Ch. 1 Linear Algebra[‡]

(October 8, 2018)

Probably the most important problem in mathematics is that of solving a system of linear equations. Using modern mathematics, it is often possible to take a sophisticated problem and reduce it to a single system of linear equations. Linear algebra and matrix theory are essentially synonymous terms for the area of mathematics that has became one of the most useful and pervasive tools in a wide rang of disciplines to solve a system of linear equations. It is also a subject of great mathematical beauty.

1 Vector Space

The concept of a vector is a very useful one. This utility from two important aspects of vectors, namely that they engender a highly geometrical insight, which is of course much to be desired, and that vector notation permits many complicated formulas to be written in a very compact form. With this economy of notation comes a greater ease in handling problem.

[‡]Editorial assistance from Qi Wang, School of Economics and Management, Tongji University, Shanghai, China for the following 11 Chapters is highly appreciated.

1.1 Vector Space Axiom

A vector space (also called a linear space) is a collection of objects called *vectors*, which may be added together and multiplied ("scaled") by numbers, called scalars in this context. That is, a vector space (over a field $\mathbb{F}$) is a set $\mathcal{V}$ admitting two "operation", called *multiplication by scalars* and *addition*:

(a). If $\mathbf{x} \in \mathcal{V}$ and α is a scalar, then $\alpha\mathbf{x} \in \mathcal{V}$.

(b). If $\mathbf{x}, \mathbf{y} \in \mathcal{V}$, then $\mathbf{x} + \mathbf{y} \in \mathcal{V}$,

They also satisfy the following conditions:

(c). $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ for any $\mathbf{x}$ and $\mathbf{y}$ in $\mathcal{V}$.

(d). $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$.

(e). There exists an element $\mathbf{0}$ in $\mathcal{V}$ such that $\mathbf{x} + \mathbf{0} = \mathbf{x}$ for each $\mathbf{x} \in \mathcal{V}$.

(f). For each $\mathbf{x} \in \mathcal{V}$, there exists an element $-\mathbf{x} \in \mathcal{V}$ such that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$.

(g). $\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$ for each real number α and any $\mathbf{x}$ and $\mathbf{y}$ in $\mathcal{V}$.

(h). $(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}$ for any real number α and β and any $\mathbf{x} \in \mathcal{V}$.

(i). $(\alpha\beta)\mathbf{x} = \alpha(\beta\mathbf{x})$ for any real number α and β and any $\mathbf{x} \in \mathcal{V}$.

(j). $1 \cdot \mathbf{x} = \mathbf{x}$ for all $\mathbf{x} \in \mathcal{V}$. □

Example.

A familiar example of a vector space is the n-dimensional Euclidean space $\mathbb{R}^n$. Here, addition and multiplication are defined as follows: If $(u_1, u_2, ..., u_n)'$ and $(v_1, v_2, ..., v_n)'$ are two elements in $\mathbb{R}^n$, then their sum is defined as $(u_1 + v_1, u_2 + v_2, ..., u_n + v_n)'$ which is also an element of $\mathbb{R}^n$. If α is a scalar, then $\alpha(u_1, u_2, ..., u_n)' = (\alpha u_1, \alpha u_2, ..., \alpha u_n)'$.[1] ■

[1] Example: The two-dimensional plane is the set of all vectors with two real-valued coordinates. We label this set $\mathbb{R}^2$. It has two important properties.
(a). $\mathbb{R}^2$ is closed under scalar multiplication; every scalar multiple of a vector in the plane is also in the plane.
(b). $\mathbb{R}^2$ is closed under addition; the sum of any two vectors is always a vector in the plane.

Example.

Let $\mathcal{V}$ be the set of all polynomials in x of degree less than or equal to k. Then $\mathcal{V}$ is a vector space. Any element in $\mathcal{V}$ can be expressed as $\sum_{i=0}^{k} a_i x^i$, where the a_i's are scalars. ■

1.2 Euclidean Vector Space

Perhaps the most elementary vector is the Euclidean vector space $\mathbb{R}^n, n = 1, 2, \ldots$. For simplicity, let us consider first $\mathbb{R}^2$.

Non-zero vector in $\mathbb{R}^2$ can be represented geometrically by directed line segments. Given a nonzero vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ (or $\mathbf{x} = (x_1\ x_2)$) we can associate it with the line segment in the plane from $(0, 0)$ to (x_1, x_2). If we equate line segment that have the same length and direction, $\mathbf{x}$ can be represented by **any** line segment from (a, b) to $(a + x_1, b + x_2)$. For example, the vector $\mathbf{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ in $\mathbb{R}^2$ could be represented by the directed line segment from $(2, 2)$ to $(4, 3)$, or from $(-1, -1)$ to $(1, 0)$.

Definition.

The Euclidean length of a vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ is the length of any directed line segment representing $\mathbf{x}$. The length of the segment from $(0, 0)$ to (x_1, x_2) is $\sqrt{x_1^2 + x_2^2}$ $(= \mathbf{x}'\mathbf{x})$. The length is also called the Euclidean norm and is denoted as $\|\mathbf{x}\|$. ■

Two basic operations are defined for vectors, **scalar multiplication** and **addition**. The geometric representation will help us to visualize how the operation of scalar multiplication and addition work in $\mathbb{R}^2$.

(a). Scalar multiplication:

For each vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and each scalar α, the product $\alpha\mathbf{x}$ is defined by

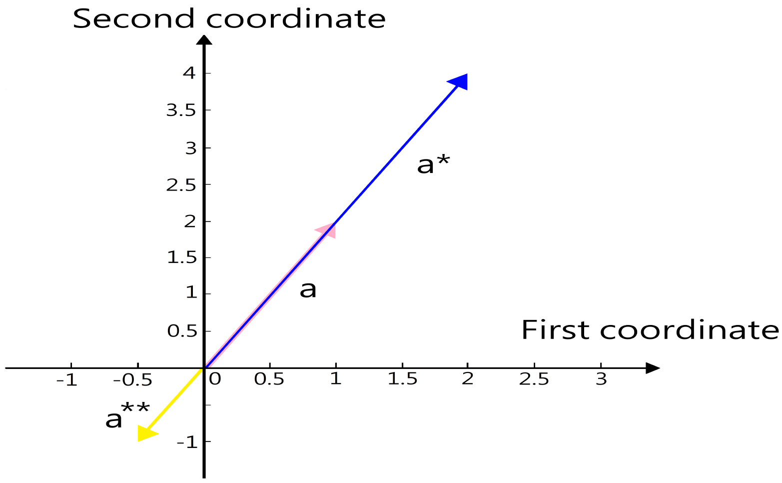
$$\alpha\mathbf{x} = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \end{bmatrix}.$$

The set of all possible scalar multiple of $\mathbf{x}$ is the line through $\mathbf{0}$ and $\mathbf{x}$. Any scalar multiple of $\mathbf{x}$ is a segment of this line.

Example.

$$\mathbf{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \mathbf{a}^* = \mathbf{2a} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \quad \mathbf{a}^{**} = -\frac{\mathbf{1}}{\mathbf{2}}\mathbf{a} = \begin{bmatrix} -\frac{1}{2} \\ -1 \end{bmatrix}.$$

The vector $\mathbf{a}^*(= \mathbf{2a})$ is in the same direction as $\mathbf{a}$, but its length is two times that of $\mathbf{a}$. The vector $\mathbf{a}^{**}(= -\frac{\mathbf{1}}{\mathbf{2}}\mathbf{a})$ has half of length as $\mathbf{a}$ but its point in the opposite direction.

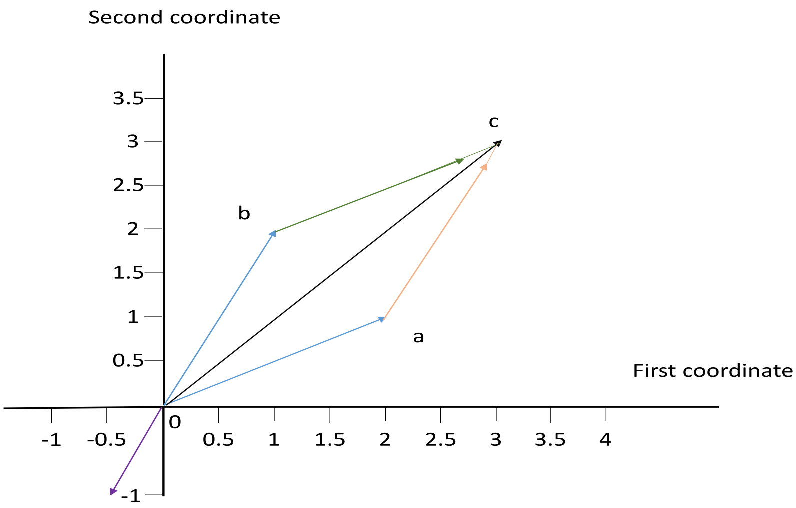


(b). Addition:

The sum of two vectors $\mathbf{a}$ and $\mathbf{b}$ is a third vector whose coordinates are the sums of the corresponding coordinates of $\mathbf{a}$ and $\mathbf{b}$. For example ,

$$\mathbf{c} = \mathbf{a} + \mathbf{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}.$$

Geometrically, $\mathbf{c}$ is obtained by moving in the distance and direction defined by $\mathbf{b}$ from the tip of $\mathbf{a}$ or, because addition is commutative, from the tip of $\mathbf{b}$ in the distance and direction of $\mathbf{a}$.



■

In a similar manner, vectors in $\mathbb{R}^3$ can be represented by directed line segments in a 3-space. Vector in $\mathbb{R}^n$ can be views as the coordinates of a point in a n-dimensional space or as the definition of the line segment connecting the origin and this point.

In general, scalar multiplication and addition in $\mathbb{R}^n$ are defined by

$$\alpha\mathbf{x} = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \cdot \\ \cdot \\ \cdot \\ \alpha x_n \end{bmatrix} \quad \text{and} \quad \mathbf{x}+\mathbf{y} = \begin{bmatrix} x_1+y_1 \\ x_2+y_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n+y_n \end{bmatrix}$$

for any $\mathbf{x}$ and $\mathbf{y} \in \mathbb{R}^n$ and any scalar α.

1.3 Linear Combination of Vectors and Basis Vectors

We can combine the two operations "addition" and "scalar multiplication" above as a whole which is known as a "linear combination".

Definition.

Let $\mathbf{u} \in \mathcal{V}$ and $\mathbf{v} \in \mathcal{V}$, the vector $\mathbf{w} = \alpha\mathbf{u} + \beta\mathbf{v} \in \mathcal{V}$ is called a linear combination of $\mathbf{u}$ and $\mathbf{v}$, where α and β are scalars. ■

Definition.

A set of vectors in a vector space is a basis for that vector space if any vector in the vector space can be written as a linear combination of them. The number of elements in this basis is called the *dimension* of the vector space. ■

Result.

Any pair of two dimensional vectors that point *in different directions* will form a basis for $\mathbb{R}^2$.

Proof.

Consider an arbitrary set of vectors in $\mathbb{R}^2$, $\mathbf{a}, \mathbf{b}$, and $\mathbf{c}$. If $\mathbf{a}$ and $\mathbf{b}$ are a basis, we can find numbers α_1 and α_2 such that $\mathbf{c} = \alpha_1\mathbf{a} + \alpha_2\mathbf{b}$. Let

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad \textit{and} \quad \mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}.$$

Then

$$\begin{aligned} c_1 &= \alpha_1 a_1 + \alpha_2 b_1, \\ c_2 &= \alpha_1 a_2 + \alpha_2 b_2. \end{aligned}$$

The solutions to this pair of equations are

$$\alpha_1 = \frac{b_2c_1 - b_1c_2}{a_1b_2 - b_1a_2}, \tag{1-1}$$

$$\alpha_2 = \frac{a_1c_2 - a_2c_1}{a_1b_2 - b_1a_2}. \tag{1-2}$$

This gives a unique solution unless $(a_1b_2 - b_1a_2) = 0$. If $(a_1b_2 - b_1a_2) = 0$, then

$a_1/a_2 = b_1/b_2$, which means that **b** is just a multiple of **a**. This returns us to our original condition, that **a** and **b** point in different direction. The implication is that if **a** and **b** are any pair of vectors for which the denominator in (1-1) and (1-2) is not zero, then any other vector **c** can be formed as a *unique* linear combination of **a** and **b**.■

The basis of a vector space is not unique, since any set of vectors that satisfy the definition will do. But for any particular basis, there is only one linear combination of them that will produce another particular vector in the vector space.

1.4 Linear Dependence

As the preceding should suggest, k vectors are required to form a basis for $\mathbb{R}^k$. However it is not every set of k vectors will suffices. As we see, to form a basis we require that this k vectors to be linearly independent.

$\boxed{\text{Definition.}}$

A sets of vectors is **linearly dependent** if any one of the vectors in the set can be written as a linear combination of the others. ■

$\boxed{\text{Definition.}}$

The vector $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ in a vector space $\mathcal{V}$ are said to be **linearly independent** if and only if the solution to

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + ... + c_n\mathbf{v}_n = \mathbf{0}$$

is

$$c_1 = c_2 = ... = c_n = 0.$$

■

Example.

The vector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ are linear independent, since if

$$c_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

then

$$\begin{aligned} c_1 + c_2 &= 0 \\ c_1 + 2c_2 &= 0 \end{aligned}$$

and the only solution to this system is $c_1 = c_2 = 0$. ■

1.5 Subspace

Given a vector space $\mathcal{V}$, it is often possible to form another vector space by taking s subset $\mathcal{S}$ of $\mathcal{V}$ and using the operations of $\mathcal{V}$.

Definition.

If $\mathcal{S}$ is a nonempty subset of a vector space $\mathcal{V}$, and $\mathcal{S}$ satisfies the following conditions:
(a). $\alpha\mathbf{x} \in \mathcal{S}$, whenever $\mathbf{x} \in \mathcal{S}$ for any scalar α.
(b). $\mathbf{x} + \mathbf{y} \in \mathcal{S}$ whenever $\mathbf{x} \in \mathcal{S}$ and $\mathbf{y} \in \mathcal{S}$, then $\mathcal{S}$ is said to be a subspace of $\mathcal{V}$. ■

Definition. (Linear Span)

Let $\mathbf{u}_1$, $\mathbf{u}_2$,..., $\mathbf{u}_n$ be n elements in a vector space $\mathcal{V}$. The collection of all linear combinations of the form $\sum_{i=1}^{n} \alpha_i \mathbf{u}_i$, where the α_i's are scalars, is called a linear span of $\mathbf{u}_1$, $\mathbf{u}_2$,..., $\mathbf{u}_n$. ■

Example.

$\mathbb{R}^k = Span(\mathbf{v_1}, ..., \mathbf{v_k})$ for a basis $(\mathbf{v_1}, ..., \mathbf{v_k})$. ■

We now consider what happens to the vector space that is spanned by linearly dependent vectors.

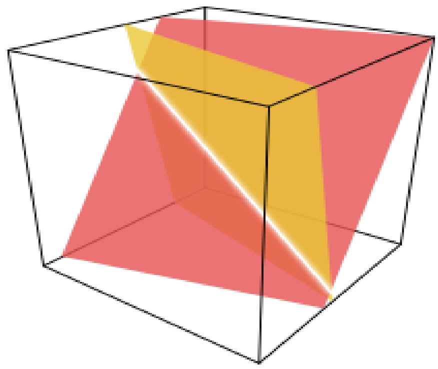
Example.

$\mathcal{S}=$ Span $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is a **one-dimensional subspace in** $\mathbb{R}^2$ since $\alpha \begin{bmatrix} 1 \\ 2 \end{bmatrix} \in \mathcal{S}$ and $\alpha \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \beta \begin{bmatrix} 1 \\ 2 \end{bmatrix} = (\alpha+\beta)\begin{bmatrix} 1 \\ 2 \end{bmatrix} \in \mathcal{S}$. Furthermore, $\mathcal{S} \subset \mathbb{R}^2$. ■

Therefore, the space spanned by a set of vectors in $\mathbb{R}^k$ has at most k dimensions. If this space has fewer than k dimensions, it is subspace, or hyperplane. But the important point in the preceding discussion is that every set of vectors spans some space; it may be the entire space in which the vector reside, or some subspace of it.

Example.

In $\mathbb{R}^3$, the intersection of two-dimensional subspaces is one-dimensional:



Exercise 1.

Let $\mathcal{S} = \{(x_1, x_2, x_3)'|x_1 = x_2\}$. Show that $\mathcal{S}$ is a subspace of $\mathbb{R}^3$. ■

1.6 Vector Projection

Definition.

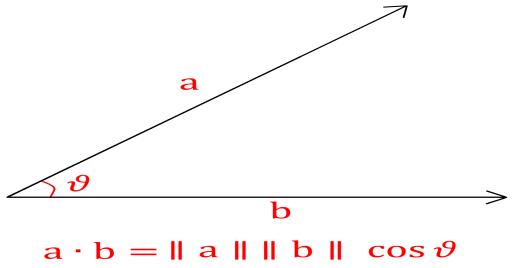
The dot product of two vectors, $\mathbf{a}_n$ and $\mathbf{b}_n$, is a scalar and is written as

$$\begin{aligned}\mathbf{a}_n \cdot \mathbf{b}_n &= a_1 \times b_1 + a_2 \times b_2 + ... + a_n \times b_n \\ &= \|\mathbf{a}_n\| \times \|\mathbf{b}_n\| \times \cos(\theta) \\ &= \mathbf{b}_n \cdot \mathbf{a}_n,\end{aligned}$$

where θ is the angle between $\mathbf{a}_n$ and $\mathbf{b}_n$. Two vectors $\mathbf{a}_n$ and $\mathbf{b}_n$ are said to be *orthogonal* if $\mathbf{a}_n \cdot \mathbf{b}_n = 0$.[2] ■

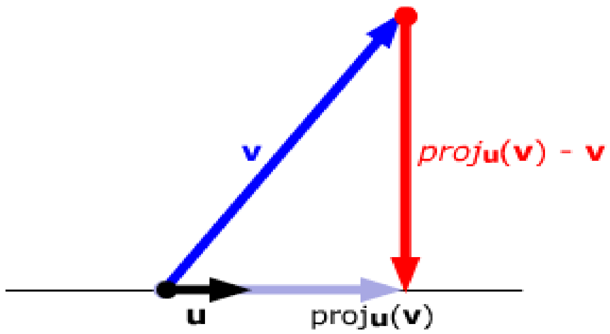
It is apparent that $\mathbf{a}_n \cdot \mathbf{a}_n = \sum a_i^2 = \|\mathbf{a}\|^2$.

[2]It is because $\cos(90°) = 0$.



Definition.

The vector projection of a vector $\mathbf{v}$ on (or onto) a nonzero vector $\mathbf{u}$ is the orthogonal projection of a onto a straight line parallel to $\mathbf{u}$. It is a vector parallel to $\mathbf{u}$, denoted as Proj$\mathbf{u}(\mathbf{v})$. ■



Result.

$$\text{Proj}\mathbf{u}(\mathbf{v}) = \left(\frac{\mathbf{v}\cdot\mathbf{u}}{\|\mathbf{u}\|^2}\right)\mathbf{u}. \tag{1-3}$$

Proof.

Let $\text{Proj}\mathbf{u}(\mathbf{v}) = t\mathbf{u}$. Because $(\text{Proj}\mathbf{u}(\mathbf{v}) - \mathbf{v})$ is orthogonal to $\mathbf{u}$,[3]

$$(\text{Proj}\mathbf{u}(\mathbf{v}) - \mathbf{v}) \cdot \mathbf{u} = 0.$$

That is

$$(t\mathbf{u} - \mathbf{v}) \cdot \mathbf{u} = 0.$$

Hence $t = \left(\frac{\mathbf{v}\cdot\mathbf{u}}{\|\mathbf{u}\|^2}\right)$, and therefore $\text{Proj}\mathbf{u}(\mathbf{v}) = t\mathbf{u} = \left(\frac{\mathbf{v}\cdot\mathbf{u}}{\|\mathbf{u}\|^2}\right)\mathbf{u}$. ■

[3] In fact, it is $-(\text{Proj}\mathbf{u}(\mathbf{v}) - \mathbf{v})$ to be orthogonal to $\mathbf{u}$.

2 Matrices

2.1 Linear Transformation

Linear mappings from one vector space to another play an important role in mathematics. In the study of vector spaces the most important types of mappings are linear transformation.

Definition. (Linear Transformation)
Let $\mathcal{U}$ and $\mathcal{V}$ be two vector spaces. A function $\mathcal{L}: \mathcal{U} \to \mathcal{V}$ is called a linear transformation if

$$\mathcal{L}(\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2) = \alpha_1 \mathcal{L}(\mathbf{u}_1) + \alpha_2 \mathcal{L}(\mathbf{u}_2) \tag{1-4}$$

for all $\mathbf{u}_1$ and $\mathbf{u}_2$ in $\mathcal{U}$ and any scalars α_1 and α_2. ■

If $\mathcal{L}$ is a linear transformation mapping a vector space $\mathcal{U}$ into $\mathcal{V}$, it follows from (1-3) that

$$\mathcal{L}(\mathbf{u}_1 + \mathbf{u}_2) = \mathcal{L}(\mathbf{u}_1) + \mathcal{L}(\mathbf{u}_2), \quad (\because \alpha_1 = \alpha_2 = 1) \tag{1-5}$$

and

$$\mathcal{L}(\alpha_1 \mathbf{u}) = \alpha_1 \mathcal{L}(\mathbf{u}), \quad (\because \mathbf{u} = \mathbf{u}_1,\ \alpha_2 = 0). \tag{1-6}$$

Conversely, if $\mathcal{L}$ satisfies (4) and (5), then

$$\begin{aligned} \mathcal{L}(\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2) &= \mathcal{L}(\alpha_1 \mathbf{u}_1) + \mathcal{L}(\alpha_2 \mathbf{u}_2) \\ &= \alpha_1 \mathcal{L}(\mathbf{u}_1) + \alpha_2 \mathcal{L}(\mathbf{u}_2). \end{aligned}$$

Thus $\mathcal{L}$ is a linear transformation on $\mathcal{U}$ if and only if $\mathcal{L}$ satisfies (1-4) and (1-5).

Example.
The operator $\mathcal{L}$ defined by

$$\mathcal{L}(\mathbf{x}) = \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix}$$

is linear, since

$$\begin{aligned}\mathcal{L}(\alpha\mathbf{x}+\beta\mathbf{y}) &= \begin{pmatrix} -(\alpha x_2+\beta y_2) \\ \alpha x_1+\beta y_2 \end{pmatrix} \\ &= \alpha\begin{pmatrix} -x_2 \\ x_1 \end{pmatrix} + \beta\begin{pmatrix} -y_2 \\ y_1 \end{pmatrix} \\ &= \alpha\mathcal{L}(\mathbf{x})+\beta\mathcal{L}(\mathbf{y}).\end{aligned}$$

The operator $\mathcal{L}$ has the effect of rotating each vector in $\mathbb{R}^2$ by $90°$ in the counterclockwise direction. ■

Example.

Let $\mathcal{T}:\mathbb{R}^2\to\mathbb{R}^1$ be defined as

$$\mathcal{T}(\mathbf{x}) = \mathcal{T}(x_1,x_2) = (x_1^2+x_2^2)^{1/2}.$$

Then $\mathcal{T}$ is **not** a linear transformation, since

$$\mathcal{T}(\alpha\mathbf{x}) = (\alpha^2x_1^2+\alpha^2x_2^2)^{1/2} = |\alpha|\mathcal{T}(\mathbf{x}).$$

It follows that

$$\alpha\mathcal{T}(\mathbf{x}) \neq \mathcal{T}(\alpha\mathbf{x})$$

whenever $\alpha<0$ and $\mathbf{x}\neq\mathbf{0}$. Therefore, $\mathcal{T}$ is not a linear transformation. ■

2.1.1 The Matrix Representation of a Linear Transformation

We will see how any linear operator between finite-dimensional space can be represented by a matrix now.

Theorem.

If $\mathcal{L}$ is a linear operator mapping $\mathbb{R}^n$ into $\mathbb{R}^m$, then there is an $m\times n$ matrix $\mathbf{A}$ such that

$$\mathcal{L}(\mathbf{x}) = \mathbf{A}\mathbf{x}$$

for each $\mathbf{x} \in \mathbb{R}^n$. In fact, the jth column vector of $\mathbf{A}$ is given by

$$\mathbf{a}_j = \mathcal{L}(\mathbf{e}_j) \quad j = 1, 2, ..., n.$$

Proof.
For $j = 1, ..., n$ define

$$\mathbf{a}_j = (a_{1j}, a_{2j}, ..., a_{mj})' = \mathcal{L}(\mathbf{e}_j).$$

Let

$$\mathbf{A} = (a_{ij}) = (\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n).$$

If

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n$$

is an arbitrary element of $\mathbb{R}^n$, then

$$\begin{aligned}
\mathcal{L}(\mathbf{x}) &= x_1\mathcal{L}(\mathbf{e}_1) + x_2\mathcal{L}(\mathbf{e}_2) + \cdots + x_n\mathcal{L}(\mathbf{e}_n) \\
&= x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n \\
&= (\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n)\begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix} \\
&= \mathbf{A}\mathbf{x}.
\end{aligned}$$

■

Since the standard basis elements $\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n$ were used for $\mathbb{R}^n$, we refer $\mathbf{A}$ as the *standard matrix representation* of $\mathcal{L}$.

Example.
Let $\mathcal{L} : \mathbb{R}^3 \to \mathbb{R}^3$ be defined as

$$\mathcal{L}(x_1, x_2, x_3) = (x_1 - x_2, x_1 + x_3, x_3).$$

Then $\mathcal{L}$ is a linear transformation, since

$$\begin{aligned}
&\mathcal{L}[\alpha(x_1,x_2,x_3)+\beta(y_1,y_2,y_3)]\\
&=\mathcal{L}(\alpha x_1+\beta y_1,\alpha x_2+\beta y_2,\alpha x_3+\beta y_3)\\
&=(\alpha x_1+\beta y_1-\alpha x_2-\beta y_2,\alpha x_1+\beta y_1+\alpha x_3+\beta y_3,\alpha x_3+\beta y_3)\\
&=\alpha(x_1-x_2,x_1+x_3,x_3)+\beta(y_1-y_2,y_1+y_3,y_3)\\
&=\alpha\mathcal{L}(x_1,x_2,x_3)+\beta\mathcal{L}(y_1,y_2,y_3).
\end{aligned}$$

We wish to find a matrix $\mathbf{A}$ such that $\mathcal{L}(\mathbf{x})=\mathbf{A}\mathbf{x}$ for each $\mathbf{x}\in\mathbb{R}^3$. To do this, one must determine $\mathcal{L}(\mathbf{e}_1)$, $\mathcal{L}(\mathbf{e}_2)$, and $\mathcal{L}(\mathbf{e}_3)$.

$$\mathcal{L}(\mathbf{e}_1)=\mathcal{L}(1,0,0)'=\begin{pmatrix}1\\1\\0\end{pmatrix}$$

$$\mathcal{L}(\mathbf{e}_2)=\mathcal{L}(0,1,0)'=\begin{pmatrix}-1\\0\\0\end{pmatrix}$$

$$\mathcal{L}(\mathbf{e}_3)=\mathcal{L}(0,0,1)'=\begin{pmatrix}0\\1\\1\end{pmatrix}.$$

We choose these vectors to be the columns of $\mathbf{A}$,

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

then

$$\mathcal{L}(\mathbf{x})=\begin{bmatrix}x_1-x_2\\x_1+x_3\\x_3\end{bmatrix}=\mathbf{A}\mathbf{x}$$

for each $\mathbf{x}\in\mathbb{R}^3$. ■

In general, let $\mathcal{T}:\ \mathcal{U}\to\mathcal{V}$ be a linear transformation, where $\mathcal{U}$ and $\mathcal{V}$ are vector spaces of dimensions n and m, respectively. Let $\mathbf{u}_1$, $\mathbf{u}_2$,...,$\mathbf{u}_n$ be a basis for $\mathcal{U}$ and $\mathbf{v}_1$, $\mathbf{v}_2$,...,$\mathbf{v}_m$ be a basis for $\mathcal{V}$. For $j=1,2,...,n$, consider $\mathcal{T}(\mathbf{u}_j)$, which can be uniquely represented as

$$\mathcal{T}(\mathbf{u}_j)=\sum_{i=1}^m a_{ij}\mathbf{v}_i,\quad j=1,2,...,n,$$

where the a_{ij}'s are scalars. These scalars completely determine all possible value of $\mathcal{T}$: If $\mathbf{u} \in \mathcal{U}$, then $\mathbf{u} = \sum_{j=1}^{n} x_j \mathbf{u}_j$, for some scalars $x_1, x_2, ..., x_n$. Then

$$\begin{aligned}
\mathcal{T}(\mathbf{u}) &= \sum_{j=1}^{n} x_j \mathcal{T}(\mathbf{u}_j) \quad (\textit{since } \mathcal{T} \textit{ is a linear transformation}) \\
&= \sum_{j=1}^{n} x_j \left(\sum_{i=1}^{m} a_{ij} \mathbf{v}_i \right) \\
&= \sum_{j=1}^{n} x_j \sum_{i=1}^{m} a_{ij} \mathbf{v}_i \\
&= \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_j \mathbf{v}_i \quad (\textit{interchange summation}) \qquad (1\text{-}7)
\end{aligned}$$

By definition, the rectangular array

$$\mathbf{A}_{m\times n} = \begin{bmatrix} a_{11} & a_{12} & . & . & . & a_{1n} \\ a_{21} & a_{22} & . & . & . & a_{2n} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ a_{m1} & a_{m2} & . & . & . & a_{mn} \end{bmatrix},$$

is called a matrix of order $m \times n$, which indicates that $\mathbf{A}$ has m rows and n columns. We could further write (1-6) as

$$\mathcal{T}(\mathbf{u}) = \sum_{i=1}^{m} y_i \mathbf{v}_i, \qquad (1\text{-}8)$$

where

$$y_i = \sum_{j=1}^{n} a_{ij} x_j, \quad i = 1, 2, ..., m. \qquad (1\text{-}9)$$

Therefore, if $\mathbf{A}$ is the $m \times n$ matrix with element a_{ij}, $\mathbf{x}$ and $\mathbf{y}$ are the n- and m-vectors with component $x_1, ..., x_n$ and $y_1, ..., y_m$, then (1-8) is equivalent to

$$\mathbf{y} = \mathbf{A}\mathbf{x}. \qquad (1\text{-}10)$$

It is therefore noted that an $m \times n$ matrix $\mathbf{A}$ define a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$, and $\mathbf{A}$ is called the **matrix representation of the $\mathcal{T}$** with respect to the particular bases. Conversely, any linear transformation between finite dimensional space can be represented by a matrix that depends on a choice of basis for the two spaces.

This is an underlying reason why matrix-vector multiplication is defined the way it is below.

2.2 Some Terminology

A matrix is a rectangular array of numbers, denoted

$$\mathbf{A}_{i\times k} = [a_{ik}] = \begin{bmatrix} a_{11} & a_{12} & . & . & . & a_{1k} \\ a_{21} & a_{22} & . & . & . & a_{2k} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ a_{i1} & a_{i2} & . & . & . & a_{ik} \end{bmatrix},$$

where a subscribed element of a matrix is always read as $a_{row,column}$. Here we confine the element to be real number.

A **vector** is a matrix with one row or one column. Therefore a row vector is $\mathbf{A}_{1\times k}$ and a column vector is $\mathbf{A}_{i\times 1}$ and commonly denoted as $\mathbf{a}^k$ and $\mathbf{a}_i$, respectively. In the followings of this course, we follow conventional custom to say that a *vector* is a *column vector* except for particular mention.

The dimension of a matrix is the numbers of rows and columns it contained. If i equals to k, then $\mathbf{A}$ is a square matrix. Several particular types of square matrices occur in econometrics:

(a). A symmetric matrix $\mathbf{A}$ is one in which $a_{ik} = a_{ki}$ for all i and k.

(b). A diagonal matrix is a square matrix whose nonzero elements appears on the main diagonal, moving from upper left to lower right.

(c). A scalar matrix is a diagonal matrix with the same values in all diagonal elements.

(d). An identity matrix is a scalar matrix with ones on the diagonal. This matrix is always denoted as $\mathbf{I}$. A subscript is sometimes included to indicate its size. for example,

$$\mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

(e). A triangular matrix is one that has only zeros either above or below the main diagonal. For example,

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 3 & 0 \\ 4 & 5 & 6 \end{bmatrix}.$$

2.3 Algebraic Manipulation of Matrices

2.3.1 Equality of Matrices

Matrices $\mathbf{A}$ and $\mathbf{B}$ are equal if and only if they have the same dimensions and each element of $\mathbf{A}$ equal the corresponding element of $\mathbf{B}$.

$$\mathbf{A}=\mathbf{B} \text{ if and only if } a_{ik} = b_{ik} \text{ for all } i \text{ and } k.$$

2.3.2 Transposition

The transpose of a matrix $\mathbf{A}$, denoted as $\mathbf{A}'$, is obtained by creating the matrix whose kth row is the kth column of the original matrix. If $\mathbf{A}$ is $i \times k$, then $\mathbf{A}'$ is $k \times i$. For example,

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 5 & 1 & 5 \\ 6 & 4 & 5 \\ 3 & 1 & 4 \end{bmatrix}, \quad \text{then} \qquad \mathbf{A}' = \begin{bmatrix} 1 & 5 & 6 & 3 \\ 2 & 1 & 4 & 1 \\ 3 & 5 & 5 & 4 \end{bmatrix}.$$

If $\mathbf{A}$ is symmetric, $\mathbf{A}=\mathbf{A}'$. It is also apparent that for any matrix $\mathbf{A}$, $(\mathbf{A}')' = \mathbf{A}$. Finally, the transpose of a column vector, $\mathbf{a}_i$ is a row vector:

$$\mathbf{a}_i' = \begin{bmatrix} a_1 & a_2 & . & . & . & a_i \end{bmatrix}.$$

2.3.3 Matrix Addition

Matrices cannot be added unless they have the same dimension. The operation of addition is extended to matrices by defining

$$\mathbf{C} = \mathbf{A} + \mathbf{B} = [a_{ik} + b_{ik}].$$

We also extend the operation of subtraction to matrices precisely as if they were scalars by performing the operation element by element. Thus,

$$\mathbf{C} = \mathbf{A} - \mathbf{B} = [a_{ik} - b_{ik}].$$

It follows that

(a). matrix addition is commutative,

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A},$$

(b). and associative,

$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C}),$$

(c). and that

$$(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'. \qquad \square$$

2.3.4 Matrix Multiplication

Matrices are multiplied by using the dot product.

Definition.

For an $n \times k$ matrix $\mathbf{A}$ and a $k \times T$ matrix $\mathbf{B}$, the product matrix,

$$\mathbf{C} = \mathbf{AB}$$

is an $n \times T$ matrix whose ikth element is the dot product of row i of $\mathbf{A}$ and column k of $\mathbf{B}$, i.e.

$$\mathbf{C} = [c_{nT}], \quad c_{ik} = \mathbf{a}^i \cdot \mathbf{b}_k.$$

Generally, $\mathbf{AB} \neq \mathbf{BA}$. ■

The product of a matrix and a vector is a vector and is written as

$$\begin{aligned} \mathbf{c} &= \mathbf{Ab} \\ &= b_1\mathbf{a}_1 + b_2\mathbf{a}_2 + ... + b_k\mathbf{a}_k, \end{aligned}$$

where b_i is ith element of vector $\mathbf{b}$ and $\mathbf{a}_i$ is ith column of matrix $\mathbf{A}$. Here we see that the right-hand side is a **linear combination** of the columns of the matrix where the coefficients are the elements of the vector.

In the calculation of a matrix product $\mathbf{C} = \mathbf{A}_{n\times k}\mathbf{B}_{k\times T}$, it can be written as

$$\begin{aligned} \mathbf{C} &= \mathbf{AB} \\ &= [\mathbf{Ab}_1 \ \ \mathbf{Ab}_2 \ \cdots \ \mathbf{Ab}_T], \end{aligned}$$

where $\mathbf{b}_i$ is ith column of matrix $\mathbf{B}$.

Some general rules for matrix multiplication are as follows:

(a). Associate law: $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$.

(b). Distributive law: $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$.

(c). Transpose of a product: $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$.

(d). Scalar multiplication: $\alpha\mathbf{A} = [\alpha a_{ik}]$ for a scalar α. □

2.3.5 Matrix Inversion

To solve the system $\mathbf{Ax} = \mathbf{b}$ for $\mathbf{x}$, something akin to division by a matrix is needed.

$\boxed{\text{Definition.}}$

A *square* matrix $\mathbf{A}$ is said to be nonsingular or invertible if there exist a unique matrix (square) $\mathbf{B}$ such that $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$. The matrix $\mathbf{B}$ is said to be a multiplicative inverse of $\mathbf{A}$. We will refer to the *multiplicative inverse* of a nonsingular matrix $\mathbf{A}$ as simply the inverse of $\mathbf{A}$ and denote it by $\mathbf{A}^{-1}$.[4] ■

Some computational results involving inverse are

$$\begin{aligned}
|\mathbf{A}^{-1}| &= \frac{1}{|\mathbf{A}|}, \\
(\mathbf{A}^{-1})^{-1} &= \mathbf{A}, \\
(\mathbf{A}^{-1})' &= (\mathbf{A}')^{-1} \\
(\mathbf{AB})^{-1} &= \mathbf{B}^{-1}\mathbf{A}^{-1}, \\
(\mathbf{A}+\mathbf{B})^{-1} &= \mathbf{A}^{-1} - \mathbf{A}^{-1}(\mathbf{B}^{-1}+\mathbf{A}^{-1})^{-1}\mathbf{A}^{-1}.
\end{aligned}$$

when both inverse matrices exist. Finally, if $\mathbf{A}$ is symmetric, then $\mathbf{A}^{-1}$ is also symmetric.

2.4 An Useful Idempotent Matrix

A fundamental matrix in statistics is the one that is used to transform data to deviations from their mean.

Definition.
An idempotent matrix is the one that is equal to its square, that is $\mathbf{M}^2 = \mathbf{MM} = \mathbf{M}$. ■

An useful idempotent matrix we will often face is the matrix

$$\mathbf{M}_i = \mathbf{I} - \mathbf{i}(\mathbf{i}'\mathbf{i})^{-1}\mathbf{i}' = \mathbf{I} - \frac{1}{n}\mathbf{i}\mathbf{i}',$$

where $\mathbf{i}$ is a column of ones's $(n \times 1)$ vector.

[4] For a full column rank $m \times n$ matrix $\boldsymbol{U}$ $(m > n)$, if there exists an $n \times m$ (here m can be equal to n) matrix, $\boldsymbol{X}$, satisfying the following conditions: (a) $\boldsymbol{UXU} = \boldsymbol{U}$, (b) $\boldsymbol{XUX} = \boldsymbol{X}$, (c) $(\boldsymbol{UX})' = \boldsymbol{UX}$, and (d) $(\boldsymbol{XU})' = \boldsymbol{XU}$, then $\boldsymbol{X}$ is called the Moore-Penrose inverse of $\boldsymbol{U}$, denoted as $\boldsymbol{U}^+$. It is well-known that $\boldsymbol{U}^+$ exists and is unique for any $m \times n$ matrix $\boldsymbol{U}$.

Result.

$$\mathbf{M}_i\mathbf{x} = \begin{bmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \cdot \\ \cdot \\ \cdot \\ x_n - \bar{x} \end{bmatrix},$$

where $\mathbf{x} = [x_1, x_2, ..., x_n]'$ and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. It is easily seen that $\mathbf{M}_i$ is a symmetric and idempotent matrix.

Proof.
As definition,

$$\mathbf{M}_i\mathbf{x} = \left(\mathbf{I} - \frac{1}{n}\mathbf{i}\mathbf{i}'\right)\mathbf{x} = \mathbf{x} - \mathbf{i}\frac{1}{n}\mathbf{i}'\mathbf{x} = \mathbf{x} - \mathbf{i}\bar{x}.$$ ■

Exercise 2.
Using the data I give to you, compute $\sum_{i=1}^{N}(X - \bar{X})^2$, where $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ from the idempotent matrix $\mathbf{M}_i$ such that $\sum_{i=1}^{N}(X - \bar{X})^2 = \mathbf{x}'\mathbf{M}_i\mathbf{x}$. ■

2.5 Trace of Matrix

Definition.
The trace of a square $k \times k$ matrix is the sums of its diagonal elements:

$$tr(\mathbf{A}) = \sum_{i=1}^{k} a_{ii}.$$ ■

Some useful results about trace are:
(a).

$$tr(c) = c \quad \text{for a constant } c,$$

(b).

$$tr(c\mathbf{A}) = c(tr(\mathbf{A})),$$

(c).

$$tr(\mathbf{A}) = tr(\mathbf{A}'),$$

(d).

$$tr(\mathbf{A} + \mathbf{B}) = tr(\mathbf{B}) + tr(\mathbf{A}),$$

(e).

$$tr(\mathbf{ABCD}) = tr(\mathbf{BCDA}) = tr(\mathbf{CDAB}) = tr(\mathbf{DABC}).$$

Exercise 3.

A matrix $\mathbf{D}$ is *skew-symmetric* if $\mathbf{D}' = -\mathbf{D}$. Now if $\mathbf{A}$ is a symmetric $n \times n$ matrix, and $\mathbf{B}$ is an $n \times n$ skew-symmetric matrix. Find $tr(\mathbf{AB})$. ■

2.6 The Nullspace of a Matrix

Definition.

Let $\mathbf{A}$ be an $m \times n$ matrix. Let $N(\mathbf{A})$ denote the set of all solutions to the homogeneous system $\mathbf{Ax} = \mathbf{0}$. Thus

$$N(\mathbf{A}) = \{\mathbf{x} \in \mathbb{R}^n | \mathbf{Ax} = \mathbf{0}\}.$$

This set of all solutions forms a subspace of $\mathbb{R}^n$ and is called the null space of $\mathbf{A}$. ■

Example.

Determine $N(\mathbf{A})$ if

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

One solution to this system is

$$\begin{aligned} x_1 &= x_3 - x_4 \\ x_2 &= -2x_3 + x_4. \end{aligned}$$

Thus, if we set $x_3 = \alpha$ and $x_4 = \beta$, then

$$\mathbf{x} = \begin{bmatrix} \alpha - \beta \\ -2\alpha + \beta \\ \alpha \\ \beta \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} -1 \\ 1 \\ 0 \\ 1 \end{bmatrix}.$$

The vector space $N(\mathbf{A})$ consists of all vector of the form

$$\alpha \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} -1 \\ 1 \\ 0 \\ 1 \end{bmatrix},$$

which is a two-dimensional subspace in $\mathbb{R}^4$. ■

Definition.

The dimension of the null space of a matrix is called the **nullity** of the matrix. ■

2.7 Rank of a Matrix

If $\mathbf{A}$ is an $m \times n$ matrix, each row of $\mathbf{A}$ is an $n-$tuple of real numbers and hence can be considered as a vector in $\mathbb{R}^{1\times n}$. The m vectors corresponding to the rows of $\mathbf{A}$ will be referred to as the row vectors of $\mathbf{A}$. Similarly, each column of $\mathbf{A}$ can be considered as a vector in $\mathbb{R}^m$ and one can associate n column vectors with the matrix $\mathbf{A}$.

Definition.

If $\mathbf{A}$ is an $m \times n$ matrix, the subspace of $\mathbb{R}^{1\times n}$ spanned by the row vectors of $\mathbf{A}$ is called the *row space* of $\mathbf{A}$. The subspace of $\mathbb{R}^m$ spanned by the column vectors of $\mathbf{A}$ is called the *column space* of $\mathbf{A}$. ■

Example.

Let

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

The row space of $\mathbf{A}$ is the set of all 3-tuples of the form

$$\alpha(1\ \ 0\ \ 0) + \beta(0\ \ 1\ \ 0) = (\alpha\ \ \beta\ \ 0).$$

The column space of $\mathbf{A}$ is the set of all vectors of the form

$$\alpha \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \gamma \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Thus the row space of $\mathbf{A}$ is a two-dimensional subspace of $\mathbb{R}^{1\times 3}$ and the column space of $\mathbf{A}$ is $\mathbb{R}^2$. ■

Theorem.

The column space and the row space of a matrix have the same dimension. ■

Definition.

The column(row) rank of a matrix is the dimension of the vector space that is spanned by its columns (rows). In short from this definition we know that the column rank is the number of *linearly independent column* of a matrix. ■

Theorem.

The column rank and row rank of a matrix are equal, that is

rank($\mathbf{A}$)=rank($\mathbf{A}$')$\leq$ min(number of rows, numbers of columns). ■

Definition.

A full (short) rank matrix is a matrix whose rank is equal (fewer) to the number of columns it contains. ■

Theorem.
If $\mathbf{A}$ is an $m \times n$ matrix, then the rank of $\mathbf{A}$ plus the nullity of $\mathbf{A}$ equals n.

Example.
Let

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

The rank of $\mathbf{A} = 2$ and nullity of $\mathbf{A} = 2$. Therefore $2 + 2 = 4$. ■

Theorem.
Let $\mathbf{A}$ be a $m \times n$ matrix. Then $N(\mathbf{A'A}) = N(\mathbf{A})$ and $rank(\mathbf{A'A}) = rank(\mathbf{A})$.

Proof.
We first prove that $N(\mathbf{A'A}) = N(\mathbf{A})$. If $\mathbf{Ax} = \mathbf{0}$, then $\mathbf{A'Ax} = \mathbf{0}$, so $N(\mathbf{A}) \subset N(\mathbf{A'A})$. If $\mathbf{A'Ax} = \mathbf{0}$, then $\mathbf{0} = \mathbf{x'A'Ax} = (\mathbf{Ax})'\mathbf{Ax}$, so that $\mathbf{Ax} = \mathbf{0}$, i.e. $N(\mathbf{A'A}) \subset N(\mathbf{A})$. Thus $N(\mathbf{A'A}) = N(\mathbf{A})$.

From last theorem,

$$rank(\mathbf{A'A}) = n - nullity(\mathbf{A'A}) = n - nullity(\mathbf{A}) = rank(\mathbf{A}). \quad ■$$

Corollary.
$rank(\mathbf{A'A}) = rank(\mathbf{A}) = rank(\mathbf{A'}) = rank(\mathbf{AA'})$. ■

Theorem.
In a product matrix $\mathbf{C} = \mathbf{AB}$, then

$$rank(\mathbf{C}) = rank(\mathbf{AB}) \leq \min(rank(\mathbf{A}), rank(\mathbf{B})).$$

Proof.
For $\mathbf{A}$ is $m \times n$ and $\mathbf{B}$ is $n \times k$. Write $\mathbf{AB} = [\mathbf{Ab}_1 \ \ \mathbf{Ab}_2 ..... \ \mathbf{Ab}_k]$, where $\mathbf{b}_i$ are the ith

column of **B**. That is, each column of **AB** can be expressed as a linear combination of the column of **A**, so the number of linearly independent columns in **AB** can not be more than the number of linearly independent columns in **A**. Thus, rank(**AB**) $\leq$ rank(**A**). Similarly, each row of **AB** can be expressed as a linear combination of the rows of **B** from which we get rank(**AB**) $\leq$ rank(**B**). ■

Corollary.

If **A** is $m \times n$ and **B** is a square matrix of rank n, then rank(**AB**)=rank(**A**).

Proof.

For any two matrix,

$$rank(\mathbf{AB}) \leq rank(\mathbf{A}).$$

If **B** is nonsingular, then

$$rank(\mathbf{A}) = rank(\mathbf{ABB^{-1}}) \leq rank(\mathbf{AB}).$$

Hence rank(**AB**)=rank(**A**). ■

Theorem.

Let **A** is $m \times n$ matrix, **B** is $m \times m$ matrix, and **C** is $n \times n$ matrix. Then if **B** and **C** are nonsingular matrices, it follows that

$$rank(\mathbf{BAC}) = rank(\mathbf{BA}) = rank(\mathbf{A}). \tag{1-11}$$

Proof.

By last corollary, $rank(\mathbf{BAC}) = rank[(\mathbf{BA})\mathbf{C}] = rank(\mathbf{BA})$. Since $rank(\mathbf{BA}) = rank(\mathbf{A'B'}) = rank(\mathbf{A'}) = rank(\mathbf{A})$, the result is obtained. ■

2.8 Determinant

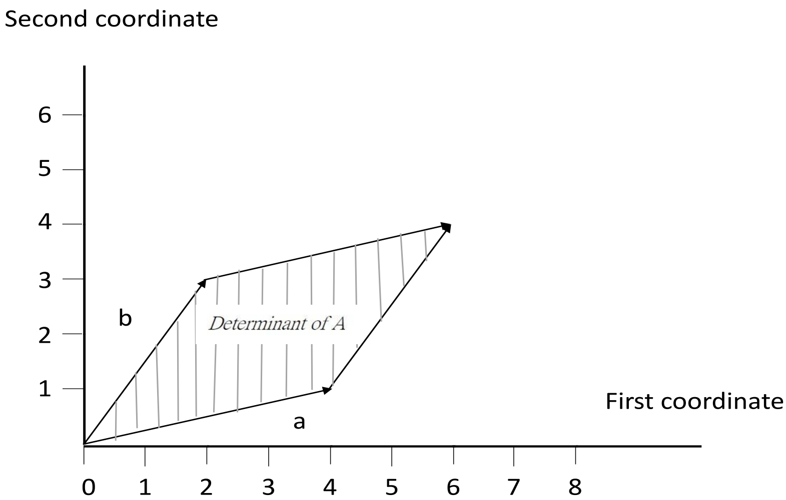
With each square matrix it is possible to associate a real number called the *determinant* of the matrix. The value of this number will tell us whether or not the matrix is singular.

Example.

The vectors of the matrix is

$$\begin{aligned}\mathbf{A} &= [\mathbf{a},\ \mathbf{b}] \\ &= \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix}.\end{aligned}$$

The area of the parallelogram formed by the columns of **A** can be obtained by manipulating congruent triangles. The result is $4(3) - 1(2) = 10$. The (absolute) area is the *determinant* of **A**, denoted as $|\mathbf{A}|$.



■

If the columns of **A** were linearly dependent, then the two vectors would lie on the same line. The "parallelogram" would collapse to a line and would have zero area. This concept implies that if the columns of a (square) matrix are linear dependent, then its determinant is zero.

Theorem.

The determinant of a matrix is nonzero if and only if it has full rank. ■

results.

(a).

$$|c\mathbf{D}| = c^k|\mathbf{D}|,\ for\ a\ constant\ c,\ and\ k \times k\ matrix\ D.$$

(b).

$$|\mathbf{CD}| = |\mathbf{C}| \cdot |\mathbf{D}|\ for\ two\ matrices\ C\ and\ D.$$

(c).

$$|\mathbf{C}| = |\mathbf{C}'|.$$

■

3 Partitioned Matrices

Often it is useful to think of a matrix as being composed of a number of submatrices. A matrix $\mathbf{A}$ can be partitioned into smaller matrices by drawing horizontal lines between the rows and vertical lines between the columns. For example, we might write

$$\mathbf{A}_{m\times n}=\begin{bmatrix}\mathbf{A}_{11} & \mathbf{A}_{12}\\ \mathbf{A}_{21} & \mathbf{A}_{22}\end{bmatrix},\text{ or }\mathbf{A}=\begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n\end{bmatrix},\text{ or }\mathbf{A}=\begin{bmatrix}\mathbf{a}^1\\ \mathbf{a}^2\\ \vdots\\ \mathbf{a}^m\end{bmatrix},$$

and note that

$$\mathbf{A}'=\begin{bmatrix}\mathbf{A}'_{11} & \mathbf{A}'_{21}\\ \mathbf{A}'_{12} & \mathbf{A}'_{22}\end{bmatrix}.$$

A common special case is the block diagonal matrix:

$$\mathbf{A}=\begin{bmatrix}\mathbf{A}_{11} & \mathbf{0}\\ \mathbf{0} & \mathbf{A}_{22}\end{bmatrix},$$

where $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ are square matrices.

3.1 Addition and Multiplication of Partitioned Matrices

For conformably partitioned matrices $\mathbf{A}_{ij}$ and $\mathbf{B}_{ij}$,

$$\mathbf{A}+\mathbf{B}=\begin{bmatrix}\mathbf{A}_{11}+\mathbf{B}_{11} & \mathbf{A}_{12}+\mathbf{B}_{12}\\ \mathbf{A}_{21}+\mathbf{B}_{21} & \mathbf{A}_{22}+\mathbf{B}_{22}\end{bmatrix}.$$

That is, for addition, the dimension of $\mathbf{A}_{ij}$ and $\mathbf{B}_{ij}$ must be the same. However, for

$$\begin{aligned}\mathbf{AB} &= \begin{bmatrix}\mathbf{A}_{11} & \mathbf{A}_{12}\\ \mathbf{A}_{21} & \mathbf{A}_{22}\end{bmatrix}\times\begin{bmatrix}\mathbf{B}_{11} & \mathbf{B}_{12}\\ \mathbf{B}_{21} & \mathbf{B}_{22}\end{bmatrix}\\ &= \begin{bmatrix}\mathbf{A}_{11}\mathbf{B}_{11}+\mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12}+\mathbf{A}_{12}\mathbf{B}_{22}\\ \mathbf{A}_{21}\mathbf{B}_{11}+\mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12}+\mathbf{A}_{22}\mathbf{B}_{22}\end{bmatrix},\end{aligned}$$

the number of columns in $\mathbf{A}_{ij}$ must equal the number of rows in $\mathbf{B}_{jk}$ for all pairs i and j.

Example.

Recall the calculation of the product of a matrix and a vector

$$\begin{aligned}
\mathbf{c} &= \mathbf{A}_{n\times k}\mathbf{b}_{k\times 1} \\
&= \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & . & . & . & \mathbf{a}_k \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ . \\ . \\ . \\ b_k \end{bmatrix} \\
&= b_1\mathbf{a}_1 + b_2\mathbf{a}_2 + ... + b_k\mathbf{a}_k,
\end{aligned}$$

and that of a matrix product $\mathbf{C} = \mathbf{A}_{n\times k}\mathbf{B}_{k\times l}$

$$\begin{aligned}
\mathbf{C} &= \mathbf{AB} \\
&= \mathbf{A}_{n\times k}\begin{bmatrix} (\mathbf{b}_1)_{k\times 1} & (\mathbf{b}_2)_{k\times 1} & . & . & . & (\mathbf{b}_l)_{k\times 1} \end{bmatrix} \\
&= [\mathbf{Ab}_1 \;\; \mathbf{Ab}_2 \cdots \mathbf{Ab}_l] \\
&= \begin{bmatrix} (\mathbf{a}^1)_{1\times k} \\ (\mathbf{a}^2)_{1\times k} \\ . \\ . \\ . \\ (\mathbf{a}^n)_{1\times k} \end{bmatrix} \mathbf{B}_{k\times l} \\
&= \begin{bmatrix} \mathbf{a}^1\mathbf{B} \\ \mathbf{a}^2\mathbf{B} \\ . \\ . \\ . \\ \mathbf{a}^n\mathbf{B} \end{bmatrix}.
\end{aligned}$$

■

Two cases frequently encountered are of the form

$$\begin{aligned}
\begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix}' \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix} &= \begin{bmatrix} \mathbf{A}_1' & \mathbf{A}_2' \end{bmatrix} \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{A}_1'\mathbf{A}_1 + \mathbf{A}_2'\mathbf{A}_2 \end{bmatrix},
\end{aligned}$$

and

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix}' \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11}'\mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22}'\mathbf{A}_{22} \end{bmatrix}.$$

3.2 Determinants of Partitioned Matrices

Results.

(a).

$$\begin{vmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{vmatrix} = |\mathbf{A}_{11}| \cdot |\mathbf{A}_{22}|.$$

(b).

$$\begin{aligned} \begin{vmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{vmatrix} &= |\mathbf{A}_{22}| \cdot |\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}| \\ &= |\mathbf{A}_{11}| \cdot |\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}|. \end{aligned}$$

■

3.3 Inverses of Partitioned Matrices

Results.

(a). The inverse of a block diagonal matrix is

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22}^{-1} \end{bmatrix}.$$

(b). For a general 2×2 partitioned matrix,

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}_{11}^{-1}(\mathbf{I} + \mathbf{A}_{12}\mathbf{F}_2\mathbf{A}_{21}\mathbf{A}_{11}^{-1}) & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{F}_2 \\ -\mathbf{F}_2\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{F}_2 \end{bmatrix},$$

where $\mathbf{F}_2 = (\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}$. ■

Exercise 4.

Show that

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}^{-1}$$

is also

$$\begin{bmatrix} \mathbf{F}_1 & -\mathbf{F}_1\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{F}_1 & \mathbf{A}_{22}^{-1}(\mathbf{I}+\mathbf{A}_{21}\mathbf{F}_1\mathbf{A}_{12}\mathbf{A}_{22}^{-1}) \end{bmatrix},$$

where $\mathbf{F}_1 = (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1}$. ■

3.4 Kronecker Products

Definition.

For general matrices $\mathbf{A}$ and $\mathbf{B}$, the Kronecker products of them is

$$\mathbf{A}\otimes\mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & . & . & . & a_{1k}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & . & . & . & a_{2k}\mathbf{B} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ a_{i1}\mathbf{B} & a_{i2}\mathbf{B} & . & . & . & a_{ik}\mathbf{B} \end{bmatrix}.$$ ■

Notice that if $\mathbf{A}$ is $i \times k$ and $\mathbf{B}$ is $m \times n$, then $\mathbf{A}\otimes\mathbf{B}$ is $(im)\times(kn)$.

Results.

Let $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ be any matrices, then

(1).

$$(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = (\mathbf{AC})\otimes(\mathbf{BD}),$$

(2).

$$(\mathbf{A}+\mathbf{B})\otimes\mathbf{C} = (\mathbf{A}\otimes\mathbf{C}) + (\mathbf{B}\otimes\mathbf{C}),$$

(3).

$$(\mathbf{A}\otimes\mathbf{B})' = \mathbf{A}'\otimes\mathbf{B}'.$$ ■

Results.

Let $\mathbf{A}$ be $m \times m$ and $\mathbf{B}$ be $k \times k$ nonsingular matrices, then

(1).

$$(\mathbf{A} \otimes \mathbf{B})^{-\mathbf{1}} = \mathbf{A}^{-\mathbf{1}} \otimes \mathbf{B}^{-\mathbf{1}}.$$

(2).

$$tr(\mathbf{A} \otimes \mathbf{B}) = tr(\mathbf{A})tr(\mathbf{B}).$$

Proof.

(1).

$$(\mathbf{A}^{-1} \otimes \mathbf{B}^{-1})(\mathbf{A} \otimes \mathbf{B}) = (\mathbf{A}^{-1}\mathbf{A} \otimes \mathbf{B}^{-1}\mathbf{B}) = \mathbf{I}_m \otimes \mathbf{I}_k = \mathbf{I}_{mk}.$$

(2).

$$tr(\mathbf{A} \otimes \mathbf{B}) = \sum_{i=1}^{m} a_{ii} tr(\mathbf{B}) = tr(\mathbf{A})tr(\mathbf{B}).$$ ■

3.5 The Vec Operator

The operator that transforms a matrix to a vector is known as the **vec** operator. If the $m \times n$ matrix $\mathbf{A}$ has $\mathbf{a}_i$ as its ith column, then vec($\mathbf{A}$) is the $m \cdot n \times 1$ vectors given by

$$vec(\mathbf{A}) = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ . \\ . \\ . \\ \mathbf{a}_n \end{bmatrix}.$$

Result.

Let $\mathbf{a}$ be $m \times 1$ and $\mathbf{b}$ be $n \times 1$ vectors, then
(1).

$$vec(\mathbf{a}) = vec(\mathbf{a}') = \mathbf{a},$$

(2).

$$vec(\mathbf{a}\mathbf{b}') = \mathbf{b} \otimes \mathbf{a}.$$

Proof.

(2).

$$vec(\mathbf{a}\mathbf{b}') = vec([b_1\mathbf{a} \;\; b_2\mathbf{a} \;\; ... \;\; b_n\mathbf{a}]) = \begin{bmatrix} b_1\mathbf{a} \\ b_2\mathbf{a} \\ . \\ . \\ . \\ b_n\mathbf{a} \end{bmatrix} = \mathbf{b} \otimes \mathbf{a}. \qquad \blacksquare$$

Result.

Let $\mathbf{A}$ and $\mathbf{B}$ both be $m \times n$ matrices, then

$$tr(\mathbf{A}'\mathbf{B}) = \{vec(\mathbf{A})\}'vec(\mathbf{B}). \tag{1-12}$$

Proof.

Let $\mathbf{a}_1$,...,$\mathbf{a}_n$ denote the columns of $\mathbf{A}$ and $\mathbf{b}_1$,...,$\mathbf{b}_n$ denote the columns of $\mathbf{B}$. Then

$$tr(\mathbf{A}'\mathbf{B}) = \sum_{i=1}^{n} (\mathbf{A}'\mathbf{B})_{ii} = \sum_{i=1}^{n} \mathbf{a}_i'\mathbf{b}_i = [\mathbf{a}_1' \; ... \;\; \mathbf{a}_n'] \begin{bmatrix} \mathbf{b}_1 \\ . \\ . \\ . \\ \mathbf{b}_n \end{bmatrix} = \{vec(\mathbf{A})\}'vec(\mathbf{B}). \blacksquare$$

Result.

Let $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ be matrices of dimension $m \times n$, $n \times p$ and $p \times q$, respectively. Then

$$vec(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})vec(\mathbf{B}). \qquad \blacksquare$$

Result.

Let $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ be matrices of dimension $m\times n$, $n\times p$, $p\times q$, and $q\times m$ respectively. Then

$$tr(\mathbf{ABCD}) = \{vec(\mathbf{A}')\}'(\mathbf{D}'\otimes\mathbf{B})vec(\mathbf{C}).$$

Proof.

Using (1-11), it follows that

$$tr(\mathbf{ABCD}) = tr\{\mathbf{A}(\mathbf{BCD})\} = \{vec(\mathbf{A}')\}'vec(\mathbf{BCD}),$$

and using (c). we have that

$$vec(\mathbf{BCD}) = (\mathbf{D}'\otimes\mathbf{B})vec(\mathbf{C}),$$

so the proof is completed. ∎

Exercise 5.

Let

$$\mathbf{\Pi} = \begin{bmatrix} \mathbf{\Pi}_1 & \mathbf{0} & . & . & \mathbf{0} \\ \mathbf{0} & \mathbf{\Pi}_2 & . & . & \mathbf{0} \\ . & & & & \\ . & & & & \\ . & & & & \\ \mathbf{0} & . & . & . & \mathbf{\Pi}_N \end{bmatrix},$$

where $\mathbf{\Pi}_i$ is $k\times p_i$ matrix. Find an expression for the matrix $\mathbf{A}$ such that

$$vec\ (\mathbf{\Pi}') = \mathbf{A}\cdot\begin{bmatrix} vec(\mathbf{\Pi}_1') \\ . \\ . \\ . \\ vec(\mathbf{\Pi}_N') \end{bmatrix}.$$ ∎

4 Diagonalization of a Matrix

Almost all vectors $\mathbf{x} \in \mathbb{R}^n$ change direction, when they are multiplied by the square matrix $\mathbf{A}_{n\times n}$. Certain exceptional vectors $\mathbf{x}$ are in the same direction as $\mathbf{Ax}$. Those are the "eigenvectors". Multiply an eigenvector by $\mathbf{A}$, and the vector $\mathbf{Ax}$ is a number λ times the original $\mathbf{x}$. The basic equation is $\mathbf{Ax} = \lambda\mathbf{x}$. The number λ is an eigenvalue of $\mathbf{A}$. The eigenvalue λ tells whether the special vector $\mathbf{x}$ is stretched or shrunk or reversed or left unchanged–when it is multiplied by $\mathbf{A}$.

4.1 Eigenvalues, Eigenvectors, and Eigenspaces

Eigenvalues and eigenvectors are special implicitly defined functions of the elements of a square matrix.

$\boxed{\text{Definition.}}$

If $\mathbf{A}$ is an $n \times n$ matrix, then any scalar λ satisfying the equation

$$\mathbf{Ax} = \lambda\mathbf{x}, \tag{1-13}$$

for some $n \times 1$ vector $\mathbf{x} \neq \mathbf{0}$, is called an *eigenvalues* of $\mathbf{A}$. The vector $\mathbf{x}$ is called an *eigenvector* of $\mathbf{A}$ corresponding to eigenvalue λ and equation (1-12) is called the *eigenvalue-eigenvector equation* of $\mathbf{A}$. ■

Equation (1-12) can be equivalently expressed as

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

Notices that if $|\mathbf{A}-\lambda\mathbf{I}| \neq \mathbf{0}$, then $(\mathbf{A}-\lambda\mathbf{I})^{-\mathbf{1}}$ would exist and so premultiplication of this equation by this inverse would lead to a contradiction of the already stated assumption that $\mathbf{x} \neq \mathbf{0}$. Thus, any eigenvalue λ must satisfy the determinantal equation

$$|\mathbf{A} - \lambda\mathbf{I}| = \mathbf{0}, \tag{1-14}$$

which is known as the characteristic equation of $\mathbf{A}$.

Example.

Let $\mathbf{A} = \begin{bmatrix} 5 & 1 \\ 2 & 4 \end{bmatrix}$, then the eigenvalues of $\mathbf{A}$ are the solution to $\begin{vmatrix} 5-\lambda & 1 \\ 2 & 4-\lambda \end{vmatrix} = 0$. Thus, the eigenvalues of $\mathbf{A}$ are $\lambda_1 = 6$ and $\lambda_2 = 3$.

To find the eigenvectors belonging to $\lambda_1 = 6$, we solve $(\mathbf{A} - 6\mathbf{I})\mathbf{x} = \mathbf{0}$ to get $\mathbf{x_1} = \begin{bmatrix} x_2 \\ x_2 \end{bmatrix}$. Thus any nonzero multiple of $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is an eigenvectors belonging to λ_1 and $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is a basis for the eigenspace corresponding to λ_1. Similarly, any nonzero multiple of $\begin{bmatrix} -1/2 \\ 1 \end{bmatrix}$ is an eigenvector belonging to λ_2. ∎

From the example above, we see that eigenvectors are not uniquely defined.[5] To remove the indeterminacy, we always (but not necessary) impose the scale constraint that

$$\mathbf{x}_i'\mathbf{x}_i = \|\mathbf{x}_i\|^2 = 1, \quad \forall i = 1, ..., n.$$

This additional equation $\mathbf{x}_i'\mathbf{x}_i = 1$ produce complete solutions for both eigenvectors in example above:

$$\begin{aligned} \text{For } \lambda_1 &= 6, \ \mathbf{x}_1 = \pm \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}, \\ \text{For } \lambda_2 &= 3, \ \mathbf{x}_2 = \pm \begin{bmatrix} 1/\sqrt{5} \\ -2/\sqrt{5} \end{bmatrix}. \end{aligned}$$

For an $n \times n$ matrix, the characteristic equation is an nth order polynomial in λ. Its solution may be n distinct values, as in the preceding example, or may contain repeated values of λ and may contain some zeros as well. However, the eigenvectors belonging to distinct eigenvalues are linear independent.

Theorem.

If $\lambda_1, \lambda_2, ..., \lambda_k$ are distinct eigenvalues of the $n \times n$ matrix $\mathbf{A}$ with corresponding eigenvectors $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k$ where $k \leq n$, then $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k$ are linear independent.

[5] Note that if a nonnull vector $\mathbf{x}$ satisfies (1-12) for a given value of λ, then so will $(\alpha\mathbf{x})$ for any nonzero scalar α.

Proof.

See p. 273 of Leon, S.J. (1990). ■

Furthermore, there is no guarantee that the eigenvalues (and eigenvectors) will be real.

Exercise 6.

Find the eigenvectors and eigenvalues of the matrix $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ -2 & -1 \end{bmatrix}$. ■

4.2 Symmetric Matrices

Many of the applications involving eigenvalues and eigenvectors in statistics are ones that deal with a symmetric matrix. Symmetric matrices have some especially nice properties regarding eigenvalues and eigenvectors. In this section we will develop some of these properties.

We have seen that a matrix may have complex eigenvalues even when the matrix itself is real. This is not the case for symmetric matrices.

Theorem.

Let $\mathbf{A}$ be an $n \times n$ real symmetric matrix. Then the eigenvalues of $\mathbf{A}$ are real, and corresponding to any eigenvalue there exist eigenvectors that are real.

Proof.

Let $\lambda = \alpha + i\beta$ be an eigenvalue of $\mathbf{A}$ and $\mathbf{x} = \mathbf{y} + i\mathbf{z}$ a corresponding eigenvector, where $i = \sqrt{-1}$. We will first show that $\beta = 0$. Substitution of these expressions for λ and $\mathbf{x}$ in the eigenvalue-eigenvector equation (1-12) yields

$$\mathbf{A}(\mathbf{y} + i\mathbf{z}) = (\alpha + i\beta)(\mathbf{y} + i\mathbf{z}). \tag{1-15}$$

Premultiplying (1-14) by $(\mathbf{y} - i\mathbf{z})'$ we get

$$(\mathbf{y} - i\mathbf{z})'\mathbf{A}(\mathbf{y} + i\mathbf{z}) = (\alpha + i\beta)(\mathbf{y} - i\mathbf{z})'(\mathbf{y} + i\mathbf{z}),$$

which simplifies to

$$\mathbf{y}'\mathbf{A}\mathbf{y} + \mathbf{z}'\mathbf{A}\mathbf{z} = (\alpha + i\beta)(\mathbf{y}'\mathbf{y} + \mathbf{z}'\mathbf{z}), \tag{1-16}$$

since $\mathbf{y}'\mathbf{A}\mathbf{z} = \mathbf{z}'\mathbf{A}\mathbf{y}$ follows from the symmetry of $\mathbf{A}$. Now $\mathbf{x} \neq 0$ implies that $(\mathbf{y}'\mathbf{y} + \mathbf{z}'\mathbf{z}) > 0$ and, consequently, we must have $\beta = 0$ since the left-hand side of (1-15) is real. Substituting $\beta = 0$ in (1-14), we find that

$$\mathbf{A}\mathbf{y} + i\mathbf{A}\mathbf{z} = \alpha\mathbf{y} + i\alpha\mathbf{z}.$$

Thus, $\mathbf{x} = \mathbf{y} + i\mathbf{z}$ will be an eigenvector of $\mathbf{A}$ corresponding to $\lambda = \alpha$ as long as $\mathbf{y}$ and $\mathbf{z}$ satisfy $\mathbf{A}\mathbf{y} = \alpha\mathbf{y}$, $\mathbf{A}\mathbf{z} = \alpha\mathbf{z}$, and at least one is not $\mathbf{0}$ so that $\mathbf{x} \neq 0$. A real eigenvector is then constructed by selecting $\mathbf{y} \neq \mathbf{0}$, such that $\mathbf{A}\mathbf{y} = \alpha\mathbf{y}$ and $\mathbf{z} = 0$. ■

We have seen that a set of eigenvectors of an $m \times m$ matrix $\mathbf{A}$ is linearly independent if the associated eigenvalues are all different from one another. We will now show that, if $\mathbf{A}$ is symmetric, we can say a bit more.

Theorem.

If $\mathbf{A}$ is an $m \times m$ real symmetric matrix with m distinct eigenvalues, then the set of corresponding eigenvectors will form a group of mutually orthogonal vectors.

Proof.

Let

$$\mathbf{A}\mathbf{x}_i = \lambda_i\mathbf{x}_i, \text{ and } \mathbf{A}\mathbf{x}_j = \lambda_j\mathbf{x}_j, \quad \forall i \neq j.$$

Since the eigenvalues of $\mathbf{A}$ are distinct, it follows that

$$\lambda_i\mathbf{x}_j'\mathbf{x}_i = \mathbf{x}_j'(\mathbf{A}\mathbf{x}_i) = (\mathbf{x}_j'\mathbf{A})\mathbf{x}_i = (\mathbf{A}'\mathbf{x}_j)'\mathbf{x}_i = (\mathbf{A}\mathbf{x}_j)'\mathbf{x}_i = (\lambda_j\mathbf{x}_j)'\mathbf{x}_i = \lambda_j\mathbf{x}_j'\mathbf{x}_i.$$

Thus, $(\lambda_i - \lambda_j)\mathbf{x}_j'\mathbf{x}_i = 0$. Because $\lambda_i \neq \lambda_j$, it implies $\mathbf{x}_j'\mathbf{x}_i = 0$ as required. ■

The above result is still possible even when $\mathbf{A}$ has multiple eigenvalues.

Theorem.

If the $m \times m$ matrix $\mathbf{A}$ is symmetric, then it is possible to construct a set of m eigenvectors of $\mathbf{A}$ such that the set is orthonormal, i.e. $\mathbf{x}_i'\mathbf{x}_i = 1$ for $\forall i$ and $\mathbf{x}_i'\mathbf{x}_j = 0$ for $i \neq j$.

Proof.

See p. 95 (Theorem 3.10) of Schoott, J.R. (1997). ■

4.3 Diagonalization of a Matrix

In this section we consider the problem of factoring an $n \times n$ matrix $\mathbf{A}$ into a product of the form $\mathbf{SDS}^{-1}$, where $\mathbf{D}$ is diagonal.

Definition.

An $n \times n$ matrix $\mathbf{A}$ is said to be *diagonalizable* if there exists a nonsingular matrix $\mathbf{S}$ and a diagnoal matrix $\mathbf{D}$ such that

$$\mathbf{S}^{-1}\mathbf{AS} = \mathbf{D}.$$

We say that $\mathbf{S}$ diagonalizes $\mathbf{A}$. ■

From now on, we focus only on the case that $\mathbf{A}$ is a symmetric matrix.[6] It is convenient to collect the n eigenvectors in a $n \times n$ matrix whose ith column is the $\mathbf{x}_i$ corresponding to λ_i,

$$\mathbf{X} = [\mathbf{x}_1 \ \ \mathbf{x}_2 ..... \ \ \mathbf{x}_n],$$

and the n-eigenvalues in the same order, in a diagonal matrix,

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & . & . & . & 0 \\ 0 & \lambda_2 & 0 & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & \lambda_n \end{bmatrix} = [\boldsymbol{\lambda}_1, \ \boldsymbol{\lambda}_2, ... \ , \boldsymbol{\lambda}_n].$$

[6] Because it guarantees $\mathbf{S}^{-1}$ exists.

It is easy to see that

$$\begin{aligned}\mathbf{AX} &= [\mathbf{Ax}_1 \;\; \mathbf{Ax}_2 ..... \;\; \mathbf{Ax}_n] \\ &= [\lambda_1 \mathbf{x}_1, \; \lambda_2 \mathbf{x}_2, ... \;\; , \lambda_n \mathbf{x}_n],\end{aligned}$$

and

$$\begin{aligned}\mathbf{X\Lambda} &= [\mathbf{X}\boldsymbol{\lambda}_1, \; \mathbf{X}\boldsymbol{\lambda}_2, ... \;\; , \mathbf{X}\boldsymbol{\lambda}_n] \\ &= [\lambda_1 \mathbf{x}_1, \; \lambda_2 \mathbf{x}_2, ... \;\; , \lambda_n \mathbf{x}_n].\end{aligned}$$

Therefore, we have the useful results that

$$\mathbf{AX} = \mathbf{X\Lambda}. \tag{1-17}$$

Since the eigenvectors are orthogonal and $\mathbf{x}_i'\mathbf{x}_i = 1$, we have

$$\begin{aligned}\mathbf{X}'\mathbf{X} = \begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_n' \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix} &= \begin{bmatrix} \mathbf{x}_1'\mathbf{x}_1 & \mathbf{x}_1'\mathbf{x}_2 & \cdots & \mathbf{x}_1'\mathbf{x}_n \\ \mathbf{x}_2'\mathbf{x}_1 & \mathbf{x}_2'\mathbf{x}_2 & \cdots & \mathbf{x}_2'\mathbf{x}_1 \\ & & \vdots & \\ \mathbf{x}_n'\mathbf{x}_1 & \mathbf{x}_n'\mathbf{x}_2 & \cdots & \mathbf{x}_n'\mathbf{x}_n \end{bmatrix} \\ &= \mathbf{I}.\end{aligned} \tag{1-18}$$

equation (1-18) implies that

$$\mathbf{X}' = \mathbf{X}^{-1}.$$

Consequently,

$$\mathbf{XX}' = \mathbf{XX}^{-1} = \mathbf{I}.$$

By premultiplying (1-17) by $\mathbf{X}'$ and using (1-18), we can extract the eigenvalues of $\mathbf{A}$.

Definition.

The *diagonalization of a symmetric matrix* $\mathbf{A}$ is

$$\mathbf{X}'\mathbf{AX}(= \mathbf{X}'\mathbf{X\Lambda} = \mathbf{I\Lambda}) = \mathbf{\Lambda}. \qquad \blacksquare$$

Alternatively, by postmultiplying (1-17) by $\mathbf{X}'$ and using (1-18), we obtain a useful representation of $\mathbf{A}$.

Definition.

The spectral decomposition of a symmetric matrix $\mathbf{A}$ is

$$\mathbf{A}(=\mathbf{AXX}') = \mathbf{X\Lambda X}'. \qquad \blacksquare$$

4.4 Rank, Trace and Determinant of a matrix

Using the results in the spectral decomposition and matrix diagnoalization, it is easy to see that

(a).

$$rank(\mathbf{A}) = rank(\mathbf{X\Lambda X}') = rank(\mathbf{\Lambda}) = \textit{ numbers of non zero eigenvalues of } A.$$

(b).

$$\begin{aligned} tr(\mathbf{A}) &= tr(\mathbf{X\Lambda X}') = tr(\mathbf{\Lambda X'X}) = tr(\mathbf{\Lambda I}) = tr(\mathbf{\Lambda}) \\ &= \sum_{i=1}^{n} \lambda_i = \textit{ summations of eigenvalues of } \mathbf{A}. \end{aligned}$$

(c).

$$\begin{aligned} |\mathbf{A}| = |\mathbf{X\Lambda X}'| &= |\mathbf{X}||\mathbf{\Lambda}||\mathbf{X}'| \\ &= |\mathbf{\Lambda}||\mathbf{X}'||\mathbf{X}| \\ &= |\mathbf{\Lambda}||\mathbf{X}'\mathbf{X}| \\ &= |\mathbf{\Lambda}||\mathbf{I}| \\ &= |\mathbf{\Lambda}| = \prod_{i=1}^{n} \lambda_i = \textit{ products of eigenvalues of } \mathbf{A}. \end{aligned}$$

Exercise 7.

Let $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 9 & 6 \\ 3 & 6 & 7 \end{bmatrix}$. Find eigenvalues and eigenvectors of $\mathbf{A}$. In addition, verify that $tr(\mathbf{A}) = \sum_{i=1}^{3} \lambda_i$, $|\mathbf{A}| = \prod_{i=1}^{3} \lambda_i$, and eigenvectors are orthonormal. Finally, is $\mathbf{A}$ a full rank matrix ? ■

Exercise 8.

Let $\mathbf{\Sigma} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$. Find eigenvalues and eigenvectors of $\mathbf{\Sigma}$. ■

4.5 Powers of a Matrix

4.5.1 Expanding a Matrix by Integer Power

We often use expressions involving powers of matrices, such as $\mathbf{AA} = \mathbf{A}^2$. For positive integer power, these expressions can be computed by repeated multiplication. Consider first

$$\begin{aligned} \mathbf{AA} = \mathbf{A}^2 &= (\mathbf{X\Lambda X'})(\mathbf{X\Lambda X'}) \\ &= \mathbf{X\Lambda X'X\Lambda X'} \\ &= \mathbf{X\Lambda I\Lambda X'} \\ &= \mathbf{X\Lambda\Lambda X'} \\ &= \mathbf{X\Lambda^2 X'}. \end{aligned} \tag{1-19}$$

It implies the following results.

Result.

For any symmetric matrix, the eigenvalus of $\mathbf{A}^2$ are the squares of those of $\mathbf{A}$, and the eigenvectors are the same. ■

Since $\mathbf{A}^3 = \mathbf{A}\mathbf{A}^2 = \mathbf{X}\mathbf{\Lambda}^3\mathbf{X}'$ and so on, (1-19) extend to any positive integer.

Result.

For any symmetric matrix, the eigenvalus of $\mathbf{A}^k$ are the λ_i^k, and the eigenvectors are the same, where λ_i are the eigenvalues of $\mathbf{A}$. ■

If $\mathbf{A}$ is nonsingular, so that all its roots λ_i are nonzero, the this proof in (1-19) can be extended to negative powers as well. If $\mathbf{A}^{-1}$ exists, then

$$\begin{aligned}
\mathbf{A}^{-1} &= (\mathbf{X}\mathbf{\Lambda}\mathbf{X}')^{-1} \\
&= (\mathbf{X}')^{-1}\mathbf{\Lambda}^{-1}\mathbf{X}^{-1} \\
&= \mathbf{X}\mathbf{\Lambda}^{-1}\mathbf{X}'.
\end{aligned}$$

It implies the following results.

Result.

If $\mathbf{A}^{-1}$ exists, the eigenvalus of $\mathbf{A}^{-1}$ are the reciprocals of those of $\mathbf{A}$, and the eigenvectors are the same. ■

By extending the notion of repeated multiplication, we have a more general result.

Theorem.

For any nonsingular symmetric matrix $\mathbf{A} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}'$, then

$$\mathbf{A}^k = \mathbf{X}\mathbf{\Lambda}^k\mathbf{X}', \quad k = ..., -2, -1, 0, 1, 2, ... \tag{1-20}$$

■

4.5.2 Square Root of a Matrix

Sometimes we may require a matrix $\mathbf{A}$ such that $\mathbf{BB} = \mathbf{A}$,[7] the $\mathbf{B}$ is denoted as $\mathbf{A}^{1/2}$, and it can be computed as

$$\begin{aligned} \mathbf{A}^{1/2} &= \mathbf{X}\mathbf{\Lambda}^{1/2}\mathbf{X}' \\ &= \mathbf{X}\begin{bmatrix} \sqrt{\lambda_1} & 0 & . & . & . & 0 \\ 0 & \sqrt{\lambda_2} & 0 & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & \sqrt{\lambda_n} \end{bmatrix}\mathbf{X}', \end{aligned}$$

as long as all λ_i are nonnegative. This equation satisfies the requirement for a square root, since

$$\begin{aligned} \mathbf{A}^{1/2}\mathbf{A}^{1/2} &= \mathbf{X}\mathbf{\Lambda}^{1/2}\mathbf{X}'\mathbf{X}\mathbf{\Lambda}^{1/2}\mathbf{X}' \\ &= \mathbf{X}\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{1/2}\mathbf{X}' \\ &= \mathbf{X}\mathbf{\Lambda}\mathbf{X}' \\ &= \mathbf{A}. \end{aligned}$$

If we continue in this fashion, we can define the powers of a matrix more generally, still assuming that all the eigenvalues are nonnegative. For example, $\mathbf{A}^{1/3} = \mathbf{X}\mathbf{\Lambda}^{1/3}\mathbf{X}'$. Combining the above results we have the following theorem.

Theorem.

For a positive definite matrix $\mathbf{A}$, $\mathbf{A}^r = \mathbf{X}\mathbf{\Lambda}^r\mathbf{X}'$, for any real number r. ■

4.6 Idempotent Matrices

A symmetric idempotent matrix is the one such that

$$\mathbf{A}^k = \mathbf{X}\mathbf{\Lambda}^k\mathbf{X}' = \mathbf{X}\mathbf{\Lambda}\mathbf{X}' = \mathbf{A}, \; \textit{for all nonnegetive integer } k.$$

[7] In this case it is also true that $\mathbf{B}'\mathbf{B} = \mathbf{A}$ since $\mathbf{B}$ is symmetric.

Therefore, $(\lambda_i)^k = \lambda_i$ for all $i = 1, .., n$. That is all the eigenvalue of an idempotent matrix are 0 or 1. An immediate results from this is that rank of idempotent matrix is equal to its trace.

4.7 Quadratic Forms and Definite Matrices

Many optimization problems involve double sums of the following form.

Definition.

Let $\mathbf{A} = [a_{ij}]$ be a symmetric matrix of dimension $n \times n$, and let $\mathbf{c} = (c_1, c_2, ..., c_n)'$ be a column vector. The function

$$\begin{aligned} q &= \mathbf{c}'\mathbf{A}\mathbf{c} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} c_i c_j \end{aligned}$$

is called a **quadratic form** in $\mathbf{c}$, and $\mathbf{A}$ is referred as the matrix of the quadratic form. ■

Example.

In optimization $z = f(x_1, x_2)$, the FOC is

$$dz = f_1 dx_1 + f_2 dx_2,$$

and the SOC is[8]

$$d^2 z = f_{11} dx_1^2 + 2 f_{12} dx_1 dx_2 + f_{22} dx_2^2,$$

which can be written as

$$d^2 z = [dx_1 \ dx_2] \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix} \begin{bmatrix} dx_1 \\ dx_2 \end{bmatrix}, \tag{1-21}$$

[8] From p.310 of Chiang, A.C. (1984): For the function $z = f(x)$, for maximum of z, $f''(x) \leq 0$ can be translated into $d^2 z \leq 0$.

and is a quadratic form. ■

Definition.

For a given symmetric matrix $\mathbf{A}$,

(a). If $\mathbf{c}'\mathbf{A}\mathbf{c} > (<)0$ for all nonzero $\mathbf{c}$, the $\mathbf{A}$ is positive (negative) definite.

(b). If $\mathbf{c}'\mathbf{A}\mathbf{c} \geq (\leq)0$ for all nonzero $\mathbf{c}$, the $\mathbf{A}$ is nonnegative definite or positive semidefinite (nonpositive definite). ■

Theorem.

Let $\mathbf{A}$ be a symmetric matrix. If all the eigenvalues of $\mathbf{A}$ are positive (negative), then $\mathbf{A}$ is positive definite (negative definite). If some of the eigenvalues are zero, then $\mathbf{A}$ is nonnegative definite if the remainder are positive. If $\mathbf{A}$ has both negative and positive roots, then $\mathbf{A}$ is indefinite.

Proof.

Recall that

$$\mathbf{A} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}',$$

therefore the quadratic form can be written as

$$\begin{aligned}
\mathbf{c}'\mathbf{A}\mathbf{c} &= \mathbf{c}'\mathbf{X}\mathbf{\Lambda}\mathbf{X}'\mathbf{c} \\
&= \mathbf{y}'\mathbf{\Lambda}\mathbf{y} \\
&= \sum_{i=1}^{n} \lambda_i y_i^2,
\end{aligned}$$

where $\mathbf{y} = \mathbf{X}'\mathbf{c}$ is a $n \times 1$ real vector. ■

4.7.1 Nonnegative Definite Matrices

Some useful results pertaining to non-negative definite matrices are in the following:

Results.

(a). If $\mathbf{A}$ is nonnegative definite, then $|\mathbf{A}| \geq \mathbf{0}$.
(b). If $\mathbf{A}$ is positive definite, so is $\mathbf{A^{-1}}$.
(c). If $\mathbf{A}$ is $n \times k$ with full rank and $n > k$, then $\mathbf{A}'\mathbf{A}$ is positive and $\mathbf{A}\mathbf{A}'$ is nonnegative.

Proof (c).

Since $\mathbf{A}$ is full rank with $n > k$, so $\mathbf{Ac} = c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + ... + c_k\mathbf{a}_k \neq \mathbf{0}$. Therefore, $\mathbf{c}'(\mathbf{A}'\mathbf{A})\mathbf{c} = (\mathbf{Ac})'(\mathbf{Ac}) = \mathbf{y}'\mathbf{y} = \sum_i y_i^2 > 0$. Hence $\mathbf{A}'\mathbf{A}$ is positive. Meanwhile, a possible "zero" solution exist in the equation $\mathbf{A}'\mathbf{c} = \mathbf{0}$, since $\mathbf{A}'\mathbf{c} = c_1\dot{\mathbf{a}}_1 + c_2\dot{\mathbf{a}}_2 + ... + c_n\dot{\mathbf{a}}_n$. Here, $\mathbf{a}_i, i = 1, 2, ..., k$ is the ith column of $\mathbf{A}$ and $\dot{\mathbf{a}}_j, j = 1, 2, ..., n$ is the jth column of $\mathbf{A}'$. ■

4.7.2 Idempotent Quadratic Forms

A quadratic form $\mathbf{c}'\mathbf{Ac}$ is called a "Idempotent Quadratic Forms" when $\mathbf{A}$ is a symmetric idempotent matrix. Some useful results pertaining to idempotent quadratic forms are in the following:

Result.

(a). Every symmetric idempotent matrix is nonnegative definite.

(b). If $\mathbf{A}$ is symmetric and idempotent $n \times n$ with rank j, then every quadratic form in $\mathbf{A}$ can be written as $\mathbf{c}'\mathbf{Ac}$=$\sum_{i=1}^{j} y_i^2$. ■

5 The Triangular Factorization

In some applications, we shall require a matrix $\mathbf{P}$ such that

$$\mathbf{P}'\mathbf{P} = \mathbf{\Omega}.$$

One choice is

$$\mathbf{P} = \mathbf{\Lambda}^{1/2}\mathbf{X}',$$

where $\mathbf{\Lambda}$ and $\mathbf{X}$ are the matrices of eigenvalues and eigenvectors of $\mathbf{\Omega}$ as in (1-16). Hence,

$$\begin{aligned}\mathbf{P}'\mathbf{P} &= (\mathbf{X}')'(\mathbf{\Lambda}^{1/2})'\mathbf{\Lambda}^{1/2}\mathbf{X}' \\ &= \mathbf{X}\mathbf{\Lambda}\mathbf{X} = \mathbf{\Omega},\end{aligned}$$

as desired. Thus the spectral decomposition of $\mathbf{\Omega}$, $\mathbf{\Omega} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}$ is a useful results for this kind of computation.[9]

The Cholesky factorization of a symmetric positive definite matrix is an alternative representation that is useful in regression. We first introduce the triangular factorization.

Theorem.

Any positive definite symmetric $(n \times n)$ matrix $\mathbf{\Omega}$ has a unique representation of the form

$$\mathbf{\Omega} = \mathbf{A}\mathbf{D}\mathbf{A}',$$

where $\mathbf{A}$ is a lower triangular matrix with 1s along the principal diagonal,

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & . & . & . & 0 \\ a_{21} & 1 & 0 & . & . & . & 0 \\ a_{31} & a_{32} & 1 & . & . & . & 0 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ a_{n1} & a_{n2} & a_{n3} & . & . & . & 1 \end{bmatrix},$$

[9] It is to be noted that here $\mathbf{P}$ is not symmetric since $\mathbf{P}' = (\mathbf{X}')'(\mathbf{\Lambda}^{1/2})' = \mathbf{X}(\mathbf{\Lambda}^{1/2})' \neq \mathbf{\Lambda}^{1/2}\mathbf{X}' = \mathbf{P}$. If we let $\dot{\mathbf{P}} = \mathbf{X}\mathbf{\Lambda}^{1/2}\mathbf{X}'$, then $\dot{\mathbf{P}}'\dot{\mathbf{P}} = \mathbf{\Omega}$ and $\dot{\mathbf{P}}$ is symmetric.

and $\mathbf{D}$ is a diagonal matrix,

$$\mathbf{D} = \begin{bmatrix} d_{11} & 0 & 0 & . & . & . & 0 \\ 0 & d_{22} & 0 & . & . & . & 0 \\ 0 & 0 & d_{33} & . & . & . & 0 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & . & d_{nn} \end{bmatrix},$$

where $d_{ii} > 0$ for all i. This is known as the triangular factorization of $\mathbf{\Omega}$.

Proof.

Consider

$$\mathbf{\Omega} = \begin{bmatrix} \Omega_{11} & \Omega_{12} & \Omega_{13} & . & . & . & \Omega_{1n} \\ \Omega_{21} & \Omega_{22} & \Omega_{23} & . & . & . & \Omega_{2n} \\ \Omega_{31} & \Omega_{32} & \Omega_{33} & . & . & . & \Omega_{3n} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ \Omega_{n1} & \Omega_{n2} & \Omega_{n3} & . & . & . & \Omega_{nn} \end{bmatrix}.$$

Our goal here is to transform $\mathbf{\Omega}$ to be a diagonal matrix. This can be accomplished in the first step by transform $\mathbf{\Omega}$ to be a matrix with zeros in all the first rows and first columns except for the $(1,1)$ element. This set of operations is $\mathbf{\Omega}$ pre-multiplied by $\mathbf{E}_1$ and post-multiplied by $\mathbf{E}_1'$ the result is

$$\mathbf{E}_1\mathbf{\Omega}\mathbf{E}_1' = \mathbf{H}, \tag{1-22}$$

where

$$\mathbf{E}_1 = \begin{bmatrix} 1 & 0 & 0 & . & . & . & 0 \\ -\Omega_{21}\Omega_{11}^{-1} & 1 & 0 & . & . & . & 0 \\ -\Omega_{31}\Omega_{11}^{-1} & 0 & 1 & . & . & . & 0 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ -\Omega_{n1}\Omega_{11}^{-1} & 0 & 0 & . & . & . & 1 \end{bmatrix} \tag{1-23}$$

and

$$
\begin{aligned}
\mathbf{H} &= \begin{bmatrix} h_{11} & 0 & 0 & . & . & . & 0 \\ 0 & h_{22} & h_{23} & . & . & . & h_{2n} \\ 0 & h_{32} & h_{33} & . & . & . & h_{3n} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 0 & h_{n2} & h_{n3} & . & . & . & h_{nn} \end{bmatrix} \\
&= \begin{bmatrix} \Omega_{11} & 0 & 0 & . & . & . & 0 \\ 0 & \Omega_{22}-\Omega_{21}\Omega_{11}^{-1}\Omega_{12} & \Omega_{23}-\Omega_{21}\Omega_{11}^{-1}\Omega_{13} & . & . & . & \Omega_{2n}-\Omega_{21}\Omega_{11}^{-1}\Omega_{1n} \\ 0 & \Omega_{32}-\Omega_{31}\Omega_{11}^{-1}\Omega_{12} & \Omega_{33}-\Omega_{31}\Omega_{11}^{-1}\Omega_{13} & . & . & . & \Omega_{3n}-\Omega_{31}\Omega_{11}^{-1}\Omega_{1n} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 0 & \Omega_{n2}-\Omega_{n1}\Omega_{11}^{-1}\Omega_{12} & \Omega_{n2}-\Omega_{n1}\Omega_{11}^{-1}\Omega_{13} & . & . & . & \Omega_{nn}-\Omega_{n1}\Omega_{11}^{-1}\Omega_{1n} \end{bmatrix}.
\end{aligned}
$$

The matrix $\mathbf{E}_1$ always exists, provided that $\Omega_{11} \neq 0$. This is ensured in the present case, because Ω_{11} is equal to $\mathbf{e}_1'\mathbf{\Omega}\mathbf{e}_1$, where $\mathbf{e}_1' = [1\ 0\ 0\ ...\ 0]$. Since $\mathbf{\Omega}$ is positive definite, $\mathbf{e}_1'\mathbf{\Omega}\mathbf{e}_1$ must be greater than zero.[10]

We next proceed in exactly the same way with the second row and second column of $\mathbf{H}$. This set of operations is $\mathbf{H}$ promultiplied by $\mathbf{E}_2$ and postmultiplied by $\mathbf{E}_2'$ the result is

$$
\mathbf{E}_2\mathbf{H}\mathbf{E}_2' = \mathbf{K}, \tag{1-24}
$$

[10]To see this, consider $n = 3$ for example, $\mathbf{E}_1$ is so constructed such that the first column of $\mathbf{E}_1\mathbf{\Omega}$ is zero, i.e.

$$
\begin{aligned}
\mathbf{E}_1\mathbf{\Omega} &= \begin{bmatrix} 1 & 0 & 0 \\ -\Omega_{21}\Omega_{11}^{-1} & 1 & 0 \\ -\Omega_{31}\Omega_{11}^{-1} & 0 & 1 \end{bmatrix} \begin{bmatrix} \Omega_{11} & \Omega_{12} & \Omega_{13} \\ \Omega_{21} & \Omega_{22} & \Omega_{23} \\ \Omega_{31} & \Omega_{32} & \Omega_{33} \end{bmatrix} \\
&= \begin{bmatrix} \Omega_{11} & \Omega_{12} & \Omega_{13} \\ 0 & \Omega_{22}-\Omega_{21}\Omega_{11}^{-1}\Omega_{12} & \Omega_{23}-\Omega_{21}\Omega_{11}^{-1}\Omega_{13} \\ 0 & \Omega_{32}-\Omega_{31}\Omega_{11}^{-1}\Omega_{12} & \Omega_{33}-\Omega_{31}\Omega_{11}^{-1}\Omega_{13} \end{bmatrix}.
\end{aligned}
$$

It is easy to see that

$$
\begin{aligned}
\mathbf{E}_1\mathbf{\Omega}\mathbf{E}_1' &= \begin{bmatrix} \Omega_{11} & \Omega_{12} & \Omega_{13} \\ 0 & \Omega_{22}-\Omega_{21}\Omega_{11}^{-1}\Omega_{12} & \Omega_{23}-\Omega_{21}\Omega_{11}^{-1}\Omega_{13} \\ 0 & \Omega_{32}-\Omega_{31}\Omega_{11}^{-1}\Omega_{12} & \Omega_{33}-\Omega_{31}\Omega_{11}^{-1}\Omega_{13} \end{bmatrix} \begin{bmatrix} 1 & -\Omega_{21}\Omega_{11}^{-1} & -\Omega_{31}\Omega_{11}^{-1} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} \Omega_{11} & 0 & 0 \\ 0 & \Omega_{22}-\Omega_{21}\Omega_{11}^{-1}\Omega_{12} & \Omega_{23}-\Omega_{21}\Omega_{11}^{-1}\Omega_{13} \\ 0 & \Omega_{32}-\Omega_{31}\Omega_{11}^{-1}\Omega_{12} & \Omega_{33}-\Omega_{31}\Omega_{11}^{-1}\Omega_{13} \end{bmatrix} \quad (\textit{using the fact that } \mathbf{\Omega} \textit{ is symmetric.})
\end{aligned}
$$

where

$$\mathbf{E}_2 = \begin{bmatrix} 1 & 0 & 0 & . & . & . & 0 \\ 0 & 1 & 0 & . & . & . & 0 \\ 0 & -h_{32}h_{22}^{-1} & 1 & . & . & . & 0 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 0 & -h_{n2}h_{22}^{-1} & 0 & . & . & . & 1 \end{bmatrix} \tag{1-25}$$

and

$$\mathbf{K} = \begin{bmatrix} h_{11} & 0 & 0 & . & . & . & 0 \\ 0 & h_{22} & 0 & . & . & . & 0 \\ 0 & 0 & h_{33} - h_{32}h_{22}^{-1}h_{23} & . & . & . & h_{3n} - h_{32}h_{22}^{-1}h_{2n} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 0 & 0 & h_{n3} - h_{n2}h_{22}^{-1}h_{23} & . & . & . & h_{nn} - h_{n2}h_{22}^{-1}h_{2n} \end{bmatrix}.$$

The matrix $\mathbf{E}_2$ always exists, provided that $h_{22} \neq 0$. But h_{22} can be calculated as $h_{22} = \mathbf{e}_2'\mathbf{H}\mathbf{e}_2$, where $\mathbf{e}_2' = [0\ 1\ 0\ ...\ 0]$. Moreover, $\mathbf{H} = \mathbf{E}_1\boldsymbol{\Omega}\mathbf{E}_1'$, where $\boldsymbol{\Omega}$ is positive definite and $\mathbf{E}_1$ is given by (1-23). Since $\mathbf{E}_1$ is lower triangular, its determinant is the product of terms along the principle diagonal, which are all unity. Thus $\mathbf{E}_1$ is nonsingular, meaning that $\mathbf{H} = \mathbf{E}_1\boldsymbol{\Omega}\mathbf{E}_1'$ is positive definite and so $h_{22} = \mathbf{e}_2'\mathbf{H}\mathbf{e}_2$ must be strictly positive. Thus the matrix in (1-24) can always be calculated.

Proceeding through each of the columns and rows with the same approach, we see that for any positive symmetric matrix $\boldsymbol{\Omega}$ there exist matrices $\mathbf{E}_1$, $\mathbf{E}_2$,...,$\mathbf{E}_{n-1}$ such that

$$\mathbf{E}_{n-1}\cdots\mathbf{E}_2\mathbf{E}_1\boldsymbol{\Omega}\mathbf{E}_1'\mathbf{E}_2'\cdots\mathbf{E}_{n-1}' = \mathbf{D}, \tag{1-26}$$

where

$$\mathbf{D} = \begin{bmatrix} \Omega_{11} & 0 & 0 & . & . & . & 0 \\ 0 & \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12} & 0 & . & . & . & 0 \\ 0 & 0 & h_{33} - h_{32}h_{22}^{-1}h_{23} & . & . & . & 0 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & . & c_{nn} - c_{n,n-1}c_{n-1,n-1}^{-1}c_{n-1,n} \end{bmatrix},$$

with all the diagonal entries of $\mathbf{D}$ strictly positive. In general, $\mathbf{E}_j$ is a matrix with nonzero value in the jth column below the principle diagonal, 1s along the principle diagonal, and zeros everywhere else.

Thus each $\mathbf{E}_j$ is lower triangular with unit determinant. Hence $\mathbf{E}_j^{-1}$ exists, and the following matrix exists:

$$\mathbf{A} = (\mathbf{E}_{n-1} \cdots \mathbf{E}_2\mathbf{E}_1)^{-1} = \mathbf{E}_1^{-1}\mathbf{E}_2^{-1} \cdots \mathbf{E}_{n-1}^{-1}.$$

If (1-26) is premultiplied by $\mathbf{A}$ and postmulptiplied by $\mathbf{A}'$, the result is

$$\mathbf{\Omega} = \mathbf{ADA}',$$

where $\mathbf{A}$ is a lower triangular matrix with 1s along the principle diagonal from the fact that the product of lower triangular matrix is also triangular and the inverse of a lower triangular matrix is also lower triangular. ■

5.1 The Cholesky Factorization

A close related factorization of a symmetric positive definite matrix $\mathbf{\Omega}$ is obtained as follows. Define $\mathbf{D}^{1/2}$ to be the $(n \times n)$ diagonal matrix whose diagonal entries are the square roots of the corresponding elements of the matrix $\mathbf{D}$ in the triangular factorization:

$$\mathbf{D}^{1/2} = \begin{bmatrix} \sqrt{d_{11}} & 0 & 0 & . & . & . & 0 \\ 0 & \sqrt{d_{22}} & 0 & . & . & . & 0 \\ 0 & 0 & \sqrt{d_{33}} & . & . & . & 0 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & . & \sqrt{d_{nn}} \end{bmatrix}.$$

since the matrix $\mathbf{D}$ is unique and has strictly positive diagonal entries, the matrix $\mathbf{D}^{1/2}$ exists and is unique. Then the triangular factorization can be written as

$$\mathbf{\Omega} = \mathbf{AD}^{1/2}\mathbf{D}^{1/2}\mathbf{A}' = \mathbf{AD}^{1/2}(\mathbf{AD}^{1/2})'$$

or

$$\mathbf{\Omega} = \mathbf{PP}', \tag{1-27}$$

where

$$
\begin{aligned}
\mathbf{P} &\equiv \mathbf{A}\mathbf{D}^{1/2} \\
&= \begin{bmatrix} 1 & 0 & 0 & . & . & . & 0 \\ a_{21} & 1 & 0 & . & . & . & 0 \\ a_{31} & a_{32} & 1 & . & . & . & 0 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ a_{n1} & a_{n2} & a_{n3} & . & . & . & 1 \end{bmatrix} \begin{bmatrix} \sqrt{d_{11}} & 0 & 0 & . & . & . & 0 \\ 0 & \sqrt{d_{22}} & 0 & . & . & . & 0 \\ 0 & 0 & \sqrt{d_{33}} & . & . & . & 0 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & . & \sqrt{d_{nn}} \end{bmatrix} \\
&= \begin{bmatrix} \sqrt{d_{11}} & 0 & 0 & . & . & . & 0 \\ a_{21}\sqrt{d_{11}} & \sqrt{d_{22}} & 0 & . & . & . & 0 \\ a_{31}\sqrt{d_{11}} & a_{32}\sqrt{d_{22}} & \sqrt{d_{33}} & . & . & . & 0 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ a_{n1}\sqrt{d_{11}} & a_{n2}\sqrt{d_{22}} & a_{n3}\sqrt{d_{33}} & . & . & . & \sqrt{d_{nn}} \end{bmatrix}.
\end{aligned}
$$

Expression (1-27) is known as the *Cholesky factorization* of $\mathbf{\Omega}$.

Exercise 9.

Let $\mathbf{\Omega} = \begin{bmatrix} 4 & 2 & -2 \\ 2 & 10 & 2 \\ -2 & 2 & 5 \end{bmatrix}$. Find lower triangular matrices $\mathbf{P}$ and $\mathbf{A}$, and a diagonal matrix $\mathbf{D}$ such that $\mathbf{\Omega} = \mathbf{P}\mathbf{P}'$ (Cholesky decomposition) and $\mathbf{\Omega} = \mathbf{A}\mathbf{D}\mathbf{A}'$ (triangular factorization). ■

6 Calculus and Matrix Algebra

6.1 Conventional Notation

We can regard a function $y = f(x_1, x_2, ..., x_n) = f(\mathbf{x})$ as *scalar-valued function of a vector*. Following the convention, the following is the results of matrix's differentiation.

Results.

(a). Differentiating a scalar with respect to a column yield a column. The vector of partial derivatives, or *gradient vector*, or simply gradient, is

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial y}{\partial x_n} \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ \cdot \\ \cdot \\ \cdot \\ f_n \end{bmatrix} \equiv \textit{gradient vector}. \quad \blacksquare$$

Example.

A linear function is $y = a_1x_1 + a_2x_2 + ... + a_nx_n = \mathbf{a}'\mathbf{x}$. Then

$$\frac{\partial \mathbf{a}'\mathbf{x}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial \mathbf{a}'\mathbf{x}}{\partial x_1} \\ \frac{\partial \mathbf{a}'\mathbf{x}}{\partial x_2} \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial \mathbf{a}'\mathbf{x}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{bmatrix} = \mathbf{a}.$$

It can also easily found that

$$\frac{\partial \mathbf{a}'\mathbf{x}}{\partial \mathbf{x}'} = \mathbf{a}'. \tag{1-28}$$

$\blacksquare$

Example.

A quadratic form is $y = \mathbf{x}'\mathbf{A}\mathbf{x}$. Then

$$\frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \begin{cases} 2\mathbf{A}\mathbf{x} & \textit{when } \mathbf{A} \textit{ is symmetric;} \\ (\mathbf{A}+\mathbf{A}')\mathbf{x} & \textit{when } \mathbf{A} \textit{ is not symmetric.} \end{cases}$$ ■

(b). Differentiating a column vector with respect to a row vector yield a matrix. A *second derivatives matrix* or *Hessian* is computed as

$$\begin{aligned} \frac{\partial \left[\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}\right]}{\partial \mathbf{x}'} &= \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & . & . & . & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & . & . & . & \frac{\partial f_2}{\partial x_n} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & . & . & . & \frac{\partial f_n}{\partial x_n} \end{bmatrix} \\ &= \begin{bmatrix} f_{11} & f_{12} & . & . & . & f_{1n} \\ f_{21} & f_{22} & . & . & . & f_{2n} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ f_{n1} & f_{n2} & . & . & . & f_{nn} \end{bmatrix} \\ &= \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}'}, \quad \textit{Hessian Matrix.} \end{aligned}$$ ■

Example.

A set of linear function is $\mathbf{y} = \mathbf{A}\mathbf{x}$. It follows that $y_i = \mathbf{a}^i\mathbf{x}$, where $\mathbf{a}^i$ is the ith row of $\mathbf{A}$. Therefore

$$\frac{\partial y_i}{\partial \mathbf{x}'} = \frac{\partial \mathbf{a}^i\mathbf{x}}{\partial \mathbf{x}'} = \mathbf{a}^i. \quad (\textit{using } (1-28))$$

Collecting all the elements, we have

$$\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}'} = \begin{bmatrix} \frac{\partial y_1}{\partial \mathbf{x}'} \\ \frac{\partial y_2}{\partial \mathbf{x}'} \\ . \\ . \\ . \\ \frac{\partial y_n}{\partial \mathbf{x}'} \end{bmatrix} = \begin{bmatrix} \mathbf{a}^1 \\ \mathbf{a}^2 \\ . \\ . \\ . \\ \mathbf{a}^n \end{bmatrix} = \mathbf{A}. \tag{1-29}$$

■

(c). Differentiating a scalar with respect to a $n \times n$ matrix yield a $n \times n$ matrix.

Example.

(i).

$$\frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{A}} = \mathbf{x}\mathbf{x}',$$

(ii).

$$\frac{\partial \ln |\mathbf{A}|}{\partial \mathbf{A}} = (\mathbf{A}^{-1})'.$$ ■

6.2 Optimization

Many economic and statistic's problems involve finding the x (or $\mathbf{x}$) where $f(x)$ (or $f(\mathbf{x})$) is maximized or minimized.

Results.

(a). The first-order (necessary) condition for an optimum when $y = f(x)$ is

$$\frac{dy}{dx} = 0$$

(b). The second-order (sufficient) condition for an optimum when $y = f(x)$ is

$$\text{for a maximum,} \quad \frac{d^2y}{dx^2} < 0;$$

$$\text{for a minimum,} \quad \frac{d^2y}{dx^2} > 0.$$ ■

Results.

(a). . The first-order (necessary) condition for an optimum when $y = f(\mathbf{x})$ is

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \mathbf{0}.$$

(b). The second-order (sufficient) condition for an optimum when $y = f(\mathbf{x})$ is

$$\text{for a maximum,} \qquad \mathbf{H} = \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}'} \text{ must be negative definite;}$$

$$\text{for a minimum,} \qquad \mathbf{H} = \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}'} \text{ must be positive definite.}$$ ■

Example.

Let's consider a two-product firm under a perfectly competitive market. Accordingly, a firm's revenue function is

$$R = 12Q_1 + 18Q_2,$$

where Q_i represents the output level of the ith product per unit of time. The firm's cost function is assumed to be

$$C = 2Q_1^2 + Q_1Q_2 + 2Q_2^2.$$

The profit function of this hypothetical firm can now be written as

$$\pi = R - C = 12Q_1 + 18Q_2 - 2Q_1^2 - Q_1Q_2 - 2Q_2^2$$

It is our task to find the levels of Q_1 and Q_2 which, in combination, will maximize π. For this purpose, we first find the first-order partial derivatives of the profit function:

$$\pi_1 \left(= \frac{\partial \pi}{\partial Q_1} \right) = 12 - 4Q_1 - Q_2 \tag{1-30}$$

$$\pi_2 \left(= \frac{\partial \pi}{\partial Q_2} \right) = 18 - Q_1 - 4Q_2. \tag{1-31}$$

Setting these both equal to zero, to satisfy the necessary condition for maximum, we have $Q_1^* = 2$ and $Q_2^* = 4$, implying an optimal profit $\pi^* = 48$ per unit of time.

To be sure that this does represent a maximum profit, let us check the second-order condition. The second partial derivatives, obtainable by partial differentiation of (1-30) and (1-31), give us the following Hessian:

$$\mathbf{H} = \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix} = \begin{bmatrix} -4 & -1 \\ -1 & -4 \end{bmatrix}.$$

Since eigenvalues of $\mathbf{H}$ is -3 and -5 and both are smaller than zero, the Hessian matrix (or d^2z) is negative definite, and the solution does maximize the profit. ■

MICHIGAN STATE
UNIVERSITY

End of this Chapter