

6. Eigenvectors are often normalized to length one. For instance, the eigenvector $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ of the matrix $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix}$ can be normalized to $\mathbf{x} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ so that

$$|\mathbf{x}| = \left[(1/\sqrt{2})^2 + (1/\sqrt{2})^2\right]^{1/2} = 1.$$

~~~
~~~

The (3×3) *case*

We take a brief look at the (3×3) case before moving to the general case. Let

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Its characteristic equation is

$$\det(\mathbf{A} - \lambda \mathbf{I}) = (a_{11} - \lambda)\det\begin{pmatrix} a_{22} - \lambda & a_{23} \\ a_{32} & a_{33} - \lambda \end{pmatrix} - a_{12}\det\begin{pmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} - \lambda \end{pmatrix} + a_{13}\det\begin{pmatrix} a_{21} & a_{22} - \lambda \\ a_{31} & a_{32} \end{pmatrix} \tag{13.1}$$

where we have expanded the determinant along the first row. The eigenvalues of $\mathbf{A}$ are the roots of this equation. As before, the roots may be real or complex, they may all be distinct or there may be repeated roots. If a root is repeated once, we say it has multiplicity 2. In the special case where $\mathbf{A}$ is (upper or lower) triangular or diagonal, the characteristic equation is particular easy to compute, since the determinant of such matrices is simply the product of its diagonal. In such cases, the eigenvalues are simply the diagonal elements of the matrix.

Without expanding the expansion any further, we note that the characteristic equation is an order 3 polynomial, which we can write as

$$\rho(\lambda) = b_3\lambda^3 + b_2\lambda^2 + b_1\lambda + b_0 \tag{13.2}$$

The third power of λ appears only in the first term of the determinant expansion and has coefficient $(-1)^3$. An order three polynomial has three roots (the eigenvalues), so we can write this equation as

$$\rho(\lambda) = (-1)^3(\lambda - \lambda_1)(\lambda - \lambda_2)(\lambda - \lambda_3) \tag{13.3}$$

From (13.1), the determinant of $\mathbf{A}$ can be found by setting $\lambda = 0$. Doing so in equation (13.3) show that

$$\det(\mathbf{A}) = \lambda_1\lambda_2\lambda_3$$

As in the (2×2) case, the $\det(\mathbf{A}) = 0$ iff one or more of the eigenvalues are zero.

Observe that the second power of λ also appears only in the first term in the expansion (13.1). Expanding the first term in (13.1) we have

$$\begin{aligned} &(a_{11} - \lambda)\left[(a_{22} - \lambda)(a_{33} - \lambda) - a_{23}a_{32}\right] \\ &\quad = (a_{11} - \lambda)(a_{22} - \lambda)(a_{33} - \lambda) - (a_{11} - \lambda)a_{23}a_{32} \end{aligned} \tag{13.4}$$

so the second power of λ in fact only appears in $(a_{11} - \lambda)(a_{22} - \lambda)(a_{33} - \lambda)$. Expanding this expression further, you can easily verify that the coefficient on λ^2 is $a_{11} + a_{22} + a_{33}$, which is $tr(\mathbf{A})$. Expanding (13.3), we see that the coefficient on λ^2 there is $\lambda_1 + \lambda_2 + \lambda_3$. Matching coefficients, we see that

$$tr(\mathbf{A}) = \lambda_1 + \lambda_2 + \lambda_3 .$$

All these results extend to the general $(n \times n)$ case: for an $(n \times n)$ matrix $\mathbf{A}$ we have

$$\det(\mathbf{A}) = \lambda_1\lambda_2\lambda_3 ... \lambda_n \text{ and } tr(\mathbf{A}) = \lambda_1 + \lambda_2 + ... + \lambda_n .$$

The arguments for results (5)-(10) in Section 2 also carry over to the $(n \times n)$ case with little or no amendment: if λ_1 is an eigenvalue of $\mathbf{A}$, then λ_1 is also an eigenvalue of $\mathbf{A}^{\mathrm{T}}$, λ_1^n is an eigenvalue of $\mathbf{A}^n$, and if $\mathbf{A}$ is invertible, then $1/\lambda_1$ is an eigenvalue of $\mathbf{A}^{-1}$; Eigenvectors associated with distinct eigenvalues are linearly independent, and $\mathbf{A}$ has n distinct eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$, with associated eigenvectors $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$, then it is diagonalizable: letting

$$\mathbf{P} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix} \quad \text{and} \quad \mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

Then

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{\Lambda}, \text{ or equivalently } \mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}.$$

The general $(n \times n)$ case

The **Spectral Theorem for Symmetric Matrices** also continues to hold: if the $(n \times n)$ matrix $\mathbf{A}$ is symmetric, then

(i) all its eigenvalues are real;

(ii) the eigenvectors the correspond to different eigenvalues are orthogonal,

(iii) there exists an orthonormal matrix $\mathbf{P}$ (i.e. $\mathbf{P}^{-1} = \mathbf{P}^{\mathrm{T}}$) comprising the unit eigenvectors such that

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{\Lambda}, \text{ or equivalently } \mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$$

The proof in the case when the eigenvalues are distinct is a simple extrapolation of the argument following result (11). For the proof when there are repeated roots, see Strang (2009) Section 6.4.

Exercise

7. Suppose $\mathbf{\Sigma}$ is symmetric, with non-zero eigenvalues. Use the spectral theorem to define the matrix $\mathbf{\Sigma}^{1/2}$ such that $\mathbf{\Sigma} = \mathbf{\Sigma}^{1/2}\mathbf{\Sigma}^{1/2}$. Similarly, find $\mathbf{\Sigma}^{-1/2}$ such that $\mathbf{\Sigma}^{-1} = \mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}^{-1/2}$.

Eigenvalues and the Rank of a Matrix

We often have to determine the rank of a matrix (the number of linear independent rows or columns contained in the matrix). The Spectral Theorem for symmetric matrices makes it very easy to do so. We use the following result:

$$\text{If } \mathbf{A} \text{ is } (m \times n) \text{ and } \mathbf{B} \text{ is } (n \times n) \text{ and full rank, then } rank(\mathbf{AB}) = rank(\mathbf{A})$$

which extends to the product $\mathbf{CAB}$ where both $\mathbf{C}$ and $\mathbf{B}$ are full rank:

$$rank(\mathbf{CAB}) = rank(\mathbf{A}).$$

This means that if $\mathbf{A}$ is a square diagonalizable matrix, then

$$rank(\mathbf{A}) = rank(\mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}) = rank(\mathbf{\Lambda})$$

The rank of $\mathbf{\Lambda}$ is simply the number of non-zero terms on the diagonal, i.e. the number of non-zero eigenvalues of $\mathbf{A}$.

Furthermore, if $\mathbf{A}$ is symmetric and idempotent, then the eigenvalues of $\mathbf{A}$ take on values 1 or 0 only, i.e., the diagonal of $\mathbf{\Lambda}$ comprises only 1's and 0's. In this case,

$$rank(\mathbf{\Lambda}) = \text{trace}(\mathbf{\Lambda})$$

(remember that the trace of a square matrix is just the sum of the elements on its diagonal.) Furthermore, because

$$\mathbf{\Lambda} = \mathbf{P}^{-1}\mathbf{AP}$$

we have

$$trace(\mathbf{\Lambda}) = trace(\mathbf{P}^{-1}\mathbf{AP}) = trace(\mathbf{PP}^{-1}\mathbf{A}) = trace(\mathbf{A})$$

where we have used the fact that $trace(\mathbf{AB}) = trace(\mathbf{BA})$ when both products exist. It is therefore very easy to find the rank of a symmetric idempotent matrix. Simply add up the elements in its diagonal – you don't even have to find the eigenvalues!

The result

$$rank(\mathbf{A}) = rank(\mathbf{P}^{-1}\mathbf{\Lambda P}) = rank(\mathbf{\Lambda})$$

always works for symmetric matrices since these are always diagonalizable. But this approach can also work with non-diagonalizable matrices, and even non-square matrices. Recall that for any matrix $\mathbf{A}$,

$$rank(\mathbf{A}^{\mathrm{T}}\mathbf{A}) = rank(\mathbf{A})$$

Since $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ is symmetric, we can compute its rank, and therefore the rank of $\mathbf{A}$, by computing and counting the number of non-zero eigenvalues possessed by $\mathbf{A}^{\mathrm{T}}\mathbf{A}$.

Exercise

8. Let $\mathbf{X}$ be an arbitrary $(n \times k)$ matrix such that $(\mathbf{X}'\mathbf{X})^{-1}$ exists. Because $\mathbf{X}'\mathbf{X}$ is symmetric (why?), we know that $(\mathbf{X}'\mathbf{X})^{-1}$ is also symmetric (why?). Use this fact to show that the matrix

$$\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})\mathbf{X}'$$

is symmetric and idempotent. Find its rank.

Positive Definiteness of Quadratic Forms

A quadratic form is an expression of the form

$$\mathbf{Q} = \mathbf{x}^{\mathrm{T}}\mathbf{Ax}$$

When $\mathbf{A}$ is (2×2), this expression is

$$\mathbf{Q} = \mathbf{x}^{\mathrm{T}}\mathbf{Ax} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = a_{11}x_1^2 + (a_{12} + a_{21})x_1x_2 + a_{22}x_2^2$$

For a general $(n \times n)$ matrix, we have

$$\mathbf{Q} = \mathbf{x}^{\mathrm{T}}\mathbf{Ax} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j$$

The following are examples of quadratic forms:

(i) $q_1 = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^2 + 2x_1x_2 + x_2^2$

(ii) $q_2 = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2x_1^2 + x_2^2$

(iii) $q_3 = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^2 + 4x_1x_2 + 3x_2^2$

(iv) $q_4 = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^2 + 6x_1x_2 + 3x_2^2$

The quadratic forms in (i), (ii) and (iii) involve symmetric matrices, whereas (iv) does not. Note, however, that for any quadratic form involving a non-symmetric matrix, there is an equivalent one using a symmetric matrix. In the (2×2) case, replace both a_{12} and a_{21} by $(a_{12}+a_{21})/2$. For example, corresponding to (iv) we have

$$q_4 = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^2 + 6x_1x_2 + 3x_2^2 .$$

In other words, we can limit ourselves to studying quadratic forms involving symmetric matrices.

For larger $(n\times n)$, matrices, replace $\mathbf{A}$ with $(\mathbf{A}+\mathbf{A}^{\mathrm{T}})/2$.

In applications, we often need to determine the sign of a quadratic form for arbitrary values of $\mathbf{x} \neq \mathbf{0}$. For instance, the quadratic form q_1 can be written as

$$q_1 = x_1^2 + 2x_1x_2 + x_2^2 = (x_1+x_2)^2$$

so we know that $q_1 \geq 0$ no matter what values of x_1 and x_2 are chosen. We call such quadratic forms positive semi-definite. The quadratic form q_2 has a similar (but stronger) behavior. This quadratic form can be written as

$$q_2 = 2x_1^2 + x_2^2 > 0$$

for all values of x_1 and x_2, not both equal to zero at the same time. This quadratic form is said to be positive definite. The quadratic form q_3, however, is "indefinite": writing

$$q_3(x_1, x_2) = x_1^2 + 4x_1x_2 + 3x_2^2$$

we have $q_3(-2,1) = 4-8+3 = -1$ whereas $q_3(2,1) = 4+8+3 = 15$. That is, the quadratic form is negative for some values of x_1 and x_2, and positive for others.

A quadratic form $\mathbf{Q} = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$ is said to be

-- positive definite if $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$;

-- positive semi-definite if $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x} \neq \mathbf{0}$;

-- negative definite if $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} < 0$ for all $\mathbf{x} \neq \mathbf{0}$;

-- negative semi-definite if $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} \leq 0$ for all $\mathbf{x} \neq \mathbf{0}$;

Although definiteness pertain to the quadratic form $\mathbf{Q} = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$, we often apply the term to the matrix $\mathbf{A}$. One application of these concepts is in optimization theory where the second order condition often involves determining the definiteness of the Hessian matrix. Another application is in 'comparing matrices'.

One interesting fact about definite symmetric matrices is they can be factorized into the product of a triangle matrix and its transpose. We state and explain this for positive definite symmetric matrices:

> **Triangular Factorization** Any positive definite symmetric $(n \times n)$ matrix $\mathbf{A}$ has a unique representation of the form
>
> $$\mathbf{A} = \mathbf{L}\mathbf{D}\mathbf{L}^{\mathrm{T}}$$
>
> where L is lower triangular with ones down the diagonal, and $\mathbf{D}$ is diagonal with positive diagonal elements.

We demonstrate this for a positive definite symmetric (3×3) matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

Because $\mathbf{A}$ is positive definite, we have $a_{11} > 0$ (...pick $\mathbf{x}^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$). We can construct the elimination matrix

$$\mathbf{E}_1 = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{a_{21}}{a_{11}} & 1 & 0 \\ -\frac{a_{31}}{a_{11}} & 0 & 1 \end{bmatrix}$$

You can easily verify that

$$\mathbf{E}_1\mathbf{A}\mathbf{E}_1^{\mathrm{T}} = \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & b_{22} & b_{23} \\ 0 & b_{32} & b_{33} \end{bmatrix}$$

where $b_{ij} = a_{ij} - a_{i1}a_{1j} / a_{11}$.

Note that if $\mathbf{A}$ is positive definite, then $\mathbf{E}_1\mathbf{A}\mathbf{E}_1^{\mathrm{T}}$ must also be positive definite: define $\mathbf{y}$ such that $\mathbf{x} = \mathbf{E}_1^{\mathrm{T}}\mathbf{y}$ which then allows us to write $\mathbf{y}^{\mathrm{T}}\mathbf{E}_1\mathbf{A}\mathbf{E}_1^{\mathrm{T}}\mathbf{y} = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$. Because $\mathbf{E}_1$ is non-singular, $\mathbf{x} = \mathbf{0} \Leftrightarrow \mathbf{y} = \mathbf{0}$. Therefore $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$ implies $\mathbf{y}^{\mathrm{T}}\mathbf{E}_1\mathbf{A}\mathbf{E}_1^{\mathrm{T}}\mathbf{y} > 0$ for all $\mathbf{y} \neq \mathbf{0}$.

Because $\mathbf{E}_1\mathbf{A}\mathbf{E}_1^{\mathrm{T}}$ is positive definite, $b_{22} > 0$ (... pick $\mathbf{y}^{\mathrm{T}} = [0 \quad 1 \quad 0]$), so we can define

$$\mathbf{E}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\frac{b_{32}}{b_{22}} & 1 \end{bmatrix}$$

You can verify that

$$\mathbf{E}_2\mathbf{E}_1\mathbf{A}\mathbf{E}_1{}^{\mathrm{T}}\mathbf{E}_2{}^{\mathrm{T}} = \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & b_{22} & 0 \\ 0 & 0 & c_{33} \end{bmatrix}$$

where $c_{33} = b_{33} - \frac{b_{32}{}^2}{b_{22}}$.

Call the resulting diagonal matrix $\mathbf{D}$. Because $\mathbf{E}_1$ and $\mathbf{E}_2$ are lower triangular with ones down the diagonal, $\mathbf{E}_2\mathbf{E}_1$ has the same structure. Furthermore, $\mathbf{L} = (\mathbf{E}_2\mathbf{E}_1)^{-1}$ exists and has the same structure. Therefore

$$\mathbf{E}_2\mathbf{E}_1\mathbf{A}\mathbf{E}_1{}^{\mathrm{T}}\mathbf{E}_2{}^{\mathrm{T}} = \mathbf{D} \iff \mathbf{A} = \mathbf{L}\mathbf{D}\mathbf{L}^{\mathrm{T}}.$$

We can go one step further. Defining

$$\mathbf{D} = \begin{bmatrix} \sqrt{a_{11}} & 0 & 0 \\ 0 & \sqrt{b_{22}} & 0 \\ 0 & 0 & \sqrt{c_{33}} \end{bmatrix},$$

we have $\mathbf{A} = \mathbf{L}\mathbf{D}\mathbf{L}^{\mathrm{T}} = \mathbf{L}\mathbf{D}^{1/2}\mathbf{D}^{1/2}\mathbf{L}^{\mathrm{T}} = \mathbf{C}\mathbf{C}^{\mathrm{T}}$ where $\mathbf{C} = \mathbf{L}\mathbf{D}^{1/2}$ is also lower triangular. This is the Cholesky factorization (or Cholesky decomposition) of $\mathbf{A}$. These decompositions are related to, but different from, the eigenvalue-eigenvector decomposition discussed earlier.

These arguments readily extend to the general case.

Proving the indefiniteness of a matrix merely requires providing counter examples, as we did for q_3 above. To prove the definiteness of a matrix is much harder because we have to show that the requisite sign holds for all $\mathbf{x} \neq \mathbf{0}$. The objective here is to develop criteria for evaluating the definiteness of a matrix. We provide two sets of results, one using principal minors, and the other using eigenvalues. It is worthwhile previewing the results with (2×2) matrices, before results for the general case are given.

For a given (2×2) symmetric matrix $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix}$, let

$$\mathbf{Q}(x_1, x_2) = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2,$$

for arbitrary $\mathbf{x} \neq \mathbf{0}$. Then

(i) $\mathbf{Q}(x_1, x_2)$ is positive semi-definite $\iff$ $a_{11} \geq 0$, $a_{22} \geq 0$, and $a_{11}a_{22} - a_{12}{}^2 \geq 0$;
(ii) $\mathbf{Q}(x_1, x_2)$ is positive definite $\iff$ $a_{11} > 0$ and $a_{11}a_{22} - a_{12}{}^2 > 0$;
(iii) $\mathbf{Q}(x_1, x_2)$ is negative semi-definite $\iff$ $a_{11} \leq 0$, $a_{22} \leq 0$, and $a_{11}a_{22} - a_{12}{}^2 \geq 0$;
(iv) $\mathbf{Q}(x_1, x_2)$ is negative definite $\iff$ $a_{11} < 0$ and $a_{11}a_{22} - a_{12}{}^2 > 0$.

Proof:

(i) Suppose $a_{11} \geq 0$, $a_{22} \geq 0$, and $a_{11}a_{22} - a_{12}{}^2 \geq 0$.

Case 1: $a_{11} = 0$. Then $a_{11}a_{22} - a_{12}{}^2 \geq 0$ implies $a_{12} = 0$, so the quadratic form is $\mathbf{Q}(x_1, x_2) = a_{22}x_2^2 \geq 0$.

Case 2: $a_{11} > 0$. Then

$$\mathbf{Q}(x_1, x_2) = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2$$

$$= a_{11}(x_1^2 + 2\frac{a_{12}}{a_{11}}x_1x_2 + \frac{a_{22}}{a_{11}}x_2^2)$$

$$= a_{11}\left[\left(x_1 + \frac{a_{12}}{a_{11}}x_2\right)^2 + \frac{a_{22}}{a_{11}}x_2^2 - \left(\frac{a_{12}}{a_{11}}\right)^2 x_2{}^2\right]$$

$$= a_{11}\left[\left(x_1 + \frac{a_{12}}{a_{11}}x_2\right)^2 + \frac{a_{11}a_{22} - a_{12}{}^2}{a_{11}{}^2}x_2{}^2\right] \qquad (*)$$

With $a_{11} > 0$ and $a_{11}a_{22} - a_{12}{}^2 \geq 0$, clearly $\mathbf{Q}(x_1, x_2) \geq 0$ for all x_1, x_2 not both equal to zero.

We now show the converse: Suppose $\mathbf{Q}(x_1, x_2) = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 \geq 0$ for all x_1, x_2 not both equal to zero. Then, in particular, we have

$$\mathbf{Q}(1,0) = a_{11} \geq 0 \text{, and } \mathbf{Q}(0,1) = a_{22} \geq 0.$$

If $a_{11} = 0$, then $\mathbf{Q}(x_1, 1) = 2a_{12}x_1 + a_{22} \geq 0$, which implies $a_{12} = 0$, since if $a_{12} > 0$, we can make $\mathbf{Q}(x_1, 1) < 0$ by choosing x_1 a large enough negative number, and if $a_{12} < 0$, we can make $\mathbf{Q}(x_1, 1) < 0$ by choosing x_1 to be a large enough positive number. Then $a_{11}a_{22} - a_{12}{}^2 = 0$. If $a_{11} > 0$, then we must have $a_{11}a_{22} - a_{12}{}^2 \geq 0$, otherwise we can make $\mathbf{Q}(x_1, x_2) < 0$ by choosing x_1 and x_2 to make

$$x_1 + \frac{a_{12}}{a_{11}}x_2 = 0.$$

Result (iii) is proved in a similar fashion.

(ii) Suppose $a_{11} > 0$ and $a_{11}a_{22} - a_{12}{}^2 > 0$. Then from (*), $\mathbf{Q}(x_1, x_2) > 0$. Suppose $\mathbf{Q}(x_1, x_2) > 0$ for all x_1, x_2 not both equal to zero. Then $\mathbf{Q}(1,0) = a_{11} > 0$. Because $a_{11} > 0$, we can write (*). Then $\mathbf{Q}(-a_{12}/a_{11}, 1) = (a_{11}a_{22} - a_{12}{}^2)/a_{11} > 0$, which implies $a_{11}a_{22} - a_{12}{}^2 > 0$.

Result (iv) is proved in similar fashion.

We re-state this result for general $(n \times n)$ symmetric matrices (without proof):

Theorem Consider the $(n \times n)$ symmetric matrix $\mathbf{A}$ and the associated quadratic form

$$Q(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j,$$

where $\mathbf{x}$ is an arbitrary non-zero n-dimensional vector. Let D_k be the k-th leading principal minor of $\mathbf{A}$, and Δ_k denote an arbitrary principal minors of order k. Then $Q(\mathbf{x})$ is

(a) positive definite $\Leftrightarrow$ $D_k > 0$ for $k = 1, 2, \ldots, n$

(b) positive semi-definite $\Leftrightarrow$ $\Delta_k \geq 0$ for $k = 1, 2, \ldots, n$.

(c) negative definite $\Leftrightarrow$ $(-1)^k D_k > 0$ for $k = 1, 2, \ldots, n$.

(d) negative semi-definite $\Leftrightarrow$ $(-1)^k \Delta_k \geq 0$ for $k = 1, 2, \ldots, n$.

It should be clear that the results stated and proved for the (2×2) case is a special case of this theorem.

(Note: to add definition of principal minors and leading principal minors. For the moment, look it up.)

An Eigenvalue Approach Eigenvalues provide a more convenient way to determine definiteness of quadratic forms. Because we are dealing with symmetric matrices, which are diagonalizable, we can rewrite our quadratic form as

$$\mathbf{Q}(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \mathbf{x}^{\mathrm{T}}\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{\mathrm{T}}\mathbf{x} = \mathbf{y}^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{y} = \lambda_1 y_1{}^2 + \lambda_2 y_2{}^2 + \ldots + \lambda_{n-1} y_{n-1}{}^2 + \lambda_n y_n{}^2$$

where $\mathbf{y} = \mathbf{P}^{\mathrm{T}}\mathbf{x}$. It should be clear that "$\mathbf{Q}(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$" is equivalent to "$\mathbf{Q}(\mathbf{x}) > 0$ for all $\mathbf{y} \neq \mathbf{0}$", which is in turn equivalent to $\lambda_r > 0$, $r = 1,\ldots,n$. Similarly, "$\mathbf{Q}(\mathbf{x}) \geq 0$ for all $\mathbf{x} \neq \mathbf{0}$" is equivalent to "$\mathbf{Q}(\mathbf{x}) \geq 0$ for all $\mathbf{y} \neq \mathbf{0}$", which is in turn equivalent to $\lambda_r \geq 0$, $r = 1,\ldots,n$. That is,

For a given symmetric $(n \times n)$ matrix $\mathbf{A}$, let $\mathbf{Q}(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$, for arbitrary $\mathbf{x} \neq \mathbf{0}$. Then

(i) $\mathbf{Q}(\mathbf{x})$ is positive semi-definite $\Leftrightarrow$ $\lambda_r \geq 0$, $r = 1,\ldots,n$

(ii) $\mathbf{Q}(\mathbf{x})$ is positive definite $\Leftrightarrow$ $\lambda_r > 0$, $r = 1,\ldots,n$

(iii) $\mathbf{Q}(\mathbf{x})$ is negative semi-definite $\Leftrightarrow$ $\lambda_r \leq 0$, $r = 1,\ldots,n$

(iv) $\mathbf{Q}(\mathbf{x})$ is negative definite $\Leftrightarrow$ $\lambda_r < 0$, $r = 1,\ldots,n$