Linear Algebra

MPhil Preliminary Course 2020–2021

Daniel Wales
Faculty of Economics
University of Cambridge

Contents

1 Introduction

The purpose of these notes is to provide you with the linear algebra tools required for the MPhil course. The results are presented in a heuristic manner in that we do not prove them. The idea is that you will be able to apply them comfortably in unseen questions. Sections marked with a (*) are for your information only and are not examinable for the Precourse exam. This means that you can safely skip them. They are included to make these notes more "complete" and some of you may find them useful during the course.

Email me with any typos, corrections and suggestions at ddgw2@cam.ac.uk. These notes are based on the notes and comments of Alexis De Boeck, Nicolas Paez, Donald Robertson and Ekaterina Smetanina. Many thanks go to them.

Before starting a course on linear algebra, it is useful to consider why economists may wish to understand these issues. Broadly, in economics the motivation for understanding concepts in linear algebra may be categorised as follows:

(i) If expressed in matrix from, the mathematical results from linear algebra may be applied directly to economic problems and proofs;

(ii) Modern economists work with vast amounts of data and matrices are often the clearest and most convenient form for describing problems and their solution;

(iii) More complicated problems can often be linearised and solved easily to a good degree of approximation.

I hope to illustrate some of these benefits through economic applications of linear algebra in these notes, though the real benefit of an understanding of linear algebra is likely to be more scattered throughout the MPhil course.

There are two primary ways to contemplate topics in linear algebra:

(i) A geometric interpretation takes vectors as arrows in some set of space (e.g. $\mathbf{R}^2$). Each vector is defined by a length and a direction in that space. This leads itself neatly to a graphical representation of concepts.

(ii) Alternatively, the building blocks and concepts in linear algebra may be thought of as involving lists of numbers, where order in the list matters.

I hope to provide an understanding of both views, presenting both graphical and purely mathematical arguments for the concepts in these notes to build knowledge of both interpretations, and encourage you to interlink both views.

2 Preliminary concepts

2.1 Definitions and notation

Definition 1. A matrix of order $n \times p$ is a rectangular array of numbers with n rows and p columns

$$A := [(A)_{ij}] = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{np} \end{bmatrix}.$$

A matrix, $\boldsymbol{r}$, of order $1 \times p$ is called a row vector and a matrix, $\boldsymbol{c}$, of order $n \times 1$ is called a column vector.

$$\underset{1\times p}{\boldsymbol{r}} = \begin{bmatrix} r_1 & r_2 & \dots & r_p \end{bmatrix}, \qquad \underset{n\times 1}{\boldsymbol{c}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}.$$

Matrices and vectors (usually in column form) are the primary building blocks for geometrical interpretations in linear algebra. We will work extensively with them throughout.

Example 2.1. The use of matrices in economics is widespread, as they may be used to store vast amounts of information in a coherent, organised manner. One such application is through the use of an Leontief input-output matrix to represent how intermediate goods from one industry or sector are used in the production process of another. Each element in the matrix $(A)_{ij}$ represents the amount of goods from sector i which demanded by sector j. One example for a simple 3 sector economy is set out below, which concatenates the production vectors from each industry.

$\boldsymbol{A} =$

From\to	Agriculture	Manufacturing	Services
Agriculture	0.2	0.05	0.01
Manufacturing	0.3	0.5	0.1
Services	0.3	0.1	0.3

Definition 2. A matrix of order $n \times p$ is said to have row dimension n and column dimension p.

Definition 3. The transpose of an $n \times p$ matrix $\boldsymbol{A}$ is defined by $\boldsymbol{A}' := [(\boldsymbol{A})_{ji}]$. It is the $p \times n$ matrix obtained by interchanging the rows and columns of $\boldsymbol{A}$.

Example 2.2. When the rows and columns of the matrix $\boldsymbol{A}$ are interchanged we observe the transpose $\boldsymbol{A}'$.

$$\underset{n\times p}{\boldsymbol{A}} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{np} \end{bmatrix} \quad \text{infers that} \quad \underset{p\times n}{\boldsymbol{A}'} = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1p} & a_{2p} & \dots & a_{np} \end{bmatrix}.$$

We will use the following notation: bold italic capital letters, e.g. $\boldsymbol{A}$ and $\boldsymbol{B}$, denote matrices, bold italic lower-case letters, e.g. $\boldsymbol{a}$ and $\boldsymbol{b}$, are vectors and scalars are all in lower case. The set of natural numbers, integers and real numbers are respectively $\mathbf{N}$, $\mathbf{Z}$ and $\mathbf{R}$. The n-dimensional Euclidean space is $\mathbf{R}^n$. Subscripts are used to indicate that we are looking at a particular element of a larger quantity. For example, we may use a_i to denote the ith element of the vector $\boldsymbol{a}$ or $\boldsymbol{a}_j$ to represent the jth column of the matrix $\boldsymbol{A}$. The ijth element of $\boldsymbol{A}$ is represented by $(\boldsymbol{A})_{ij}$. All vectors are column vectors; if we need to work with the rows of a matrix we denote these by using a superscript, e.g. $\boldsymbol{a}^i$ is the ith row of $\boldsymbol{A}$.

2.2 Special matrices

Below is a list of matrices which will be of particular interest during these lectures. We are running ahead of ourselves as we have not yet defined any operations on vectors or matrices, but you can return to this section after completing the notes.

Definition 4. A matrix is said to be square if it has the same number of columns as rows, i.e. $n = p$.

Example 2.3. Examples of square matrices include a scalar, $\boldsymbol{a}$, a 2×2 matrix, $\boldsymbol{B}$, and a 3×3 matrix, $\boldsymbol{C}$.

$$\underset{1\times 1}{\boldsymbol{a}} = 1, \qquad \underset{2\times 2}{\boldsymbol{B}} = \begin{bmatrix} 1 & 3 \\ 3 & 4 \end{bmatrix}, \qquad \underset{3\times 3}{\boldsymbol{C}} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}.$$

Definition 5. A matrix $\boldsymbol{A}$ is symmetric if it is equal to its transpose, i.e. $\boldsymbol{A} = \boldsymbol{A}'$. A scalar will be symmetric by definition.

Example 2.4. In the example above, the matrix $\boldsymbol{B}$ is symmetric, as $\boldsymbol{B} = \boldsymbol{B}'$, while the matrix $\boldsymbol{C}$ is clearly not since $\boldsymbol{C} \neq \boldsymbol{C}'$.

Exercise 2.1. Given that $\boldsymbol{A}$ is symmetric and $\boldsymbol{B}$ is skew-symmetric ($\boldsymbol{B} = -\boldsymbol{B}'$), find a, b, c, u, v, w, x, y

and z.

$$\boldsymbol{A} = \begin{bmatrix} 3 & a & -1 \\ 2 & 5 & c \\ b & 8 & 2 \end{bmatrix}, \qquad \boldsymbol{B} = \begin{bmatrix} u & 3 & v \\ w & x & y \\ -2 & 6 & z \end{bmatrix}.$$

Definition 6. A square matrix $\boldsymbol{D}$ is diagonal if it is of the form

$$\boldsymbol{D} := \text{diag}(d_1, \ldots, d_n) = \begin{bmatrix} d_1 & 0 & \ldots & 0 \\ 0 & d_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & d_n \end{bmatrix},$$

where the elements off the main diagonal are all zero. Diagonal matrices are always symmetric.

Example 2.5. The identity matrix is the diagonal $n \times n$ matrix with ones as diagonal elements

$$\boldsymbol{I}_n := \text{diag}(1, \ldots, 1) = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix}.$$

Furthermore, pre- or post-multiplying any square matrix, $\underset{n\times n}{\boldsymbol{A}}$, by the identity matrix, $\boldsymbol{I}_n$ will leave that matrix unchanged such that $\boldsymbol{A}\boldsymbol{I}_n = \boldsymbol{A} = \boldsymbol{I}_n\boldsymbol{A}$. As the identity matrix, $\boldsymbol{I}_n$, is a diagonal matrix it is also symmetric.

Definition 7. A square matrix, $\boldsymbol{L}$, is lower triangular if all its elements above the main diagonal are zero. The transpose of a lower triangular matrix, $\boldsymbol{U} = \boldsymbol{L}'$ is known as an upper triangular matrix.

$$\boldsymbol{L} = \begin{bmatrix} a_{11} & 0 & \ldots & 0 \\ a_{21} & a_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}, \qquad \boldsymbol{U} = \boldsymbol{L}' = \begin{bmatrix} a_{11} & a_{21} & \ldots & a_{n1} \\ 0 & a_{22} & \ldots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}.$$

Triangular matrices are useful in determining the solution to a series of simultaneous linear equations.

Definition 8. A square matrix $\boldsymbol{A}$ is orthogonal if $\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{I}_n = \boldsymbol{A}\boldsymbol{A}'$. Hence, $\boldsymbol{A}^{-1} = \boldsymbol{A}'$.

Definition 9. A square matrix is called idempotent if it is equal to its square: $\boldsymbol{A} = \boldsymbol{A}^2$.

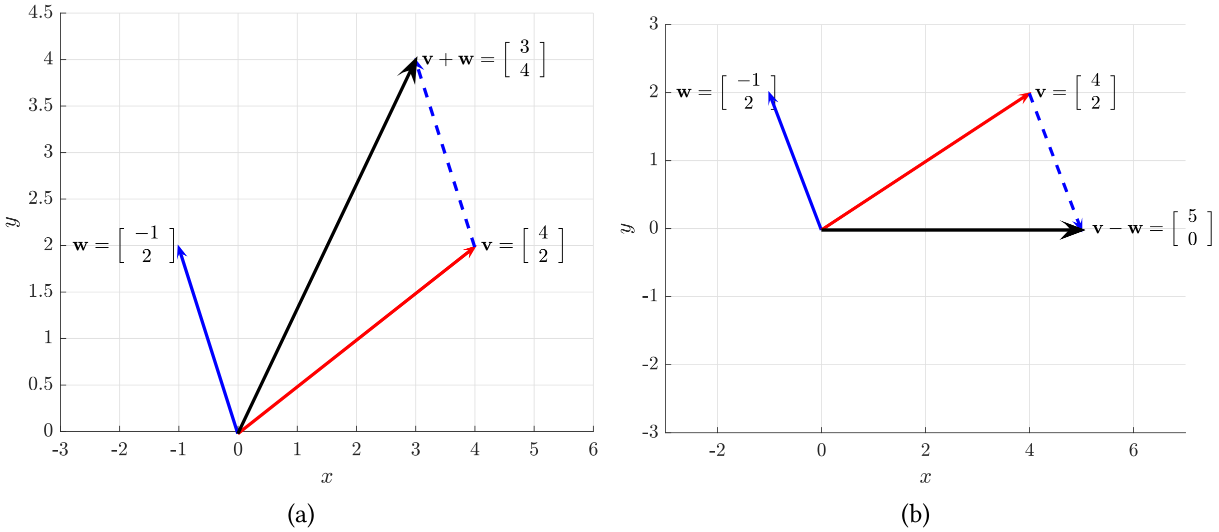


Figure 1: (a) vector addition; (b) vector subtraction.

Exercise 2.2. Show that the identity matrix of any dimension may be said to be square, symmetric, diagonal, lower triangular, upper triangular, orthogonal and idempotent.

2.3 Operations with vectors and matrices

We would like to discuss basic operations such as addition, subtraction and multiplication between matrices and in order to do so we need the notion of conformability. Two matrices are conformable if their dimensions allow the desired operation. Analogously, the operation just does not make any sense for non-conformable matrices.

2.3.1 Addition and scalar multiplication

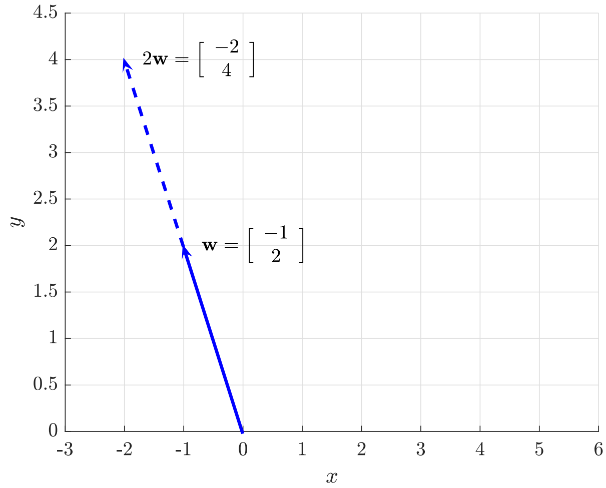
Vector and matrix addition (and equivalently subtraction) are defined element-wise. Therefore, both vectors and matrices need to be of the same dimension to be conformable. Adding two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ together amounts to adding the ith element of $\boldsymbol{a}$ to the ith element of $\boldsymbol{b}$ for all $i = 1, \ldots, n$:

$$\boldsymbol{a} + \boldsymbol{b} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{bmatrix}.$$

To understand why we define addition and subtraction element-wise we need to turn to the geometric interpretation of these concepts. Figure 1 provides a visual illustration, in $\mathbf{R}^2$, of vector addition (panel a) and vector subtraction (panel b) for the two vectors:

$$\boldsymbol{v} = \begin{bmatrix} 4 \\ 2 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{w} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}.$$

The geometric illustration says, first move along the vector $\boldsymbol{v}$, then move along $\boldsymbol{w}$. Logically, the resulting



(a)

Figure 2: Scalar multiplication of a vector.

point should be the addition of these vectors. Of course this operation could have been conducted in reverse, starting by moving along the vector $\boldsymbol{w}$. Seen in this way, we can understand precisely why addition is only conformable for vectors of the same length. For example, it makes little sense to first move along a vector in $\mathbf{R}^2$, and then one in $\mathbf{R}^3$.

Vector subtraction works in a similar fashion. We first move along vector $\boldsymbol{v}$ and then backwards along the vector $\boldsymbol{w}$. The result is shown in Figure 1 (panel b).

Matrix addition is defined equivalently. In order to add two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ we add the ijth element of $\boldsymbol{A}$ to the ijth element of $\boldsymbol{B}$ for all $i = 1, \ldots, n$ and $j = 1, \ldots, p$:

$$\begin{aligned}\boldsymbol{A} + \boldsymbol{B} &= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{bmatrix} \\ &= \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1p} + b_{1p} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2p} + b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} + b_{n1} & a_{n2} + b_{n2} & \cdots & a_{np} + b_{np} \end{bmatrix}.\end{aligned}$$

Scalar multiplication of a vector by a constant is also defined element-wise. A representation of this is shown in Figure 2. Geometrically this shows how one may consider the product $2\boldsymbol{w}$ as saying move along the vector $\boldsymbol{w}$ and then move in the same direction, for the same length, once again.

If we multiply a matrix (or a vector) by a constant $c \in \mathbf{R}$, then we multiply each element by that scalar such that for a matrix $\boldsymbol{A} \in \mathbf{R}^{n \times p}$ and a scalar $c \in \mathbf{R}$,

$$c\boldsymbol{A} = \begin{bmatrix} ca_{11} & ca_{12} & \dots & ca_{1p} \\ ca_{21} & ca_{22} & \dots & ca_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{n1} & ca_{n2} & \dots & ca_{np} \end{bmatrix}.$$

We focus exclusively on the properties of matrix scalar multiplication, but these carry over to the vector case without loss of generality.

Properties 1 (Matrix addition)**.** For three conformable matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$:

(i) $\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A}$;

(ii) $c(\boldsymbol{A} + \boldsymbol{B}) = c\boldsymbol{A} + c\boldsymbol{B}$, for any $c \in \mathbf{R}$;

(iii) $\boldsymbol{A} + (\boldsymbol{B} + \boldsymbol{C}) = (\boldsymbol{A} + \boldsymbol{B}) + \boldsymbol{C}$.

Exercise 2.3. Show the three properties above.

Exercise 2.4. Show that $(\boldsymbol{A} + \boldsymbol{B})' = \boldsymbol{A}' + \boldsymbol{B}'$ for any two conformable matrices $\boldsymbol{A}$ and $\boldsymbol{B}$.

Example 2.6. An input-output matrix (example 2.1) may also be adapted to show the use of vector addition. Converting the columns of an input-output matrix to vectors we have that total (internal) demand, $\boldsymbol{d}$, for goods produced by each sector of the economy is given as the row sum.

$$\boldsymbol{d} = \begin{bmatrix} 0.2 \\ 0.3 \\ 0.3 \end{bmatrix} + \begin{bmatrix} 0.05 \\ 0.5 \\ 0.1 \end{bmatrix} + \begin{bmatrix} 0.01 \\ 0.1 \\ 0.1 \end{bmatrix} = \begin{bmatrix} 0.26 \\ 0.9 \\ 0.5 \end{bmatrix}.$$

The impact of scalar multiplication may also be shown here. Suppose the size of (internal) demand from the manufacturing industry were to half, as it becomes more productive. The impact on internal demand may be calculated by computing a new vector, $\boldsymbol{f}$, given as

$$\boldsymbol{f} = \begin{bmatrix} 0.2 \\ 0.3 \\ 0.3 \end{bmatrix} + 0.5 \begin{bmatrix} 0.05 \\ 0.5 \\ 0.1 \end{bmatrix} + \begin{bmatrix} 0.01 \\ 0.1 \\ 0.1 \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.3 \\ 0.3 \end{bmatrix} + \begin{bmatrix} 0.025 \\ 0.25 \\ 0.05 \end{bmatrix} + \begin{bmatrix} 0.01 \\ 0.1 \\ 0.1 \end{bmatrix} = \begin{bmatrix} 0.235 \\ 0.65 \\ 0.45 \end{bmatrix}.$$

2.3.2 Vector multiplication

First, consider the vector case. When talking about multiplying two vectors, we usually mean the inner product (or dot product). For two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ the inner product is defined as follows

$$\boldsymbol{a} \cdot \boldsymbol{b} := \sum_{i=1}^{n} a_i b_i.$$

Hence, both vectors need to be of the same dimension in order to be conformable. Often, you will also see the notation: $\boldsymbol{a}'\boldsymbol{b}$ or $\langle \boldsymbol{a}, \boldsymbol{b} \rangle$.

Properties 2 (Inner product)**.** For three conformable vectors $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$:

(i) $\boldsymbol{a} \cdot \boldsymbol{b} = \boldsymbol{b} \cdot \boldsymbol{a}$;

(ii) $\boldsymbol{a} \cdot (\boldsymbol{b} + \boldsymbol{c}) = \boldsymbol{a} \cdot \boldsymbol{b} + \boldsymbol{a} \cdot \boldsymbol{c}$;

(iii) $(c_1\boldsymbol{a}) \cdot (c_2\boldsymbol{b}) = c_1 c_2 (\boldsymbol{a} \cdot \boldsymbol{b})$, for any $c_1, c_2 \in \mathbf{R}$.

Exercise 2.5. Show the three properties above.

The notion of an inner product allows us to compute the Euclidean norm of a vector[1]

$$\|\boldsymbol{a}\| = \sqrt{\boldsymbol{a} \cdot \boldsymbol{a}} = \sqrt{\sum_{i=1}^{n} a_i^2},$$

or the Euclidean distance between two vectors

$$\|\boldsymbol{a} - \boldsymbol{b}\| = \sqrt{(\boldsymbol{a} - \boldsymbol{b}) \cdot (\boldsymbol{a} - \boldsymbol{b})} = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2}.$$

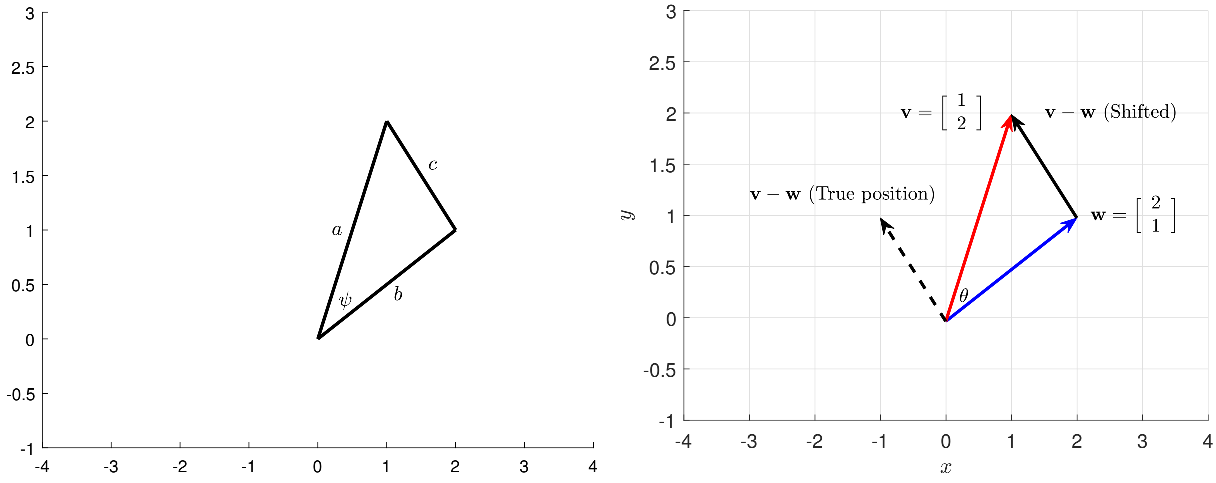
Example 2.7. A firm demands a number of inputs, $\boldsymbol{x}$, in order to produce output. At given market prices, $\boldsymbol{p}$, we may compute total costs, $\boldsymbol{c}$, for the firm.

$$\boldsymbol{c} = \boldsymbol{p} \cdot \boldsymbol{x} = \sum_{i=1}^{n} \boldsymbol{p}_i \boldsymbol{x}_i = p_1 x_1 + \cdots + p_n x_n.$$

2.3.3 Geometric interpretation of the dot product (*)

These arguments are taken from Binmore and Davies (2002). Consider any triangle, with lengths a, b and c, as shown in Figure 3, panel (a). The generalised Pythagorean theorem may be used to give a relationship between the length of the sides and one of the angles. In particular the cosine rule state

[1]This is not the only way to define the norm of a vector, but probably the one you are the most familiar with. If you want to know more look up "ℓ_p-norms".



(a) (b)

Figure 3: (a) Generalised Pythagorean theorem; (b) Geometric dot product.

that:

$$c^2 = a^2 + b^2 - 2ab\cos\psi.$$

Now, instead consider the same vector space we used above, with vectors $\boldsymbol{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. We are able to construct a triangle by moving the position of the composite vector $\boldsymbol{v} - \boldsymbol{w}$, and considering the angle, θ, between the vectors. This is shown in Figure 3, panel (b). In this case the cosine rule becomes:

$$||\boldsymbol{v} - \boldsymbol{w}||^2 = ||\boldsymbol{v}||^2 + ||\boldsymbol{w}||^2 - 2||\boldsymbol{v}||\,||\boldsymbol{w}|| \cos\theta,$$

where we restrict $0 \leq \theta \leq \pi$. Using the definition of the dot product we may expand this expression to show:

$$||\boldsymbol{v} - \boldsymbol{w}||^2 = \langle \boldsymbol{v} - \boldsymbol{w}, \boldsymbol{v} - \boldsymbol{w} \rangle,$$

$$||\boldsymbol{v} - \boldsymbol{w}||^2 = \langle \boldsymbol{v}, \boldsymbol{v} \rangle - \langle \boldsymbol{w}, \boldsymbol{v} \rangle - \langle \boldsymbol{v}, \boldsymbol{w} \rangle + \langle \boldsymbol{w}, \boldsymbol{w} \rangle,$$

$$||\boldsymbol{v} - \boldsymbol{w}||^2 = ||\boldsymbol{v}||^2 + ||\boldsymbol{w}||^2 - 2\langle \boldsymbol{v}, \boldsymbol{w} \rangle,$$

and hence by comparison to above:

$$||\boldsymbol{v}||^2 + ||\boldsymbol{w}||^2 - 2||\boldsymbol{v}||\,||\boldsymbol{w}|| \cos\theta = ||\boldsymbol{v}||^2 + ||\boldsymbol{w}||^2 - 2\langle \boldsymbol{v}, \boldsymbol{w} \rangle,$$

$$\cos\theta = \frac{\langle \boldsymbol{v}, \boldsymbol{w} \rangle}{||\boldsymbol{v}||\,||\boldsymbol{w}||},$$

for $0 \leq \theta < \pi$.

Heuristically, in $\mathbf{R}^2$, this means that:

(i) Whenever $\boldsymbol{v} \cdot \boldsymbol{w} > 0$ the two vectors have "similar" directions, as the angle between them satisfies $0 < \theta < \frac{\pi}{2}$;

(ii) Whenever $\boldsymbol{v} \cdot \boldsymbol{w} < 0$ the two vectors have "dissimilar" directions, as the angle between these vectors satisfies $\frac{\pi}{2} < \theta < \pi$;

(iii) Whenever $\boldsymbol{v} \cdot \boldsymbol{w} = 0$ the two vectors are perpendicular as the angle between them is $\theta = \frac{\pi}{2}$.

2.3.4 Matrix multiplication

There are multiple ways to define the notion of multiplication for matrices. Let us first review the "standard" way; the one you are all probably most familiar with. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be matrices of order $n \times p$ and $k \times m$ respectively. If $p = k$ then $\boldsymbol{A}$ and $\boldsymbol{B}$ are conformable to matrix multiplication. Let

$$\boldsymbol{C} = \boldsymbol{AB} := \begin{bmatrix} \sum_{s=1}^{k} a_{1s}b_{s1} & \sum_{s=1}^{k} a_{1s}b_{s2} & \dots & \sum_{s=1}^{k} a_{1s}b_{sm} \\ \sum_{s=1}^{k} a_{2s}b_{s1} & \sum_{s=1}^{k} a_{2s}b_{s2} & \dots & \sum_{s=1}^{k} a_{2s}b_{sm} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{s=1}^{k} a_{ns}b_{s1} & \sum_{s=1}^{k} a_{ns}b_{s2} & \dots & \sum_{s=1}^{k} a_{ns}b_{sm} \end{bmatrix},$$

where $\boldsymbol{C}$ is an $n \times m$ matrix. The ijth element of $\boldsymbol{C}$ is the nothing else than the inner product between the ith row of $\boldsymbol{A}$ and the jth column of $\boldsymbol{B}$.

$$(\boldsymbol{C})_{ij} := \boldsymbol{a}^{i} \cdot \boldsymbol{b}_{j} := \sum_{s=1}^{k} a_{is}b_{sj}.$$

In general, $\boldsymbol{AB} \neq \boldsymbol{BA}$. Indeed, $\boldsymbol{BA}$ is not even defined unless $n = m$ and in that case it is a $k \times p$ matrix. Matrix multiplication does obey the following properties.

Properties 3 (Matrix multiplication)**.** For three conformable matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$:

(i) $\boldsymbol{A}(\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{AB} + \boldsymbol{AC}$;

(ii) $(\boldsymbol{A} + \boldsymbol{B})\boldsymbol{C} = \boldsymbol{AC} + \boldsymbol{BC}$;

(iii) $\boldsymbol{A}(\boldsymbol{BC}) = (\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{ABC}$;

(iv) $c(\boldsymbol{AB}) = (c\boldsymbol{A})\boldsymbol{B} = \boldsymbol{A}(c\boldsymbol{B})$;

(v) $(\boldsymbol{AB})' = \boldsymbol{B}'\boldsymbol{A}'$.

Property (iii) means that you can multiply $\boldsymbol{BC}$ first or $\boldsymbol{AB}$ first.[2] To see where this is useful: consider a square matrix $\boldsymbol{A}$, then $\boldsymbol{A}^2\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^2$. By induction, $\boldsymbol{A}^p\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^p$ such that matrix powers follow the same

[2] The proof is a little awkward, but if you are curious, refer to the "hints" in problem 2.4.37 of Strang (2009).

rule as the real numbers:

$$A^p = \underbrace{A \cdots A}_{p \text{ factors}} \qquad (A^p)(A^q) = A^{p+q} \qquad (A^p)^q = A^{pq}.$$

The ability to break up these multiplications may have computational benefits.

Exercise 2.6.

(i) Prove property (v) above. Using this, or otherwise, prove that $(ABC)' = C'B'A'$.

(ii) Under what condition is $(AB)' = A'B'$?

(iii) Find an example of two distinct matrices A and B such that $AB = BA$.

Exercise 2.7. Consider the matrices: A is 3×5, B is 5×3, C is 5×1 and D is 3×1. Which of the following matrix operations are allowed?

(i) BA;

(ii) AB;

(iii) ABC;

(iv) DBA;

(v) $A(B + C)$.

Exercise 2.8. Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 4 & 5 & 6 \\ 1 & 2 & 3 \end{bmatrix}.$$

Compute AB', BA', $A'B$ and $B'A$.

Exercise 2.9. For square matrices A and B, which of the following statements are true, and why?

(i) $(A + B)^2 = (B + A)^2$;

(ii) $(A + B)^2 = A^2 + 2AB + B^2$;

(iii) $(A + B)^2 = A(A + B) + B(A + B)$;

(iv) $(A + B)^2 = (A + B)(B + A)$;

(v) $(A + B)^2 = A^2 + AB + BA + B^2$.

Exercise 2.10. For square matrices A and B, which of the following statements are true, and why?

(i) $(\boldsymbol{A}-\boldsymbol{B})^2 = (\boldsymbol{B}-\boldsymbol{A})^2$;

(ii) $(\boldsymbol{A}-\boldsymbol{B})^2 = \boldsymbol{A}^2 - \boldsymbol{B}^2$;

(iii) $(\boldsymbol{A}-\boldsymbol{B})^2 = \boldsymbol{A}^2 - 2\boldsymbol{A}\boldsymbol{B} + \boldsymbol{B}^2$;

(iv) $(\boldsymbol{A}-\boldsymbol{B})^2 = \boldsymbol{A}(\boldsymbol{A}-\boldsymbol{B}) - \boldsymbol{B}(\boldsymbol{A}-\boldsymbol{B})$;

(v) $(\boldsymbol{A}-\boldsymbol{B})^2 = \boldsymbol{A}^2 - \boldsymbol{A}\boldsymbol{B} - \boldsymbol{B}\boldsymbol{A} + \boldsymbol{B}^2$.

There are three other ways to view matrix multiplication that all yield the same result. Let $\boldsymbol{A}$ be an $n \times p$ matrix and let $\boldsymbol{B}$ be a $p \times m$ matrix.

(1) *Matrix $\boldsymbol{A}$ times columns of $\boldsymbol{B}$*: each column of $\boldsymbol{AB}$ is a linear combination of the columns of $\boldsymbol{A}$ such that:

$$\boldsymbol{AB} = \begin{bmatrix} \boldsymbol{A}\begin{pmatrix} | \\ \boldsymbol{b}_1 \\ | \end{pmatrix} & \boldsymbol{A}\begin{pmatrix} | \\ \boldsymbol{b}_2 \\ | \end{pmatrix} & \cdots & \boldsymbol{A}\begin{pmatrix} | \\ \boldsymbol{b}_m \\ | \end{pmatrix} \end{bmatrix},$$

where $\boldsymbol{b}_j$ is the jth column of $\boldsymbol{B}$.

(2) *Rows of $\boldsymbol{A}$ times matrix $\boldsymbol{B}$*: each row of $\boldsymbol{AB}$ is a linear combination of the rows of $\boldsymbol{B}$ such that:

$$\boldsymbol{AB} = \begin{bmatrix} \begin{pmatrix} \text{---} \ \boldsymbol{a}^1 \ \text{---} \end{pmatrix} \boldsymbol{B} \\ \begin{pmatrix} \text{---} \ \boldsymbol{a}^2 \ \text{---} \end{pmatrix} \boldsymbol{B} \\ \vdots \\ \begin{pmatrix} \text{---} \ \boldsymbol{a}^n \ \text{---} \end{pmatrix} \boldsymbol{B} \end{bmatrix},$$

where $\boldsymbol{a}^i$ is the ith row of $\boldsymbol{A}$.

(3) *Block matrix multiplication*: each block of $\boldsymbol{C}$ is the sum of block-rows of $\boldsymbol{A}$ times block-columns of $\boldsymbol{B}$

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

Example 2.8. As an important special case, let the blocks of $\boldsymbol{A}$ be its p columns and the blocks of $\boldsymbol{B}$ be its p rows, then $\boldsymbol{AB}$ is the sum of the columns of $\boldsymbol{A}$ times the rows of $\boldsymbol{B}$:

$$\begin{aligned}
\boldsymbol{AB} &= \begin{bmatrix} | & & | \\ \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_p \\ | & & | \end{bmatrix} \begin{bmatrix} \text{——} \boldsymbol{b}^1 \text{——} \\ \vdots \\ \text{——} \boldsymbol{b}^p \text{——} \end{bmatrix} \\
&= \begin{bmatrix} | \\ \boldsymbol{a}_1 \\ | \end{bmatrix} \begin{bmatrix} \text{——} \boldsymbol{b}^1 \text{——} \end{bmatrix} + \cdots + \begin{bmatrix} | \\ \boldsymbol{a}_p \\ | \end{bmatrix} \begin{bmatrix} \text{——} \boldsymbol{b}^p \text{——} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{a}_1 & \ldots & \boldsymbol{a}_p \end{bmatrix} \begin{bmatrix} \boldsymbol{b}^1 \\ \vdots \\ \boldsymbol{b}^p \end{bmatrix},
\end{aligned}$$

where $\boldsymbol{a}_i$ is the ith column of $\boldsymbol{A}$ and $\boldsymbol{b}^i$ is the ith row of $\boldsymbol{B}$. Setting the above example aside, block matrices arise naturally in many circumstances. In particular, matrices with identity matrices as their blocks are frequently encountered. Then, it can be much easier to work with the blocks rather than the entire matrix.

Exercise 2.11. Which rows or columns or matrices do you multiply to find

(i) the third column of $\boldsymbol{AB}$;

(ii) the first row of $\boldsymbol{AB}$;

(iii) the entry in row three, column 4 of $\boldsymbol{AB}$;

(iv) the entry in row 1, column 1 of $\boldsymbol{CDE}$.

Exercise 2.12. An econometrician collects data on the number of years of education and the marital status of n different individuals. They organise the data for the ith individual as a 2×1 vector $\boldsymbol{x}_i$ and arranges these n 2-dimensional vectors into an $n \times 2$ matrix $\boldsymbol{X}$. Show that $\boldsymbol{X}'\boldsymbol{X} = \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'$.

Exercise 2.13. The econometrician also collects income data on the same n individuals. Let $\boldsymbol{y}$ be the $n \times 1$ vector with ith element y_i. Show that $\boldsymbol{X}'\boldsymbol{y} = \sum_{i=1}^{n} \boldsymbol{x}_i y_i$.

Exercise 2.14. Suppose that for every $i = 1, \ldots, n$, the econometrician applies a weight $1/\sigma_i$ to the ith observation and arranges these n 2-dimensional observations into a matrix $\boldsymbol{Z}$ of order $n \times 2$ such that the ijth element of $\boldsymbol{Z}$ is $(\boldsymbol{X})_{ij}/\sigma_i$. Let $\boldsymbol{\Omega}$ be the diagonal matrix with $(\boldsymbol{\Omega})_{ii} = \sigma_i^2$. Show that

$$\boldsymbol{Z}'\boldsymbol{Z} = \sum_{i=1}^{n} (\boldsymbol{x}_i/\sigma_i)(\boldsymbol{x}_i/\sigma_i)' = \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'/\sigma_i^2 = \boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}$$

and

$$\sum_{i=1}^{n}(\boldsymbol{x}_i/\sigma_i)y_i/\sigma_i = \sum_{i=1}^{n}\boldsymbol{x}_i y_i/\sigma_i^2 = \boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{y}.$$

Exercise 2.15. Multiply a 3×3 matrix $\boldsymbol{A}$ and $\boldsymbol{I}_3$ using columns of $\boldsymbol{A}$ times rows of $\boldsymbol{I}_3$.

Exercise 2.16. Multiply $\boldsymbol{AB}$ below using *columns times rows*:

$$\boldsymbol{AB} = \begin{bmatrix} 1 & 0 \\ 2 & 4 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 3 & 3 & 0 \\ 1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \begin{bmatrix} 3 & 3 & 0 \end{bmatrix} + \underline{\hspace{3cm}} = \underline{\hspace{3cm}}.$$

Example 2.9. Again returning to the input-output matrix described earlier (exercise 2.1). We may recreate the increase in productivity of the manufacturing sector using a matrix, $\boldsymbol{S}$. We translate the quantity of each input required into a new production matrix in which manufactured goods require half the number of inputs as previously stated. This translation of the system may be written as

$$\boldsymbol{A} = \begin{bmatrix} 0.2 & 0.05 & 0.01 \\ 0.3 & 0.5 & 0.1 \\ 0.3 & 0.1 & 0.1 \end{bmatrix}, \quad \boldsymbol{S} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{where} \quad \boldsymbol{B} = \boldsymbol{AS} = \begin{bmatrix} 0.2 & 0.025 & 0.01 \\ 0.3 & 0.25 & 0.1 \\ 0.3 & 0.05 & 0.1 \end{bmatrix}$$

such that the internal demand, given as the row sum, is as calculated above.

2.4 Matrices as operators

Another way to view matrices is as linear operators, which can be made to act through matrix multiplication. The linear operator $\boldsymbol{A}$ (a matrix) acts upon an input which is either a vector $\boldsymbol{x}$ or a matrix $\boldsymbol{B}$ and yields another vector $\boldsymbol{Ax}$ or matrix $\boldsymbol{AB}$.

Example 2.10. Consider the matrix

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

This matrix is called a *permutation* matrix and it acts on a vector $\boldsymbol{x}$ according to

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \Rightarrow \quad \boldsymbol{b} = \boldsymbol{Ax} = \begin{bmatrix} x_2 \\ x_1 \end{bmatrix}$$

The linear operator $\boldsymbol{A}$ permutes the elements of $\boldsymbol{x}$.

Definition 10. A permutation matrix is a matrix obtained by reordering the rows of the identity matrix.

Exercise 2.17. Construct all the permutation matrices of order $n = 3$. How many permutation matrices of order $n = 4$ are there?

Example 2.11. In Figure 4, we plot 101 points on a circle with radius 2 and use the matrix

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

to map these to 101 points on an ellipse. The matrix $\boldsymbol{A}$ is an operator and $\boldsymbol{B}$ is the 2×101 matrix of coordinates on the circle. $\boldsymbol{C} = \boldsymbol{AB}$ produces a 2×101 matrix of coordinates on the ellipse.

To understand what happened, we need to consider each of the points in the original circle, and decompose these into their corresponding (x, y)-coordinates. Here the matrix $\boldsymbol{A}$, above tells us that the x-coordinate should be mapped to the vector

$$\hat{\boldsymbol{x}} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

while the y-coordinate should be mapped to the vector

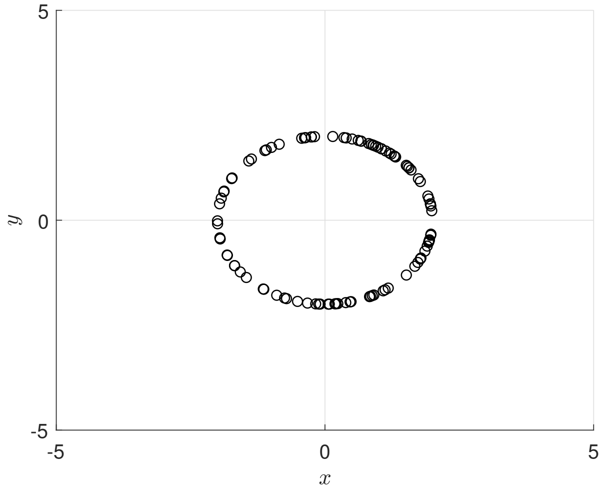
$$\hat{\boldsymbol{y}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

A graphical representation of this is shown in Figure 4 (panel c).

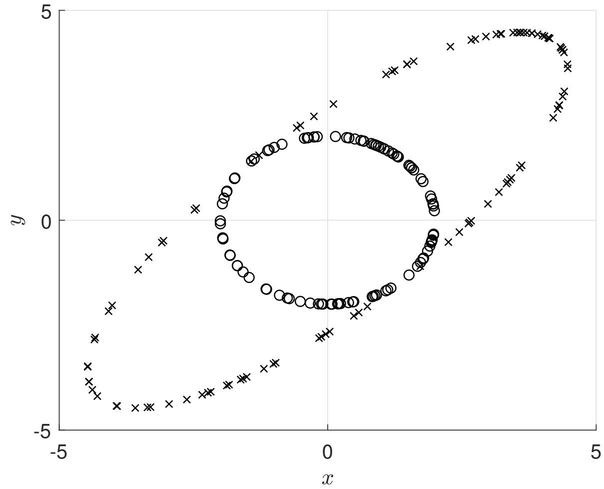
Exercise 2.18. Suppose you are given 101 points on the surface of a sphere. What is the dimension of the matrix you would require in order to map those points to 101 points on the surface of a 3-dimensional ellipsoid?

Exercise 2.19. Consider the matrix

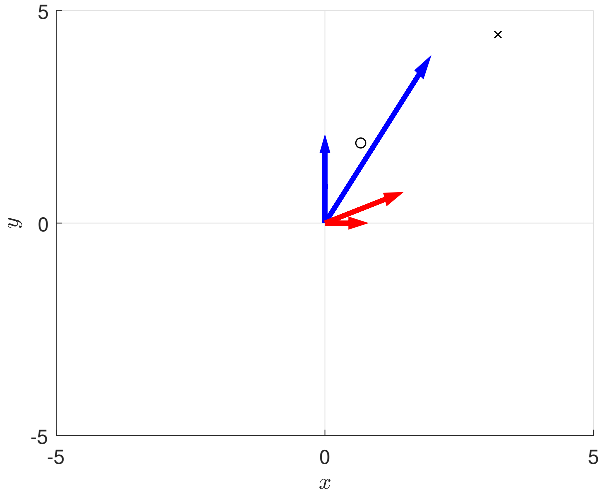
$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \\ x_{41} & x_{42} & x_{43} \\ x_{51} & x_{52} & x_{53} \end{bmatrix}.$$



(a) Initial Circle.



(b) Resultant Ellipse.



(c) Transformation of single point.

Figure 4: Panel (a) shows 101 randomly drawn points from a circle with radius 2 and the origin as centre; Panel (b) points on circle mapped into an ellipse by the linear transformation $\boldsymbol{A}$; Panel (c) shows how this may be operation may be interpreted point-by-point.

Construct a matrix $\boldsymbol{A}$ such that

$$\boldsymbol{AX} = \begin{bmatrix} x_{21} - x_{11} & x_{22} - x_{12} & x_{23} - x_{13} \\ x_{31} - x_{21} & x_{32} - x_{22} & x_{33} - x_{23} \\ x_{41} - x_{31} & x_{42} - x_{32} & x_{43} - x_{33} \\ x_{51} - x_{41} & x_{52} - x_{42} & x_{53} - x_{43} \end{bmatrix}.$$

2.4.1 Elementary matrices

Certain permutations of the identity matrix have particularly simple interpretations as operators. These matrices can be quite useful in statistics for analysing the effects on various statistical procedures of, for instance, measurement error or logarithmic transformations. We will call such matrices *elementary matrices.*

Definition 11. An elementary matrix, $\boldsymbol{E}$, is a matrix which differs from the identity matrix by a single row operation. Row operations refer to:

(i) Row interchanging. $\boldsymbol{E}_{ij}$ will denote the identity matrix with rows i and j interchanged;

(ii) Row multiplication. $\boldsymbol{E}_i(\gamma)$ will denote the identity matrix whose ith row is multiplied by $\gamma \neq 0$;

(iii) Row addition. $\boldsymbol{E}_i(\gamma|j)$ will denote the identity matrix where γ times row j is added to a row i ($i \neq j$).

Example 2.12. For $n = 3$, consider

$$\boldsymbol{E}_{13} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \qquad \boldsymbol{E}_2(7) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \boldsymbol{E}_3(5|2) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 5 & 1 \end{bmatrix}.$$

Elementary matrices satisfy the following.

Properties 4 (Elementary matrices)**.** For an elementary matrix:

(i) $\boldsymbol{E}'_{ij} = \boldsymbol{E}_{ij}$;

(ii) $\boldsymbol{E}'_i(\gamma) = \boldsymbol{E}_i(\gamma)$;

(iii) $\boldsymbol{E}'_i(\gamma|j) = \boldsymbol{E}_j(\gamma|i)$;

The proof of these results can be found in exercises 6.4 and 6.5 in Abadir and Magnus (2005, p. 134).

It may be possible to decompose a more complicated matrix into the product of elementary matrices.

Example 2.13. Consider the 3×3 matrix, $\boldsymbol{A}$, which initially appears difficult to interpret

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

The matrix $\boldsymbol{A}$ may be decomposed into a series of two steps. First add the first row to the second and then double the first row. As such we have that

$$\boldsymbol{A} = \begin{bmatrix} 2 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \boldsymbol{E}_1(2)\boldsymbol{E}_1(1|2).$$

such that $\boldsymbol{A}$ is the product of two elementary matrices.

Exercise 2.20. Decompose the following matrices into the product of series of elementary matrices and state the operations each matrix represents.

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \quad \boldsymbol{C} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

3 Systems of linear equations

3.1 Linear combinations

Definition 12 (Linear combination)**.** A linear combination of vectors is the sum of a scalar multiple of each vector.

Example 3.1. A linear combination of p vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p$ may be written as

$$c_1\boldsymbol{v}_1 + \ldots + c_p\boldsymbol{v}_p,$$

where $c_i \in \mathbf{R}$ for $i = 1, \ldots, p$.

In light of example 3.1 adding $\boldsymbol{v}$ and $\boldsymbol{w}$ is equivalent to taking a linear combination $c\boldsymbol{v} + d\boldsymbol{w}$ with $c = d = 1$. Other special linear combinations are

(i) $1\boldsymbol{v} + (-1)\boldsymbol{w}$ is the difference of vectors, as depicted in Figure 1 (b);

(ii) $0\boldsymbol{v} + 0\boldsymbol{w} = \mathbf{0}$;

(iii) $c\boldsymbol{v} + 0\boldsymbol{w}$ is a stretch or contraction of $\boldsymbol{v}$. (Scalar multiplication).

Exercise 3.1. Let $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{z}$ be vectors of the same order. Show that:

(i) $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$;

(ii) $(\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{z} = \boldsymbol{v} + (\boldsymbol{u} + \boldsymbol{z})$.

Exercise 3.2. For vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ of the same order and scalars c and d. Show that

(i) $(c + d)(\boldsymbol{u} + \boldsymbol{v}) = c\boldsymbol{v} + c\boldsymbol{u} + d\boldsymbol{v} + d\boldsymbol{u}$;

(ii) (Harder) the zero vector is uniquely determined by the condition that $c\mathbf{0} = \mathbf{0}$ for all scalars $c \in \mathbf{R}$.

We will use these linear combination to rewrite systems of linear equations, which are of interest in many economic applications. For example, consider a system with three linear equations and three unknowns.

$$\begin{aligned} x + 2y + 3z &= 6 \\ 2x + 5y + 2z &= 4 \\ 6x - 3y + z &= 2 \end{aligned}$$

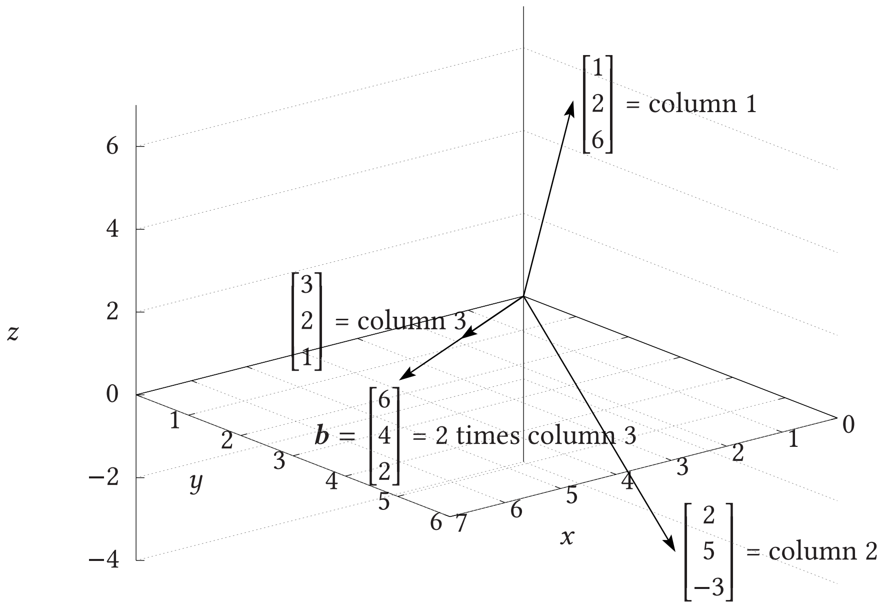


Figure 5: graphical representation of the three vectors in the system.

Notice, that we can write this system as a linear combination

$$\begin{bmatrix}1\\2\\6\end{bmatrix} x + \begin{bmatrix}2\\5\\-3\end{bmatrix} y + \begin{bmatrix}3\\2\\1\end{bmatrix} z = \begin{bmatrix}6\\4\\2\end{bmatrix},$$

or even more compactly in the form $\boldsymbol{Ax} = \boldsymbol{b}$, where

$$\boldsymbol{A} = \begin{bmatrix}1 & 2 & 3\\2 & 5 & 2\\6 & -3 & 1\end{bmatrix}, \quad \boldsymbol{x} = \begin{bmatrix}x\\y\\z\end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix}6\\4\\2\end{bmatrix}.$$

Hence, $\boldsymbol{Ax}$ is a linear combination between the columns of $\boldsymbol{A}$ weighted by the scalar components of $\boldsymbol{x}$. We wish to find the x, y and z that solve the above simultaneous equations.

A way to visualise the system is to plot the three equations in x, y, z-space, with the columns of $\boldsymbol{A}$ providing the coordinates. Each equation in the system defines a plane in x, y, z-space. Two planes meet in a line, and three planes meet at a point. The coordinates of this point provide the values of x, y and z that solve our system.

In the exercises below you may assume that a unique solution exits.

Exercise 3.3. Solve the system above.

Exercise 3.4. Write the following system in matrix form and draw the corresponding column picture.

By simple inspection, identify the $\boldsymbol{x}$ vector that solves the system

$$\begin{aligned} 2x - y &= 0 \\ -x + 2y &= 3. \end{aligned}$$

Exercise 3.5. Write the following system in matrix form and draw the corresponding column picture. By simple inspection, identify the $\boldsymbol{x}$ vector that solves the system

$$\begin{aligned} 2x - y &= 0 \\ -x + 2y - z &= -1 \\ -3y + 4z &= 4. \end{aligned}$$

3.2 Linear dependence and rank

Up to now, we have taken for granted that a unique solution to a system of equations exists. We will provide conditions under which the existence of a unique solution for $\boldsymbol{x}$ is guaranteed. The existence and uniqueness of a solution to a system of linear equations may be ascertained using a quantity called the rank of a matrix. Central to the rank of a matrix is the linear (in)dependence of a set of vectors.

Definition 13 (Linear dependence)**.** A set of vectors $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_p\}$ with $\boldsymbol{v}_i \in \mathbf{R}^n$ is called linearly *dependent* if there exits scalars $c_1, \dots, c_p$ not all zero such that

$$c_1\boldsymbol{v}_1 + \cdots + c_p\boldsymbol{v}_p = \mathbf{0}.$$

Conversely, if no such scalars exist then the set of vectors is called linearly *independent.*

Exercise 3.6. Consider the following sets of vectors, $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{w}$ and state whether they are linear independent. If they are linearly dependent write down a set of associated scalars c_1, c_2, c_3 to show this.

(i) $\boldsymbol{u} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\boldsymbol{v} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$;

(ii) $\boldsymbol{u} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, $\boldsymbol{v} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$;

(iii) $\boldsymbol{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $\boldsymbol{v} = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$;

(iv) $\boldsymbol{u} = \begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix}$, $\boldsymbol{v} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} 3 \\ 6 \\ 8 \end{bmatrix}$;

(v) $\boldsymbol{u} = \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix}$, $\boldsymbol{v} = \begin{bmatrix} 3 \\ 7 \\ -7 \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} -1 \\ -1 \\ 5 \end{bmatrix}$;

(vi) $\boldsymbol{u} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$, $\boldsymbol{v} = \begin{bmatrix} 2 \\ 2 \\ -4 \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

Exercise 3.7. Let $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ be linearly independent vectors of order n, show that $(\boldsymbol{x}+\boldsymbol{y})$, $(\boldsymbol{x}+\boldsymbol{z})$, and $(\boldsymbol{y}+\boldsymbol{z})$ are also linearly independent.

Exercise 3.8. Let $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ be a set of linearly independent vectors. Consider another vector $\boldsymbol{v}$ such that the vectors in $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p, \boldsymbol{v}\}$ are linearly dependent. Show that $\boldsymbol{v}$ can be expressed as a unique linear combination of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p$.

Definition 14. The column rank of a matrix $\boldsymbol{A}$ is the maximum number of linearly independent columns of $\boldsymbol{A}$. Likewise, the row rank is the maximum number of linearly independent rows.

Theorem 1 (Rank theorem)**.** *The column rank and row rank of any matrix $\boldsymbol{A} \in \mathbb{R}^{n\times p}$ are equal.*

Proof. See exercise 4.5 in Abadir and Magnus (2005, p. 77). ■

From Theorem 1 it readily follows that the concept of rank in unambiguous. Therefore, we will just refer to it as rank, denoted by rk, and drop the reference to "column" and "row" unless absolutely necessary.

Properties 5 (Rank)**.** For two conformable matrices $\boldsymbol{A}$ and $\boldsymbol{B}$:

(i) $0 \leq \operatorname{rk} \boldsymbol{A} \leq \min\{n, p\}$;

(ii) $\operatorname{rk} \boldsymbol{A} = \operatorname{rk} \boldsymbol{A}'$;

(iii) $\operatorname{rk} \boldsymbol{A} = 0$ if and only if $\boldsymbol{A} = \boldsymbol{O}_{n\times p}$;

(iv) $\operatorname{rk} \boldsymbol{I}_n = n$;

(v) $\operatorname{rk}(c\boldsymbol{A}) = \operatorname{rk} \boldsymbol{A}$, for any $c \in \mathbb{R}$;

(vi) $\operatorname{rk}(\boldsymbol{A}+\boldsymbol{B}) \leq \operatorname{rk} \boldsymbol{A} + \operatorname{rk} \boldsymbol{B}$;

(vii) $\operatorname{rk}(\boldsymbol{A}-\boldsymbol{B}) \geq |\operatorname{rk} \boldsymbol{A} - \operatorname{rk} \boldsymbol{B}|$;

(viii) $\operatorname{rk}(\boldsymbol{A}\boldsymbol{B}) \leq \min\{\operatorname{rk} \boldsymbol{A}, \operatorname{rk} \boldsymbol{B}\}$.

A matrix that has rank $\min\{n, p\}$ is said to be full of rank. Otherwise, we say that the matrix is rank deficient.

Example 3.2.

$$\operatorname{rk}\begin{bmatrix}1 & 2\\ 2 & 4\end{bmatrix} = 1 \quad \text{whilst} \quad \operatorname{rk}\begin{bmatrix}1 & 0\\ 0 & 2\end{bmatrix} = 2.$$

Example 3.3.

$$\text{(i)}\ \operatorname{rk}\begin{bmatrix}1 & 2 & 4\\ 2 & 4 & 8\\ 3 & 6 & 12\end{bmatrix} = 1, \quad \text{(ii)}\ \operatorname{rk}\begin{bmatrix}1 & 2 & 0\\ 2 & 4 & 7\\ 3 & 6 & 2\end{bmatrix} = 2, \quad \text{(iii)}\ \operatorname{rk}\begin{bmatrix}1 & 2 & 0\\ 2 & 3 & 7\\ 3 & 3 & 2\end{bmatrix} = 3, \quad \text{(iv)}\ \operatorname{rk}\begin{bmatrix}2 & 0\\ 3 & 7\\ 3 & 2\end{bmatrix} = 2.$$

The matrices (i) and (ii) are rank deficient, whilst (iii) and (iv) are full rank.

Exercise 3.9. Find the rank of the following matrices

$$\text{(i)}\ \boldsymbol{A} = \begin{bmatrix}0 & 3 & 0\\ 2 & 0 & 0\\ 0 & 0 & 7\end{bmatrix}, \quad \text{(ii)}\ \boldsymbol{B} = \begin{bmatrix}1 & 0 & 2 & 1\\ 0 & 2 & 4 & 2\\ 0 & 2 & 2 & 1\end{bmatrix}, \quad \text{(iii)}\ \boldsymbol{C} = \begin{bmatrix}2 & -1 & 3\\ 1 & 0 & 1\\ 0 & 2 & -1\\ 1 & 1 & 4\end{bmatrix}, \quad \text{(iv)}\ \boldsymbol{D} = \begin{bmatrix}1 & 4 & 0\\ 0 & 3 & 12\\ 0 & 2 & 8\end{bmatrix}.$$

We will now use the concept of rank to determine whether a system of equations has a solution. First, consider the following example.

Example 3.4. Consider the system of equations $\boldsymbol{Ax} = \boldsymbol{b}$ given by

$$\begin{bmatrix}1 & 1 & 2\\ 2 & 1 & 3\\ 3 & 1 & 4\\ 4 & 1 & 5\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix} = \begin{bmatrix}b_1\\ b_2\\ b_3\\ b_4\end{bmatrix}.$$

Does $\boldsymbol{Ax} = \boldsymbol{b}$ have a solution for every possible $\boldsymbol{b}$? The answer is no, but there are some vectors $\boldsymbol{b}$ that we can solve for, for instance:

(i) $\boldsymbol{b} = (0, 0, 0, 0)' \Rightarrow \boldsymbol{x} = (0, 0, 0)'$

(ii) $\boldsymbol{b} = (1, 2, 3, 4)' \Rightarrow \boldsymbol{x} = (1, 0, 0)'$

(iii) $\boldsymbol{b} = (1, 1, 1, 1)' \Rightarrow \boldsymbol{x} = (0, 1, 0)'$

(iv) $\boldsymbol{b} = (2, 3, 4, 5)' \Rightarrow \boldsymbol{x} = (0, 0, 1)'$

Whenever $\boldsymbol{b}$ is a linear combination of the columns of $\boldsymbol{A}$, the system admits at least one solution.

Formally, this requires that $\text{rk}(\boldsymbol{A}|\boldsymbol{b}) = \text{rk}\,\boldsymbol{A}$, or

$$\text{rk}\left[\begin{array}{ccc|c} 1 & 1 & 2 & b_1 \\ 2 & 1 & 3 & b_2 \\ 3 & 1 & 4 & b_3 \\ 4 & 1 & 5 & b_4 \end{array}\right] = \text{rk}\begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{bmatrix}.$$

The condition for the existence of at least one solution to a system of linear equations is that the rank of the coefficient matrix is equal to the rank of the augmented matrix, $\boldsymbol{A}|\boldsymbol{b}$, where the augmented matrix is the coefficient matrix with $\boldsymbol{b}$ attached as an extra column.

However, this still does not guarantee the uniqueness of the solution. For a unique solution, the coefficient matrix $\boldsymbol{A}$ must be full column rank, otherwise there are infinitely many solutions.

Example 3.5. Consider the previous system, but now look at a particular $\boldsymbol{b}$ vector

$$\begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Does $\boldsymbol{Ax} = \boldsymbol{0}$ have a solution? It has many, for example:

(i) $\boldsymbol{x} = (1, 1, -1)'$

(ii) $\boldsymbol{x} = (2, 2, -2)'$

In fact, it has infinitely many solutions, all of the form $\boldsymbol{x} = (c, c, -c)'$ for any $c \in \mathbf{R}$.

Example 3.6. Similarly, $\boldsymbol{x} = (1, 0, 0)'$ solves

$$\begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix},$$

but so does any $\boldsymbol{x}$ of the form

$$\boldsymbol{x} = \begin{bmatrix} 1 + c \\ c \\ -c \end{bmatrix},$$

where $c \in \mathbf{R}$.

The above examples illustrates how a system of equations with a column rank deficient coefficient matrix has infinitely many solutions.

For an $n \times p$ matrix $\boldsymbol{A}$, a p-vector $\boldsymbol{x}$ and an n-vector $\boldsymbol{b}$, the four possible scenarios are stated here without proof:

(i)	$\text{rk}\,\boldsymbol{A} = n = p$	(A is square and invertible)	$\boldsymbol{Ax} = \boldsymbol{b}$ has one solution
(ii)	$\text{rk}\,\boldsymbol{A} = n$ and $\text{rk}\,\boldsymbol{A} < p$	(A is short and wide)	$\boldsymbol{Ax} = \boldsymbol{b}$ has ∞ solutions
(iii)	$\text{rk}\,\boldsymbol{A} < n$ and $\text{rk}\,\boldsymbol{A} = p$	(A is tall and thin)	$\boldsymbol{Ax} = \boldsymbol{b}$ has 0 or 1 solution
(iv)	$\text{rk}\,\boldsymbol{A} < \min\{n, p\}$	(A is not full rank)	$\boldsymbol{Ax} = \boldsymbol{b}$ has 0 or ∞ solutions.

Note that, in the box above, the first three situations (i), (ii) and (iii) refer to a settings where the matrix $\boldsymbol{A}$ is full rank. In case (ii) and (iii) the matrix $\boldsymbol{A}$ is not square. The final case, (iv), refers to any situation where the matrix $\boldsymbol{A}$ is not full rank.

Exercise 3.10. Let $\boldsymbol{A}$ be an $n \times p$ matrix. Suppose you know that there exists an $n \times 1$ vector $\boldsymbol{a}$ such that if $\boldsymbol{a}$ is added as an additional column to $\boldsymbol{A}$, the rank of increases by 1, i.e. $\text{rk}(\boldsymbol{A}|\boldsymbol{a}) = \text{rk}(\boldsymbol{A}) + 1$. Show that this implies the rows of $\boldsymbol{A}$ are linearly dependent. *Hint: prove by contradiction.*

Exercise 3.11. Show that it is not true in general that $\text{rk}\,\boldsymbol{AB} = \text{rk}\,\boldsymbol{BA}$ for two square matrices $\boldsymbol{A}$ and $\boldsymbol{B}$.

3.2.1 Application: complete asset markets

The concepts of linear dependence and rank are widely used in both finance and economics when contemplating the financial structure of the economy, and in particular the concept of complete asset markets.

Definition 15 (Complete asset markets)**.** Asset markets are said to be *complete* when the payoff vectors of available assets span all states of nature (i.e. the payoff matrix is full column rank).

Example 3.7. Suppose a discrete number, n, of possible states of nature exist in the future. Financial assets each pay out different amounts for each state. Suppose there are p possible financial assets to purchase. This information may be summarised in a returns matrix, $\boldsymbol{R}$.

$$\boldsymbol{R} = \begin{array}{c} \\ \text{State 1} \\ \\ \text{State } n \end{array} \begin{array}{c} \begin{array}{ccc} \text{Asset 1} & & \text{Asset p} \end{array} \\ \begin{bmatrix} r_{11} & \dots & r_{1p} \\ \vdots & \ddots & \vdots \\ r_{n1} & \dots & r_{np} \end{bmatrix} \end{array}$$

As long as $\operatorname{rk} \boldsymbol{R} = n$, any consumer is able to use a combination of assets to purchase any desired level of consumption across states of nature.

Exercise 3.12. Suppose two states of nature exist $S \in \{\text{Rain}, \text{Sun}\}$, such that $n = 2$. Two financial assets exist, one which pays out 2 units in rain and another which pays out 1 unit in any state of the world. By considering the returns matrix, $\boldsymbol{R}$, argue why financial markets are complete. If a household desires 4 units of consumption in the sunny state, but none when it rains, what are the quantities of each asset they much purchase for their portfolio?

4 Square matrices

From Section 2.2 we know that a square matrix is one whose row and column dimension are the same. Throughout this section, we will discuss concepts which apply uniquely to such matrices. During your course you will encounter these regularly.

4.1 Determinant

The determinant is a single value associated with a square matrix. It has several uses and here we will focus on just a few which will be discussed in Sections 4.3 and 4.4. Before we can introduce the determinant, we need to know what a minor of a square matrix is.

Definition 16 (Minor)**.** The i, j-minor, M_{ij}, of an $n \times n$ matrix $\boldsymbol{A}$ is the determinant of the $(n-1)\times(n-1)$ submatrix created by deleting the ith row and jth column of $\boldsymbol{A}$.

Example 4.1. The 1, 1-minor of

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

is $M_{11} = \det[d] = d$. Similarly, $M_{12} = c$, $M_{21} = b$ and $M_{22} = a$.

Definition 17 (Determinant by cofactor expansion)**.** The determinant of a square $n \times n$ matrix $\boldsymbol{A}$ defined by the *cofactor expansion* is

$$\det \boldsymbol{A} := \sum_{i=1}^{n} (-1)^{i+j} a_{ij} M_{ij} \equiv \sum_{j=1}^{n} (-1)^{i+j} a_{ij} M_{ij}, \tag{1}$$

where the left-hand side of the equivalence is for fixed j and the right-hand side is for fixed i.

The definition above states that we can compute the determinant by expanding in *cofactors*, $(-1)^{i+j} M_{ij}$, either along a fixed row (fixed i) or column (fixed j).

Example 4.2. Consider again the 2×2 matrix

$$\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Use definition 17 by expanding along the first column (fix $j = 1$),

$$\begin{aligned} \det \boldsymbol{A} &= (-1)^{1+1} a_{11} M_{11} + (-1)^{2+1} a_{21} M_{21} \\ &= ad - bc. \end{aligned}$$

You can try by yourself to expand along any other column or row and you should get the same answer.

Example 4.3. Consider again the 3×3 matrix

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

Expand along the first row (fix $i = 1$) such that

$$\begin{aligned} \det \boldsymbol{A} &= (-1)^{1+1} a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + (-1)^{1+2} a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + (-1)^{1+3} a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}. \end{aligned}$$

Applying the following properties will make your life easier when computing determinants.

Properties 6 (Determinant)**.** For two $n \times n$ matrices $\boldsymbol{A}$ and $\boldsymbol{B}$:

(i) $\det \boldsymbol{I}_n = 1$;

(ii) $\det \boldsymbol{A}' = \det \boldsymbol{A}$;

(iii) $\det \boldsymbol{A}^{-1} = (\det \boldsymbol{A})^{-1}$;

(iv) $\det \boldsymbol{AB} = \det \boldsymbol{A} \det \boldsymbol{B}$;

(v) $\det(c\boldsymbol{A}) = c^n \det \boldsymbol{A}$, for any $c \in \mathbf{R}$;

(vi) for a lower or upper triangular matrix, including a diagonal matrix, $\boldsymbol{A}$

$$\det \boldsymbol{A} = \prod_{i=1}^{n} a_{ii},$$

(vii) the determinant changes signs when two rows of $\boldsymbol{A}$ are interchanged;

(viii) subtracting a multiple of one row of $\boldsymbol{A}$ from another leaves $\det \boldsymbol{A}$ unchanged;

(ix) if $\boldsymbol{A}$ has a row of zeros then $\det \boldsymbol{A} = 0$;

(x) the determinant is a function of each row separately, i.e.

$$\begin{vmatrix} ca_{11} & ca_{12} & \dots & ca_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} = c \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}$$

and

$$\begin{vmatrix} a_{11}+a'_{11} & a_{12}+a'_{12} & \dots & a_{1n}+a'_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} + \begin{vmatrix} a'_{11} & a'_{12} & \dots & a'_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}.$$

Remark. By property (ii), all the results above still hold when changing "row" to "column".

Definition 18. A singular matrix is a square matrix whose determinant is zero.

Exercise 4.1. Evaluate the determinants of the following matrices using the cofactor expansion method, along an appropriate row or column

$$\boldsymbol{A} = \begin{bmatrix} 2 & 5 & 1 \\ 1 & 0 & 2 \\ 7 & 1 & 1 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 7 & 5 & 2 & 3 \\ 2 & 0 & 0 & 0 \\ 11 & 2 & 0 & 0 \\ 23 & 57 & 1 & -1 \end{bmatrix}, \quad \boldsymbol{C} = \begin{bmatrix} 1 & 2 & 1 & 0 \\ 3 & 2 & 1 & 0 \\ 0 & 1 & 6 & 5 \\ 0 & 1 & 1 & 1 \end{bmatrix}, \quad \boldsymbol{D} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{E} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 3 & 2 \\ 0 & 0 & 0 & 2 & 9 & 3 \\ 0 & 0 & 1 & 0 & 7 & 4 \\ 0 & 6 & 9 & 8 & 7 & 5 \\ 1 & 3 & 4 & 2 & 9 & 6 \end{bmatrix},$$

$$\boldsymbol{F} = \begin{bmatrix} 3 & t & -2 \\ -1 & 5 & 3 \\ 2 & 1 & 1 \end{bmatrix}.$$

Exercise 4.2. Using property (vii), prove that if two rows of a matrix $\boldsymbol{A}$ are equal then $\det \boldsymbol{A} = 0$. *Hint: proof by contradiction by supposing that* $\boldsymbol{A}$ *has two equal rows, but* $det(A) \neq 0$.

Exercise 4.3. Prove property (viii) by using the result from exercise 4.2 and property (x). *Hint: Use both parts of property (x) to separate the relevant determinant into two components and apply exercise 4.2.*

Exercise 4.4. Prove property (ix) by using property (x).

Exercise 4.5. Prove or disprove the statement: $\det(\boldsymbol{A}+\boldsymbol{B}) = \det \boldsymbol{A} + \det \boldsymbol{B}$.

Exercise 4.6. By considering the matrices

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \qquad \text{and} \qquad B = \begin{bmatrix} 0 & 0 \\ 3 & 4 \end{bmatrix}$$

and the properties of determinants, show that $AB = O$ does not imply that either A or B is the zero matrix, but that it does imply that at least one of them is singular.

4.1.1 Geometry of the determinant (*)

The determinant of a matrix may look like a completely abstract concept, but it has a clear geometric interpretation. A generalisation of a parallelogram in n-dimensional Euclidean space is called a parallelepiped. It turns out that the volume of a parallelepiped, V, with as edges the vectors $v_1, \dots, v_n$ is

$$V = |\det(v_1, \dots, v_n)|.$$

In the case of a two dimensional matrix,

$$A = \begin{bmatrix} 4 & -1 \\ 2 & 2 \end{bmatrix},$$

this claim may be verified diagrammatically, as shown in Figure 6. Notice the vectors $v = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$ and $v = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$ are the same vectors as in Figure 1 (a).

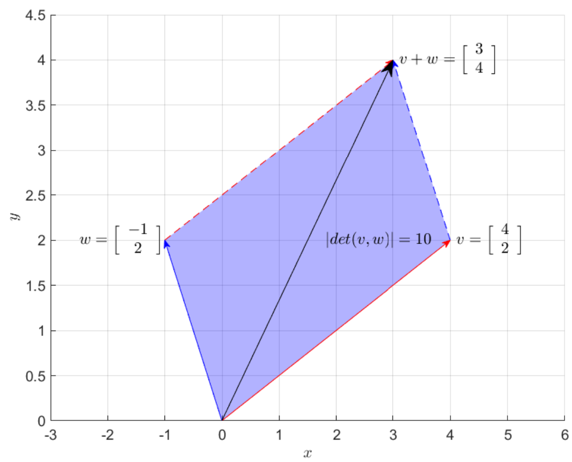


Figure 6: The absolute value of the determinant of the matrix with columns v and w is the area of the parallelogram spanned by these vectors.

Similarly, when this matrix is viewed as a basis of $\mathbf{R}^2$, we observe that the determinant is a useful way to summarise the relative area spanned by the set of basis vectors. This example is therefore shown as the difference between the standard basis, displayed in Figure 7, and the alternative basis, displayed in Figure 8.

Intuitively, we realise that it will be useful to know how much area in the standard basis translates to area in the alternative. The answer is precisely given by the determinant. Any unitary square in the standard basis is mapped into a square of size $\det(\boldsymbol{A}) = 10$ in the alternative.

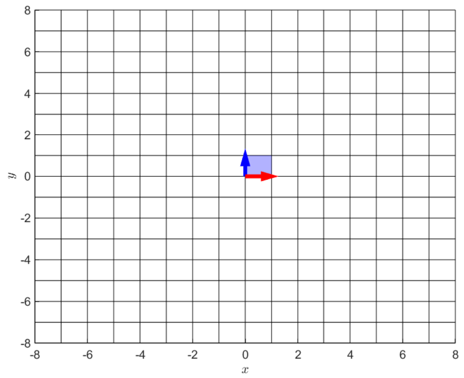


Figure 7: Standard basis

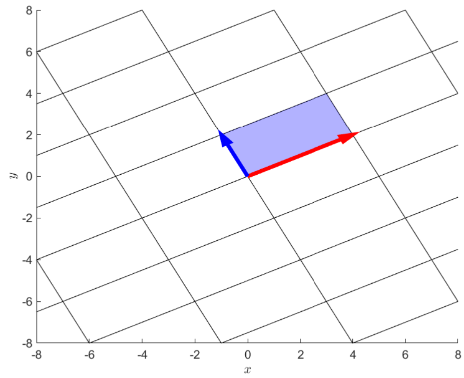


Figure 8: Alternative basis

The alternative basis arises from using the vectors in the columns of matrix $\boldsymbol{A}$ as a basis of $\mathbf{R}^2$.

4.2 Trace

Another value associated with square matrices is the trace. It is defined as follows.

Definition 19. The trace of an $n \times n$ matrix $\boldsymbol{A}$ is the sum of the diagonal elements

$$\operatorname{tr} \boldsymbol{A} = \sum_{i=1}^{n} a_{ii}.$$

It satisfies the following properties.

Properties 7. For two conformable matrices $\boldsymbol{A}$ and $\boldsymbol{B}$:

(i) $\operatorname{tr}(\boldsymbol{A} + \boldsymbol{B}) = \operatorname{tr} \boldsymbol{A} + \operatorname{tr} \boldsymbol{B}$;

(ii) $\operatorname{tr}(c\boldsymbol{A}) = c \operatorname{tr} \boldsymbol{A}$, for any $c \in \mathbf{R}$;

(iii) if $\boldsymbol{A} \in \mathbf{R}^{n\times p}$ and $\boldsymbol{B} \in \mathbf{R}^{p\times n}$, then

$$\operatorname{tr} \boldsymbol{AB} = \operatorname{tr} \boldsymbol{BA}.$$

(iv) $\operatorname{tr}(\boldsymbol{A}') = \operatorname{tr} \boldsymbol{A}$;

(v) $\operatorname{tr} c = c$, for $c \in \mathbf{R}$.

Properties (i) and (ii) make the trace a linear operator. Furthermore, in property (iii), the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ need not be square, nor do $\boldsymbol{AB}$ and $\boldsymbol{BA}$ need to be of the same dimension.

Exercise 4.7. Prove the properties above.

Exercise 4.8. Using property (iii), show that for three conformable matrices we have

$$\operatorname{tr}(\boldsymbol{ABC}) = \operatorname{tr}(\boldsymbol{CAB}) = \operatorname{tr}(\boldsymbol{BCA}).$$

Exercise 4.9. Referring to your computations in exercise 2.8 show that

$$\operatorname{tr}(\boldsymbol{A}'\boldsymbol{B}) = \operatorname{tr}(\boldsymbol{B}\boldsymbol{A}') = \operatorname{tr}(\boldsymbol{A}\boldsymbol{B}') = \operatorname{tr}(\boldsymbol{B}'\boldsymbol{A}).$$

4.3 Inverses

In Section 2.4 above, we discussed matrices $\boldsymbol{A}$ that operate on another matrix $\boldsymbol{B}$ to produce a new one $\boldsymbol{C}$ through the equation $\boldsymbol{AB} = \boldsymbol{C}$. Whatever operation is performed by $\boldsymbol{A}$ can be undone by its inverse matrix $\boldsymbol{A}^{-1}$ provided that such a matrix exists.

Definition 20. A matrix $\boldsymbol{A} \in \mathbf{R}^{n \times n}$ is invertible if there exists a matrix $\boldsymbol{A}^{-1} \in \mathbf{R}^{n \times n}$ such that

$$\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_n = \boldsymbol{A}\boldsymbol{A}^{-1}.$$

Therefore, if $\boldsymbol{AB} = \boldsymbol{C}$, then

$$\boldsymbol{A}^{-1}\boldsymbol{C} = \boldsymbol{A}^{-1}\boldsymbol{AB} = \boldsymbol{IB} = \boldsymbol{B}.$$

Note that for rectangular matrices it cannot be true that the *left inverse* is equal to the *right inverse*; the dimensions would not allow it. Rectangular matrices have inverses, but we do not discuss them here. If you are interested, please look up the “Moore-Penrose inverse”.

Exercise 4.10. Show (by multiplying $\boldsymbol{A}\boldsymbol{A}^{-1}$) that in general if $\boldsymbol{A}$ is a 2×2 matrix given by $\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, and $(ad - cb) \neq 0$ then $\boldsymbol{A}^{-1} = \frac{1}{ad-cb} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$, noting that $(ad - cb)$ is the determinant of $\boldsymbol{A}$.

Exercise 4.11. Write down the inverse of the matrices

$$\text{(i)} \quad \boldsymbol{A} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}, \quad \text{(ii)} \quad \boldsymbol{B} = \begin{bmatrix} 5 & -3 \\ 2 & 1 \end{bmatrix}, \quad \text{(iii)} \quad \boldsymbol{C} = \begin{bmatrix} 6 & 1 \\ 1 & 2 \end{bmatrix}.$$

Theorem 2. *If an inverse exists, then it is unique.*

Proof. Suppose not such that there exist two distinct inverses $\boldsymbol{B}$ and $\boldsymbol{C}$ for a matrix $\boldsymbol{A}$. Then,

$$\begin{aligned} \boldsymbol{B} = \boldsymbol{B}\boldsymbol{I} &= \boldsymbol{B}(\boldsymbol{A}\boldsymbol{C}) \\ &= (\boldsymbol{B}\boldsymbol{A})\boldsymbol{C} \\ &= \boldsymbol{I}\boldsymbol{C} = \boldsymbol{C}, \end{aligned}$$

a contradiction. ■

Theorem 3. *If there exists a non-null vector $\boldsymbol{x}$ such that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$, then $\boldsymbol{A}$ is not invertible.*

Proof. Suppose not such that there exists a non-null vector $\boldsymbol{x}$ with $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ and $\boldsymbol{A}$ is invertible. By definition definition 20, $\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x} = \boldsymbol{0}$, but $\boldsymbol{x}$ is non-null. Hence, a contradiction. ■

By Theorem 3, singular matrices take some non-zero vector into zero. There is no $\boldsymbol{A}^{-1}$ which can recover that vector.

Theorem 4. *If $\boldsymbol{A}$ is an $n \times n$ invertible matrix, then the system of equations*

$$\sum_{j=1}^{n} a_{ij}x_j = b_i \qquad (1 \leq i \leq n),$$

has a unique solution for each choice of b_i.

Exercise 4.12. Use Theorem 2 [uniqueness of an inverse] to prove Theorem 4 [unique solution for each choice of b_i].

Exercise 4.13. Construct a matrix $\boldsymbol{A}$ that multiplies the vector $(3, -1)'$ to produce the zero vector $(0, 0)'$. What do you notice about the matrix $\boldsymbol{A}$? Compute its determinant.

Theorem 5. *A square matrix $\boldsymbol{A}$ of order n is invertible if and only if $\operatorname{rk}(\boldsymbol{A}) = n$.*

Proof. See exercise 4.21 in Abadir and Magnus (2005, p. 83).

■

Corollary 5.1. *A square matrix $\boldsymbol{A}$ is invertible if and only if it is non-singular.*

Exercise 4.14. Consider the matrix

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 4 & 7 \\ 3 & 6 & 2 \end{bmatrix}.$$

Show, using Theorem 5 that the matrix $\boldsymbol{A}$ is not invertible.

Theorem 6. *The product $\boldsymbol{AB}$ has an inverse if and only if $\boldsymbol{A}$ and $\boldsymbol{B}$ are non-singular and of the same dimension.*

Exercise 4.15. Suppose that $\boldsymbol{A}$ and $\boldsymbol{B}$ are non-singular matrices of the same order. Show that

$$(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1},$$

and thereby proving Theorem 6.

The inverse of a matrix $\boldsymbol{A}$ satisfies the following properties.

Properties 8 (Inverse)**.** For a non-singular matrix $\boldsymbol{A}$:

(i) $(c\boldsymbol{A})^{-1} = (1/c)\boldsymbol{A}^{-1}$, for any $c \in \mathbf{R}_0$;

(ii) $(\boldsymbol{A}')^{-1} = (\boldsymbol{A}^{-1})'$;

(iii) $(\boldsymbol{A}^{-1})^{-1} = \boldsymbol{A}$.

Exercise 4.16. Prove the properties above.

4.3.1 Computing the inverse

Definition 20 does not immediately reveal how to compute the inverse. However, you could show that for an arbitrary $n \times n$ matrix $\boldsymbol{A}$, the formula for its inverse is

$$\boldsymbol{A}^{-1} = \frac{1}{\det \boldsymbol{A}}\boldsymbol{A}^{\#}, \tag{2}$$

where $\boldsymbol{A}^{\#}$ is the *adjoint* matrix of $\boldsymbol{A}$.

Definition 21. The adjoint matrix of $\boldsymbol{A}$ is the transpose of its cofactor matrix (recall definition 16).

Formally,

$$(\boldsymbol{A}^{\#})_{ij} = (-1)^{i+j} M_{ji}.$$

We see that (2) is consistent with corollary 5.1.

Exercise 4.17. Find the inverse of

$$\text{(i)} \quad \boldsymbol{A} = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 2 \\ 0 & 1 & 1 \end{bmatrix} \quad \text{(ii)} \quad \boldsymbol{B} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix} \quad \text{(iii)} \quad \boldsymbol{C} = \begin{bmatrix} -2 & 1.6 & -0.4 \\ 1.6 & -0.9 & 0.4 \\ 1.3 & -1.2 & 1.6 \end{bmatrix} \quad \text{Note: Harder.}$$

Exercise 4.18. Show that a matrix with a column (or row) of zeros is not invertible.

4.3.2 Special inverses (*)

The inverse of elementary matrices (introduced in section 2.4.1) are particularly simple to compute, and obey the following properties. These elementary matrices satisfy the following.

Properties 9 (Inverse of elementary matrices)**.** For an elementary matrix:

(i) $\boldsymbol{E}_{ij}^{-1} = \boldsymbol{E}_{ij}$;

(ii) $\boldsymbol{E}_i^{-1}(\gamma) = \boldsymbol{E}_i(\gamma^{-1})$;

(iii) $\boldsymbol{E}_i^{-1}(\gamma|j) = \boldsymbol{E}_i(-\gamma|j)$.

The proof of these results can be found in exercises 6.4 and 6.5 in Abadir and Magnus (2005, p. 134).

Exercise 4.19. Suppose $\boldsymbol{A}$ is invertible and you exchange its first two rows to obtain $\boldsymbol{B}$. Explain why the new matrix is invertible. How would you find $\boldsymbol{B}^{-1}$ from $\boldsymbol{A}^{-1}$.

4.3.3 Application: ISLM

A simple way to demonstrate how inverse of matrices may be useful in economics is to consider the standard closed economy ISLM model. Conveniently, the equilibrium of a simplified version of this model is often expressed in linear form as the system:

$$Y = C + I + G_0, \qquad \text{(Accounting identity)}$$

$$C = \alpha + \beta Y, \qquad \text{(Keynesian consumption function)}$$

$$I = \gamma - \delta i, \qquad \text{(Investment equation)}$$

$$M_0 = \varepsilon Y - \zeta i, \qquad \text{(Money market equilibrium)}$$

where Y, C, I and i are endogenous variables, G_0 and M_0 are exogenous variables and α, β, γ, δ, ε and ζ are structural parameters of the model. Typically two policy multipliers are of interest: the Keynesian government spending multiplier and the money multiplier. Noting that to simplify we may eliminate consumption, C, and the system may be re-written as:

$$\begin{bmatrix} 1-\beta & -1 & 0 \\ 0 & 1 & \delta \\ \varepsilon & 0 & -\zeta \end{bmatrix} \begin{bmatrix} Y \\ I \\ i \end{bmatrix} = \begin{bmatrix} \alpha \\ \gamma \\ 0 \end{bmatrix} + \begin{bmatrix} G_0 \\ 0 \\ M_0 \end{bmatrix}, \quad \text{or alternatively} \quad \boldsymbol{Ay} = \boldsymbol{a} + \boldsymbol{x}.$$

where $\boldsymbol{y}$ is a vector of the endogenous variables, and $\boldsymbol{x}$ is a vector of the exogenous variables. We are able to see immediately how computing an inverse will uncover the multipliers of interest in the form:

$$\boldsymbol{y} = \boldsymbol{A}^{-1}(\boldsymbol{a} + \boldsymbol{x}).$$

To do this we proceed with the recipe, and firstly compute the matrix of minors, $\boldsymbol{M}$, corresponding to the each element in the matrix $\boldsymbol{A}$:

$$\boldsymbol{M} = \begin{bmatrix} -\zeta & -\delta\varepsilon & -\varepsilon \\ \zeta & -\zeta(1-\beta) & \varepsilon \\ -\delta & \delta(1-\beta) & (1-\beta) \end{bmatrix}.$$

Next, we use cofactors and the transpose to convert this matrix into the adjoint matrix, $\boldsymbol{A}^{\#}$.

$$\boldsymbol{A}^{\#} = \begin{bmatrix} -\zeta & -\zeta & -\delta \\ \delta\varepsilon & \zeta(\beta-1) & \delta(\beta-1) \\ -\varepsilon & -\varepsilon & 1-\beta \end{bmatrix}.$$

Then calculate the determinant of the matrix

$$\det \boldsymbol{A} = (\beta - 1)\zeta - \delta\varepsilon$$

Before finally combining to give the inverse as:

$$\boldsymbol{A}^{-1} = \frac{1}{\det \boldsymbol{A}} \boldsymbol{A}^{\#} = \frac{1}{(\beta-1)\zeta - \delta\varepsilon} \begin{bmatrix} -\zeta & -\zeta & -\delta \\ \delta\varepsilon & \zeta(\beta-1) & \delta(\beta-1) \\ -\varepsilon & -\varepsilon & 1-\beta \end{bmatrix}$$

Hence we may read off the equation for output, Y, in terms of exogenous variables and structural parameters only as:

$$Y = \frac{\zeta(\alpha + \gamma + G_0)}{\delta\varepsilon + (1-\beta)\zeta} + \frac{\delta M_0}{\delta\varepsilon + (1-\beta)\zeta},$$

and deduce the multipliers as:

$$\frac{\zeta}{\delta\varepsilon + (1-\beta)\zeta} \quad \text{and} \quad \frac{\delta}{\delta\varepsilon + (1-\beta)\zeta}.$$

4.4 Eigenvalues and eigenvectors

In Section 2.4, we briefly discussed matrices as linear operators and linked matrices to functions, which take a vector $\boldsymbol{x}$ as input and produce as output another vector $\boldsymbol{Ax}$. In this section we are particularly interested in vectors whose direction is invariant to the linear transformation $\boldsymbol{A}$. This brings us to the concepts of eigenvalues and eigenvectors.

Definition 22. For a square matrix $\boldsymbol{A}$, the vector $\boldsymbol{x} \in \mathbf{R}^n \backslash \{\mathbf{0}\}$ and the corresponding scalar $\lambda \in \mathbf{R}$ are called an eigenvector and eigenvalue if they satisfy

$$\boldsymbol{Ax} = \lambda \boldsymbol{x}. \tag{3}$$

Remark. There is no reason why we should restrict the eigenvalues or eigenvectors to be real. Even for real matrices, the eigenvalues or eigenvectors can be complex, see example 4.6. However, in this course and the MPhil programme you will not come across complex valued eigenvalues and eigenvectors (often). Furthermore, in the definition above we have explicitly restricted the eigenvectors not to be the zero vector. Of course, $\boldsymbol{x} = \mathbf{0}$ always satisfies (3), but we rule out this trivial case.

Definition 22 immediately gives us a way to find the eigenvalues and eigenvectors. Rewrite (3) as

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

then from Theorem 3 we know that the matrix $\boldsymbol{A} - \lambda \boldsymbol{I}$ needs to be singular since we have ruled out $\boldsymbol{x} = \mathbf{0}$. Hence, we can find the eigenvalues by solving

$$\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0. \tag{4}$$

We refer to (4) as the characteristic polynomial. This is a polynomial in λ of order n. Hence, from the "fundamental theorem of algebra" there are n eigenvalues, but these are not necessarily distinct. The eigenvectors associated with each eigenvalue can then be found by solving the system in (3). From this we can see that the eigenvectors are unique up to scaling. It is therefore natural to write down the eigenvectors with unit length, $\|\boldsymbol{x}\| = 1$. These are called the *normalised* eigenvectors.

Example 4.4. Consider the permutation matrix in example 2.10. It is not difficult to spot two vectors $\boldsymbol{x}$ that yield $\boldsymbol{Ax} = \lambda \boldsymbol{x}$

$$\boldsymbol{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \Rightarrow \quad \boldsymbol{Ax} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \boldsymbol{x} \quad (\lambda = 1),$$

and

$$\boldsymbol{x} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad \Rightarrow \quad \boldsymbol{Ax} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -\boldsymbol{x} \quad (\lambda = -1).$$

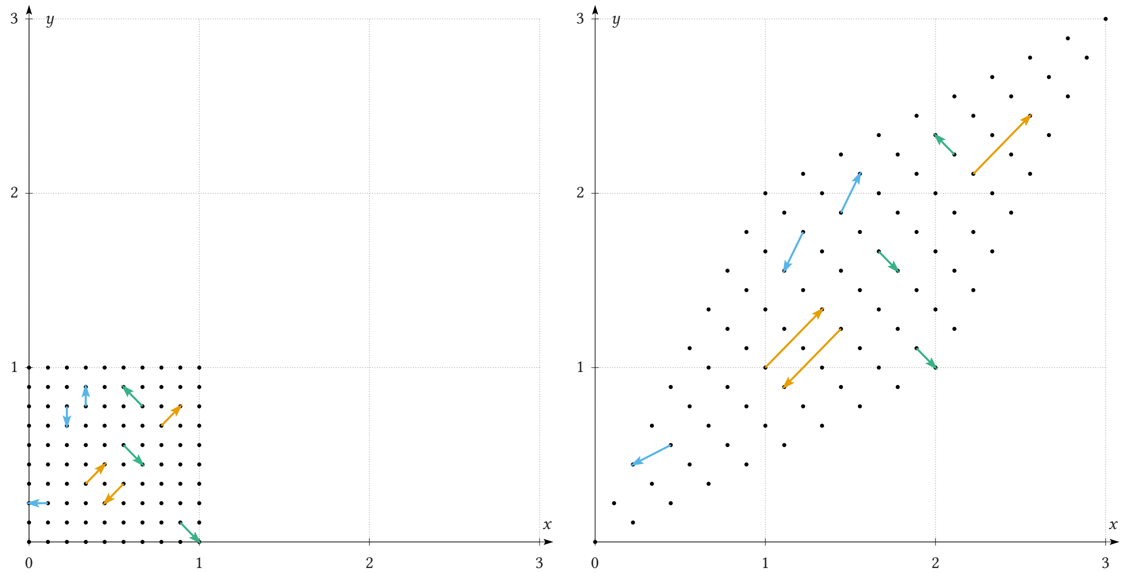


Figure 9: (a) 100 points on a grid in $[0, 1]^2$; (b) 100 points in $[0, 1]^2$ transformed by $\boldsymbol{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. The direction of the vectors parallel to $(1, -1)'$ (green) and $(1, 1)'$(orange) remains unchanged with the latter also being scaled by a factor of 3. The other vectors (blue) change direction after a transformation by $\boldsymbol{A}$.

Example 4.5. We can gain some intuition by looking at what the eigenvalues and eigenvectors mean graphically. Consider the linear transformation

$$\boldsymbol{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix},$$

which has eigenvectors

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

associated with the eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 3$, respectively. It is a good exercise to see if you can find these for yourself. As we can see in Figure 9, the direction of the vectors parallel to $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ remains unchanged after a transformation by $\boldsymbol{A}$. Furthermore, the vectors parallel to $\boldsymbol{x}_2$ are scaled by a factor of 3, the value of λ_2. Whereas the direction of the vectors not parallel to either of the eigenvectors changes.

Example 4.6 (Complex eigenvalues)**.** Consider the matrix

$$\boldsymbol{A} = \begin{bmatrix} 6 & -13 \\ 1 & 0 \end{bmatrix},$$

which has only real entries, but both its eigenvalues are complex. To see this,

$$\begin{aligned} \det(\boldsymbol{A} - \lambda \boldsymbol{I}) &= \lambda(\lambda - 6) + 13 \\ &= \lambda^2 - 6\lambda + 13. \end{aligned}$$

The discriminant of this polynomial is $\Delta = -16$, thus it has no real solutions. Hence,

$$\begin{aligned}\lambda_{1,2} &= \frac{6 \pm \sqrt{\Delta}}{2} \\ &= 3 \pm 2i.\end{aligned}$$

Exercise 4.20. Find the two eigenvalues of the matrix

$$\boldsymbol{B} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}.$$

Deduce two eigenvectors of $\boldsymbol{B}$ by inspection.

Exercise 4.21. Let $\boldsymbol{A}$ be the permutation matrix in example 4.4, and write the matrix $\boldsymbol{B}$ of exercise 4.20 as $\boldsymbol{B} = \boldsymbol{A} + 3\boldsymbol{I}$. Algebraically, deduce that the eigenvalues of $\boldsymbol{A}$ are three less than the eigenvalues of $\boldsymbol{B}$ and that the eigenvectors are unchanged.

Exercise 4.22. Show that:

(i) λ^2 is an eigenvalue of $\boldsymbol{A}^2$;

(ii) λ^{-1} is an eigenvalue of $\boldsymbol{A}^{-1}$;

(iii) $\lambda + 1$ is an eigenvalue of $\boldsymbol{A} + \boldsymbol{I}$.

Exercise 4.23. Show that the eigenvalues of idempotent matrices are either 0 or 1.

Exercise 4.24. Let

$$\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

(i) Show that $\det(\boldsymbol{A} - \lambda\boldsymbol{I}) = \lambda^2 - \operatorname{tr}(\boldsymbol{A})\lambda + \det\boldsymbol{A}$;

(ii) Find an expression for the two eigenvalues of $\boldsymbol{A}$ in terms of a, b, c and d. These characteristic roots satisfy

$$(\lambda - \lambda_1)(\lambda - \lambda_2) = \lambda^2 - \operatorname{tr}(\boldsymbol{A})\lambda + \det\boldsymbol{A} = 0,$$

and $\lambda_1\lambda_2 = \det\boldsymbol{A}$ and $\lambda_1 + \lambda_2 = \operatorname{tr}\boldsymbol{A}$.

4.4.1 Application: principal components (*)

Suppose that you have a data set with p variables and n observations on each of these variables and store them in the $n \times p$ matrix $\boldsymbol{X}$. These data points can be plotted in $\mathbf{R}^p$ space. A question we can ask is: "In which direction does the data vary the most?" To answer this, consider the *empirical* variance-covariance matrix $\tilde{X}'\tilde{X}/n$, where $\tilde{X} = X - \bar{X}$. We want to find the vectors which maximise this variance

$$\max_{v \in \mathbb{R}^p} v'\tilde{X}'\tilde{X}v \quad \text{subject to } \|v\| = 1. \tag{5}$$

It turns out that the solution to (5) is the eigenvector associated with the largest eigenvalue of $\tilde{X}'\tilde{X}$. We refer to this as the first principal component. The other $n-1$ principal components are the remaining eigenvectors with the last principal component being the eigenvector associated with the smallest eigenvalue.

Figure 10 plots the two principal components from a simulated data set with 10,000 observations drawn from a random vector

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 & -0.8 \\ -0.8 & 1 \end{bmatrix}\right).$$

There is a very large literature on Principal Components Analysis and for those of you who are interested, a good place to start is James et al. (2013).

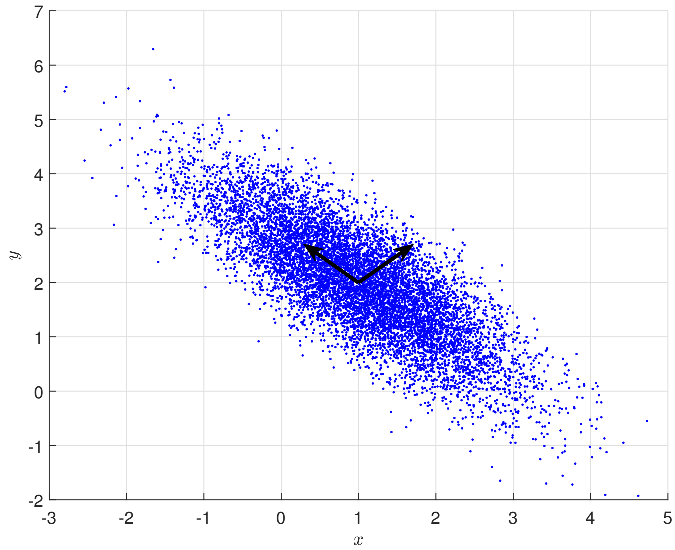


Figure 10: The two principal components from a simulated dataset.

5 Diagonalisation and powers of A

In this section we consider one of the many useful applications of eigenvalues and eigenvectors, called the eigendecomposition. Let $\boldsymbol{Q}$ be the matrix whose columns are the eigenvectors of $\boldsymbol{A}$. Suppose $\boldsymbol{A}$ possesses n linearly independent eigenvectors such that $\boldsymbol{Q}$ is invertible. By definition 22,

$$\boldsymbol{AQ} = \boldsymbol{A}\begin{bmatrix} | & & | \\ \boldsymbol{x}_1 & \dots & \boldsymbol{x}_n \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ \lambda_1\,\boldsymbol{x}_1 & \dots & \lambda_n\,\boldsymbol{x}_n \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ \boldsymbol{x}_1 & \dots & \boldsymbol{x}_n \\ | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} := \boldsymbol{Q\Lambda}.$$

Observe that by premultiplying by $\boldsymbol{Q}^{-1}$ we obtain

$$\boldsymbol{\Lambda} = \boldsymbol{Q}^{-1}\boldsymbol{AQ}.$$

Therefore, showing that $\boldsymbol{Q}^{-1}\boldsymbol{AQ}$ diagonalises $\boldsymbol{A}$, in the sense that it produces the diagonal matrix $\boldsymbol{\Lambda}$ with the eigenvalues $\lambda_1, \dots, \lambda_n$ along its main diagonal. Analogously, by postmultiplying by $\boldsymbol{Q}^{-1}$ we have the eigendecomposition

$$\boldsymbol{A} = \boldsymbol{Q\Lambda Q}^{-1}. \tag{6}$$

There is a wide variety of matrix decompositions, but this decomposition is powerful enough for our purposes and allows us to prove some interesting properties of square matrices.

Lemma 1. *Let $\boldsymbol{A}$ be diagonalisable according to* (6)*, then*

$$\det \boldsymbol{A} = \prod_{i=1}^{n} \lambda_i,$$

and

$$\operatorname{tr} \boldsymbol{A} = \sum_{i=1}^{n} \lambda_i.$$

Exercise 5.1. Prove lemma 1.

Remark. You could try and prove this directly from $\det(\boldsymbol{A} - \lambda \boldsymbol{I})$ showing that these results hold even for matrices which are not diagonalisable through the eigendecomposition. This is slightly more involved.

5.1 Difference Equations

Another interesting use for the eigendecomposition is computing powers of $\boldsymbol{A}$. We know that the kth power of a diagonal matrix $\boldsymbol{D}$ is

$$\boldsymbol{D}^k = \text{diag}(d_1^k, \dots, d_n^k).$$

From (6), it is easy to show that

$$\boldsymbol{A}^k = \boldsymbol{Q}\boldsymbol{\Lambda}^k\boldsymbol{Q}^{-1}. \tag{7}$$

This can lead to considerable computational improvements when k is large. Moreover, the following result will be useful in the lectures on difference equations.

Theorem 7. *Let $\boldsymbol{A}$ be a matrix of order n with n linearly independent eigenvectors, then $\boldsymbol{A}^k$ converges to the zero matrix if and only if the eigenvalues of $\boldsymbol{A}$ satisfy $|\lambda_i| < 1$ for all $i = 1, \dots, n$.*

Proof. Both directions follow easily from (7). ∎

Theorem 7 is important for analysing the convergence properties of certain multivariate difference equations of the form

$$\boldsymbol{u}_{k+1} = \boldsymbol{A}\boldsymbol{u}_k,$$

and by recursive substitution we have

$$\boldsymbol{u}_{k+1} = \boldsymbol{A}^{k+1}\boldsymbol{u}_0,$$

for some initialisation vector $\boldsymbol{u}_0$.

5.1.1 Application: labour markets

In labour economics, the total population (labour force) of an economy may be characterised into two states. At any given time period-t each person may be employed, e_t or unemployed, u_t. The movement between each state may also be classified. The probability that a worker moves from employment to unemployment is known as the job separation rate, s, while the probability of movement from unemployment to employment is known as the job finding rate, f. Collectively these labour market flow rates are often referred to as labour market 'churn'. The dynamic relationship between these variables may therefore be described as

$$\begin{bmatrix} u_{t+1} \\ e_{t+1} \end{bmatrix} = \begin{bmatrix} 1-f & s \\ f & 1-s \end{bmatrix} \begin{bmatrix} u_t \\ e_t \end{bmatrix},$$

which we will summarise in the form

$$u_{t+1} = Au_t,$$

An interesting policy question is therefore, given a starting composition of the labour force, $\boldsymbol{u}_0$, what will the labour force composition look like in 2 years, $\boldsymbol{u}_2$, 5 years, $\boldsymbol{u}_5$, and in the 'long run'? A concrete example will help to answer these questions.[3]

Example 5.1. Suppose the economy currently has 10 unemployed workers and 90 employed workers. The job separation rate is 0.05 while the job finding rate is 0.5. We note that this infers the main matrix of interest, $\boldsymbol{A}$ may be written as

$$\boldsymbol{A} = \begin{bmatrix} 0.5 & 0.05 \\ 0.5 & 0.95 \end{bmatrix},$$

We apply the procedure and diagonalise the matrix as follows. Firstly calculate the eigenvalues and eigenvectors of the matrix.

$$\begin{aligned} \det(\boldsymbol{A} - \lambda \boldsymbol{I}) &= 0, \\ (0.5 - \lambda)(0.95 - \lambda) - 0.025 &= 0, \\ 0.45 - 1.45\lambda + \lambda^2 &= 0, \end{aligned}$$

which infers $\lambda_1 = 1$ and $\lambda_2 = 0.45$ and in turn these imply $\boldsymbol{v}_1 = \begin{bmatrix} 0.1 \\ 1 \end{bmatrix}$, and $\boldsymbol{v}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$. Such that we have

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{-1} = \begin{bmatrix} 0.1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0.45 \end{bmatrix} \begin{bmatrix} \frac{10}{11} & \frac{10}{11} \\ -\frac{10}{11} & \frac{1}{11} \end{bmatrix},$$

as the diagonalised form of the matrix $\boldsymbol{A}$. Noting that we are given the initial vector, $\boldsymbol{u}_0 = \begin{bmatrix} 10 \\ 90 \end{bmatrix}$, we may now turn to the economics questions of interest. With the time period set to a year.

$$\begin{aligned} \boldsymbol{u}_2 = \boldsymbol{A}^2\boldsymbol{u}_0 = \boldsymbol{Q}\boldsymbol{\Lambda}^2\boldsymbol{Q}^{-1} &= \begin{bmatrix} 0.1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0.45 \end{bmatrix}^2 \begin{bmatrix} \frac{10}{11} & \frac{10}{11} \\ -\frac{10}{11} & \frac{1}{11} \end{bmatrix} \begin{bmatrix} 10 \\ 90 \end{bmatrix} = \begin{bmatrix} 9.275 \\ 90.725 \end{bmatrix}, \\ \boldsymbol{u}_5 = \boldsymbol{A}^5\boldsymbol{u}_0 = \boldsymbol{Q}\boldsymbol{\Lambda}^5\boldsymbol{Q}^{-1} &= \begin{bmatrix} 0.1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0.45 \end{bmatrix}^5 \begin{bmatrix} \frac{10}{11} & \frac{10}{11} \\ -\frac{10}{11} & \frac{1}{11} \end{bmatrix} \begin{bmatrix} 10 \\ 90 \end{bmatrix} = \begin{bmatrix} 9.1077 \\ 90.8923 \end{bmatrix} \end{aligned}$$

The behaviour of the system in the "long run" may be seen as we take the limit of the dynamic system as $t \longrightarrow \infty$ and the associated vector $\boldsymbol{u}_\infty$. Having diagonalised the matrix this behaviour is easy to note,

[3] Notice that the matrix being used here, $\boldsymbol{A}$, is actually known as a Markov matrix, as all entries are non-negative and each column sum is 1.

as the limit of each eigenvalue is take in turn. As such we have that

$$\boldsymbol{u}_\infty = \begin{bmatrix} 0.1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}^2 \begin{bmatrix} \frac{10}{11} & \frac{10}{11} \\ -\frac{10}{11} & \frac{1}{11} \end{bmatrix} \begin{bmatrix} 10 \\ 90 \end{bmatrix} = \begin{bmatrix} 9.0909 \\ 90.9091 \end{bmatrix},$$

and we conclude that the initial unemployment rate, of 10%, was above its equilibrium value. The dynamics clearly show a smooth transition back. A graph of this situation is shown in Figure 11. Of course we could also have been clever in the final example an noted that the "long run" will be characterised by an equation of the form

$$\boldsymbol{u}_\infty = \boldsymbol{A}\boldsymbol{u}_\infty,$$

and therefore claimed that the "long run" distribution of workers between states will be the eigenvector associated with the unitary eigenvalue.

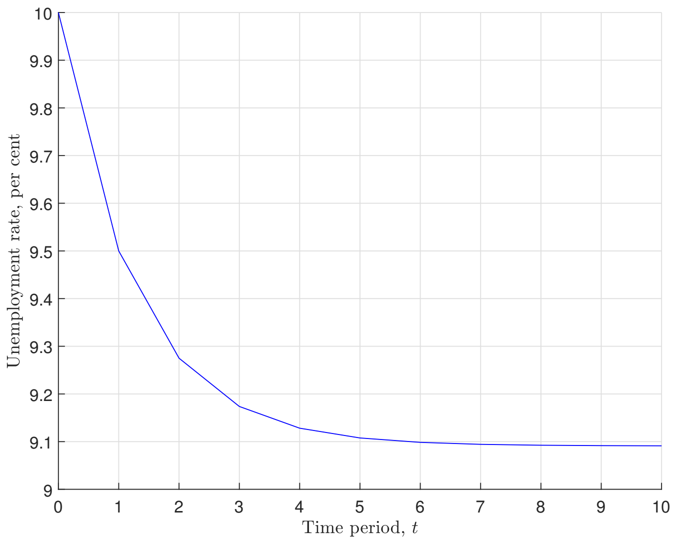


Figure 11: Transition path for the unemployment rate.

5.1.2 Application: vector autoregression

Vector autoregressions (VARs) are a popular econometrics tool, used by academics, policymakers and in business to analyse complex economic phenomena in a time series context. The core concept required to apply these models makes use of the diagonalised form of a matrix. A vector autoregression (VAR) may be written in the following form

$$\boldsymbol{x}_t = \boldsymbol{A}\boldsymbol{x}_{t-1} + \boldsymbol{B}\boldsymbol{e}_t$$

where $\boldsymbol{x}_t$ is a vector of economic variables (data), $\boldsymbol{A}$ is a matrix of coefficients describing the evolution of the economy, $\boldsymbol{e}_t$ represent structural economic shocks of interest and $\boldsymbol{x}_0 = 0$. A typical question is to ask how the economy responds after a one-time impulse (typically normalised to one standard deviation) that happens in period $t = 1$. This is known as an impulse response function and can be mapped quickly and efficiently using the equation

$$\boldsymbol{x}_{t+n} = \boldsymbol{A}^n \boldsymbol{B} \boldsymbol{e}_1 = \boldsymbol{Q} \boldsymbol{\Lambda}^n \boldsymbol{Q}^{-1} \boldsymbol{B} \boldsymbol{e}_1$$

after first using econometric techniques to estimate the matrices of interest from the data.

5.1.3 Application: Fibonacci sequence (*)

The following example comes from Strang (2009).

Example 5.2. The first 8 numbers of the Fibonacci sequence are $0, 1, 1, 2, 3, 5, 8, 13, \ldots$ where the kth element of the sequence can be deduced from the $(k-1)$th and $(k-2)$th elements as $F_k = F_{k-1} + F_{k-2}$. One useful question to ask is: "How fast are the Fibonacci numbers growing?" The answer to this question lies in the eigenvalues. Let

$$\boldsymbol{u}_k = \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix}.$$

Then, the Fibonacci sequence satisfies

$$\boldsymbol{u}_{k+1} = \boldsymbol{A} \boldsymbol{u}_k \quad \text{with} \quad \boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

We have

$$\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = \lambda^2 - \lambda - 1.$$

Setting this to zero and solving gives

$$\lambda_1 = \frac{1}{2}(1 + \sqrt{5}) \quad \text{and} \quad \lambda_2 = \frac{1}{2}(1 - \sqrt{5}).$$

Since the largest eigenvalue is the one controlling the growth, the Fibonacci numbers are growing at the rate $(1 + \sqrt{5})/2$.

Exercise 5.2. Assuming you can write the initialisation vector $\boldsymbol{u}_0$ as a linear combination of the eigenvectors, find an expression for the kth element of an arbitrary sequence of the form $\boldsymbol{u}_{k+1} = \boldsymbol{A} \boldsymbol{u}_k$.

Exercise 5.3. Using the expression you found in exercise 5.2, find the 100th element of the Fibonacci sequence in example 5.2.

5.2 Symmetric and positive definite matrices

We know from Section 2.2 that symmetric matrices are equal to their transpose. Let us first remind you what an orthogonal matrix is.

Definition 23. A square matrix $\boldsymbol{Q}$ of order n is orthogonal if for any two columns

$$\boldsymbol{q}_i'\boldsymbol{q}_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}.$$

Thus, an orthogonal matrix has orthonormal columns.

Then in relation to the eigendecomposition, we have the following result for symmetric matrices.

Theorem 8 (Spectral decomposition)**.** *Let $\boldsymbol{A}$ be a symmetric $n \times n$ matrix, then there exists a matrix $\boldsymbol{Q}$ with $\boldsymbol{Q}^{-1} = \boldsymbol{Q}'$ (i.e. $\boldsymbol{Q}$ is orthogonal) and a diagonal matrix, $\boldsymbol{\Lambda}$, with diagonal elements the eigenvalues of $\boldsymbol{A}$ such that*

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}'.$$

We then refer to the eigendecomposition of $\boldsymbol{A}$ as the spectral decomposition.

Proof. See exercise 7.46 in Abadir and Magnus (2005, p. 177). ■

Next, we ask: "what does $\boldsymbol{A} > \boldsymbol{O}$ mean?" This question is answered by the the definiteness of a matrix which generalises the notion of positivity and negativity to matrices. Together with the eigendecomposition, the definiteness lets us prove many interesting results.

Definition 24 (Definiteness)**.** A square matrix $\boldsymbol{A}$ is called positive definite if for any $\boldsymbol{x} \in \mathbf{R}^n \backslash \{\mathbf{0}\}$,

$$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0. \tag{8}$$

The matrix $\boldsymbol{A}$ is positive semi-definite if (8) holds with weak inequality. The notion of negative and negative semi-definite is similarly defined by reversing the inequalities. A matrix which is neither positive or negative definite is indefinite.

Definition 24 formalises statements like: matrix $\boldsymbol{A}$ is larger than matrix $\boldsymbol{B}$, in the sense that $\boldsymbol{A} - \boldsymbol{B}$ is positive (semi)definite. Here, we will solely focus on positive (semi)definiteness for ease of exposition.

Lemma 2. *A symmetric matrix is positive definite if and only if all its eigenvalues are positive. Moreover, a symmetric matrix is positive semi-definite if and only if all its eigenvalues are non-negative.*

Proof. Left as an exercise. ■

Lemma 2 allows us to check for positive definiteness by checking the signs of the eigenvalues. If a symmetric matrix has two eigenvalues with different signs then it is indefinite. Some of the properties below follow directly from this lemma and are left as an exercise.

Properties 10 (Positive definite)**.** For a symmetric positive definite matrix $\boldsymbol{A}$:

(i) $\det \boldsymbol{A} > 0$ such that it is also non-singular;

(ii) $\operatorname{tr} \boldsymbol{A} > 0$;

(iii) if $\boldsymbol{A} \geq \boldsymbol{O}$ and $\boldsymbol{B} \geq \boldsymbol{O}$, then $\boldsymbol{A} + \boldsymbol{B}$ is at least positive semi-definite.

(iv) $\boldsymbol{A} > \boldsymbol{O}$ if and only if $\boldsymbol{A}^{-1} > \boldsymbol{O}$.

Exercise 5.4. Show the properties above.

Exercise 5.5. Show that there exists a matrix which is both positive semi-definite and negative semi-definite, but that a matrix cannot be positive and negative definite at the same time.

Exercise 5.6. Let $\boldsymbol{\Omega}$ be a symmetric positive definite matrix. Use the spectral decomposition to show that there exists a non-singular matrix $\boldsymbol{L}$ such that $\boldsymbol{\Omega} = \boldsymbol{L}'\boldsymbol{L}$ and such that $\boldsymbol{L}^{-1} = (\boldsymbol{L}')^{-1}$. What is $\boldsymbol{L}^{-1}$?

Exercise 5.7. Find the $\boldsymbol{X}$ that solves

$$2\boldsymbol{X}^2 - 3\boldsymbol{X} + \boldsymbol{I}_n = \boldsymbol{O}_n,$$

where $\boldsymbol{X}$ is a real symmetric matrix.

6 Vector and matrix calculus

In this section, we present a brief foray into vector and matrix calculus. Please consult other textbooks if you want to learn more; see e.g. Abadir and Magnus (2005); Magnus and Neudecker (1999).

Broadly there are two learning objectives to this section of the course:

(i) Ability to compute the Jacobian and Hessian matrices;

(ii) Knowledge of what the Jacobian and Hessian matrices represent.

The following notation is maintained throughout: let φ, $\boldsymbol{f}$ and $\boldsymbol{F}$ respectively represent scalar, vector and matrix-valued functions possibly taking a scalar, vector or matrix as their inputs. For example,

$$\begin{aligned} \varphi(x) &= \sin(x), & \varphi(\boldsymbol{x}) &= \sqrt{\boldsymbol{x}'\boldsymbol{x}}, & \varphi(\boldsymbol{X}) &= \det \boldsymbol{X} \\ \boldsymbol{f}(x) &= (x, x-3)', & \boldsymbol{f}(\boldsymbol{x}) &= \boldsymbol{A}\boldsymbol{x}, & \boldsymbol{f}(\boldsymbol{X}) &= \boldsymbol{X}\boldsymbol{a} \\ \boldsymbol{F}(x) &= x\boldsymbol{A}, & \boldsymbol{F}(\boldsymbol{x}) &= \boldsymbol{x}\boldsymbol{x}', & \boldsymbol{F}(\boldsymbol{X}) &= \boldsymbol{X}'\boldsymbol{X}. \end{aligned}$$

6.1 The Jacobian and the Hessian

Our first step in using vector and matrix calculus will be to introduce the Jacobian and Hessian matrices, and their computation.

Definition 25 (Jacobian matrix)**.** Let $\boldsymbol{f}$ be an $m \times 1$ vector function with input an n-vector $\boldsymbol{x}$, then define the $m \times n$ Jacobian matrix as

$$D\boldsymbol{f}(\boldsymbol{x}) := \frac{\mathrm{d}\boldsymbol{f}(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}'} = \left[\frac{\partial f_i(\boldsymbol{x})}{\partial x_j}\right].$$

This matrix collects all the first-order partial derivatives of the components of $\boldsymbol{f}$ into an $m \times n$ matrix. Note that (rather unhelpfully) the literature refers to both the Jacobian matrix and its determinant as the Jacobian.

Definition 26 (Hessian matrix)**.** Let $\varphi(\boldsymbol{x})$ be a scalar-valued function with input an n-vector $\boldsymbol{x}$, then the Hessian matrix is the $n \times n$ matrix of second-order partial derivatives

$$\boldsymbol{H}(\boldsymbol{x}) := \begin{bmatrix} \frac{\partial^2 \varphi(\boldsymbol{x})}{\partial x_1^2} & \frac{\partial^2 \varphi(\boldsymbol{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 \varphi(\boldsymbol{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 \varphi(\boldsymbol{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 \varphi(\boldsymbol{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 \varphi(\boldsymbol{x})}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \varphi(\boldsymbol{x})}{\partial x_n \partial x_1} & \frac{\partial^2 \varphi(\boldsymbol{x})}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 \varphi(\boldsymbol{x})}{\partial x_n^2} \end{bmatrix}.$$

Exercise 6.1. Assume $\varphi(\boldsymbol{x}) = \varphi(x_1, x_2) = 2x_1^2 x_2^3$, find the Jacobian and Hessian matrix of this (scalar) function (which has the vector input $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$).

Exercise 6.2. Assuming an input an n-vector $\boldsymbol{x}$, find both the Jacobian and Hessian matrix of the scalar function $\varphi(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{x}$.

Exercise 6.3. Assume $\boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix} x + \cos(y) \\ y + \cos(x) \end{bmatrix}$ and find the Jacobian of this function.

6.1.1 Geometric interpretation of the Jacobian matrix

We already know that the Jacobian stores all of the partial derivatives of a function, when differentiating with respect to a vector. It may be viewed as *the ratio of an infinitesimal change in the variables of one coordinate system to another.* You will all have already encountered the Jacobian matrix when using the change of variable approach to integration.

Example 6.1. Consider the integral $\int_0^1 6x^2\sqrt{2x^3+5}dx$. We pick an easy example, such that we already know how to approach the answer.

Using the change of variable approach, we may transform the integral as follows. Let $u = 2x^3 + 5$, such that $du = 6x^2dx$, such that the transformed integral will be given by $\int_5^7 u^{\frac{1}{2}}dx$, where we also remember to change the limits of integration! The integral looks considerably easier.

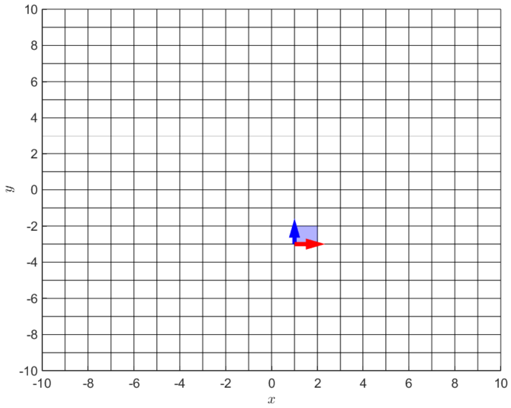
Notice, to make this substitution we invoke the Jacobian of the problem and, in this example, we have that $\frac{du}{dx} = 6x^2$.

Example 6.2. We wish to extend the approach in Example 6.1 away from the scalar setting, and will use the function given in Exercise 6.3. The function

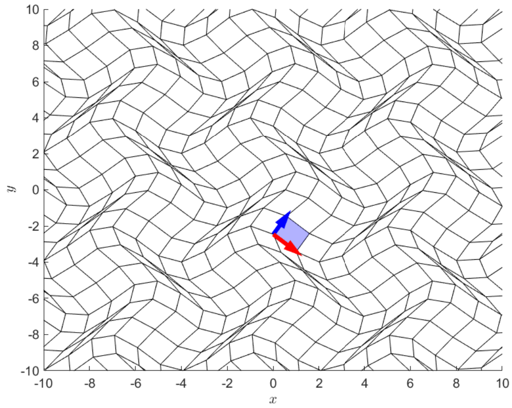
$$\boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \begin{bmatrix} x + \cos(y) \\ y + \cos(x) \end{bmatrix}$$

may be viewed as a way to map every element in the $(x, y)-$coordinate system into an alternative $(f_1, f_2)-$coordinate system. In the current example we may explicitly perform this change of variables, as shown in Figure 12, which maps a series of points to their position after the function $\boldsymbol{f}$ has been applied.

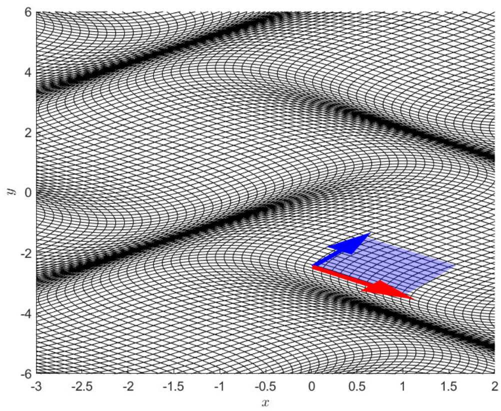
The interpretation of the Jacobian is then more straightforward to understand. Consider a small change in x, call this dx. After the function, $\boldsymbol{f}(\boldsymbol{x})$, has been applied this initial change in x results in a change in both f_1 and f_2 (shown in red on the Figure). Similarly, an initially small change in y, call this dy will result in a small change in both f_1 and f_2 (shown in blue on the Figure). The Jacobian matrix records these changes, for any value of (x, y). We plot these as changes from the point $x = [1, -3]'$.



(a) Standard Basis



(b) Non-linear Transformation



(c) Non-linear Transformation (Zoom)

Figure 12: A set of point in the $(x, y)-$coordinate system (a), and their equivalent position after the function f has been applied (b).

6.1.2 Geometric interpretation of the Hessian matrix

We now turn to the geometric representation of the Hessian matrix. As with the Jacobian, we already know that the Hessian stores all of the second-order partial derivates of a scalar-valued function and therefore describes the *local curvature of a function.* Stationary (or turning) points may be described as (local) minimum, (local) maximum or inflection point. By using available information in the Hessian, we are able to identify each type of stationary point.

Example 6.3. To demonstrate how a Hessian matrix may identify different types of stationary point, we again proceed by starting with a scalar example. Consider the three functions and their associated (scalar) Hessian matrices:

$$\begin{aligned} \varphi_1(x) &= x^2, \quad \boldsymbol{H}_1(x) = 2, \\ \varphi_2(x) &= -x^2, \quad \boldsymbol{H}_2(x) = -2, \\ \varphi_3(x) &= x^3, \quad \boldsymbol{H}_3(x) = 6x, \end{aligned}$$

Observe that the nature of the single stationary point which arises in all three cases, at $x = 0$, differs substantially. In the first case (with $\boldsymbol{H}(x) > 0$) is the stationary point is the (global) minimum point, in the second (with $\boldsymbol{H}(x) < 0$) it is the (global) maximum point, and in the third it is a point of inflection ($\boldsymbol{H}(x)$ evaluated at the stationary point $x = 0$ gives $\boldsymbol{H}(0) = 0$) . These functions, and their stationary points, are shown in Figure 14.

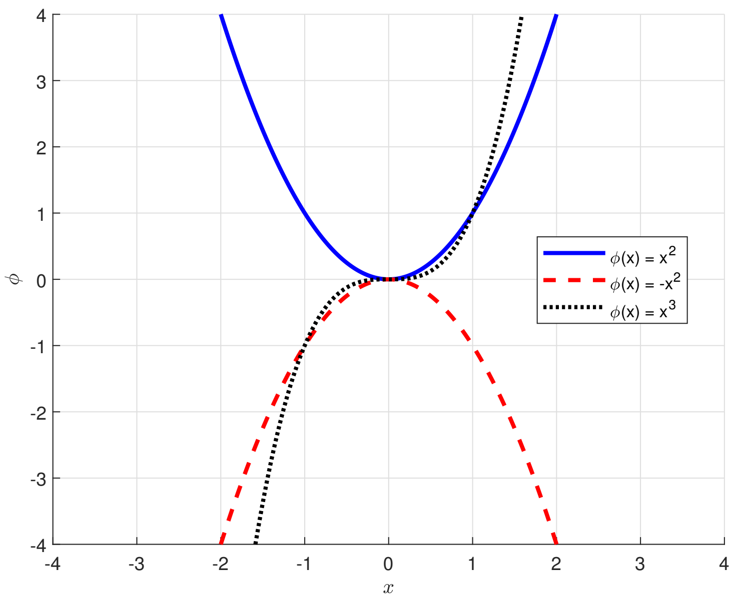


Figure 13: Three very different scalar functions.

Example 6.4. Extending the analysis of Example 6.3 to additional dimensions is straightforward. We consider the following scalar-valued functions, and their Hessians, which take the vector $\boldsymbol{x} = [x, y]$ as an input

$$\varphi_1(\boldsymbol{x}) = x^2 + y^2, \quad \boldsymbol{H}_1(\boldsymbol{x}) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix},$$

Minimum point

$$\varphi_2(\boldsymbol{x}) = -x^2 - y^2, \quad \boldsymbol{H}_2(\boldsymbol{x}) = \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix},$$

Maximum point

$$\varphi_3(\boldsymbol{x}) = x^3 + y^3, \quad \boldsymbol{H}_3(\boldsymbol{x}) = \begin{bmatrix} 6x & 0 \\ 0 & 6y \end{bmatrix},$$

Inflection point

Again we notice that the single stationary point is the same in all three cases, and arises at $\boldsymbol{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. In the first case this stationary point is the (global) minimum point, in the second it is the (global) maximum point, and in the third it is a point of inflection. These functions, and their stationary points, are shown in Figure 14.

Now, as the Hessian is a matrix it may not be evaluated as being either positive or negative. However the concepts introduced in Section 5.2 gives us a procedure which we may use to identify related properties of the Hessian matrices, namely:

(i) $H_1(\boldsymbol{x})$ is positive definite. We may show this by examining the eigenvalues of the matrix

$$\det(\boldsymbol{H}_1(\boldsymbol{x}) - \lambda \boldsymbol{I}_2) = (2 - \lambda)^2$$

which gives $\lambda = 2$ (twice). As $\lambda > 0$, $\boldsymbol{H}_1(\boldsymbol{x})$ is positive definite, telling us that the function $\varphi_1(\boldsymbol{x})$ is convex, and we have indeed reached a (global) minimum point;

(ii) $H_2(\boldsymbol{x})$ is negative definite. We may show this by examining the eigenvalues of the matrix

$$\det(\boldsymbol{H}_2(\boldsymbol{x}) - \lambda \boldsymbol{I}_2) = (-2 - \lambda)^2$$

which gives $\lambda = -2$ (twice). As $\lambda < 0$, $\boldsymbol{H}_2(\boldsymbol{x})$ is negative definite, telling us that the function $\varphi_2(\boldsymbol{x})$ is concave, and we have indeed reached a (global) maximum point;

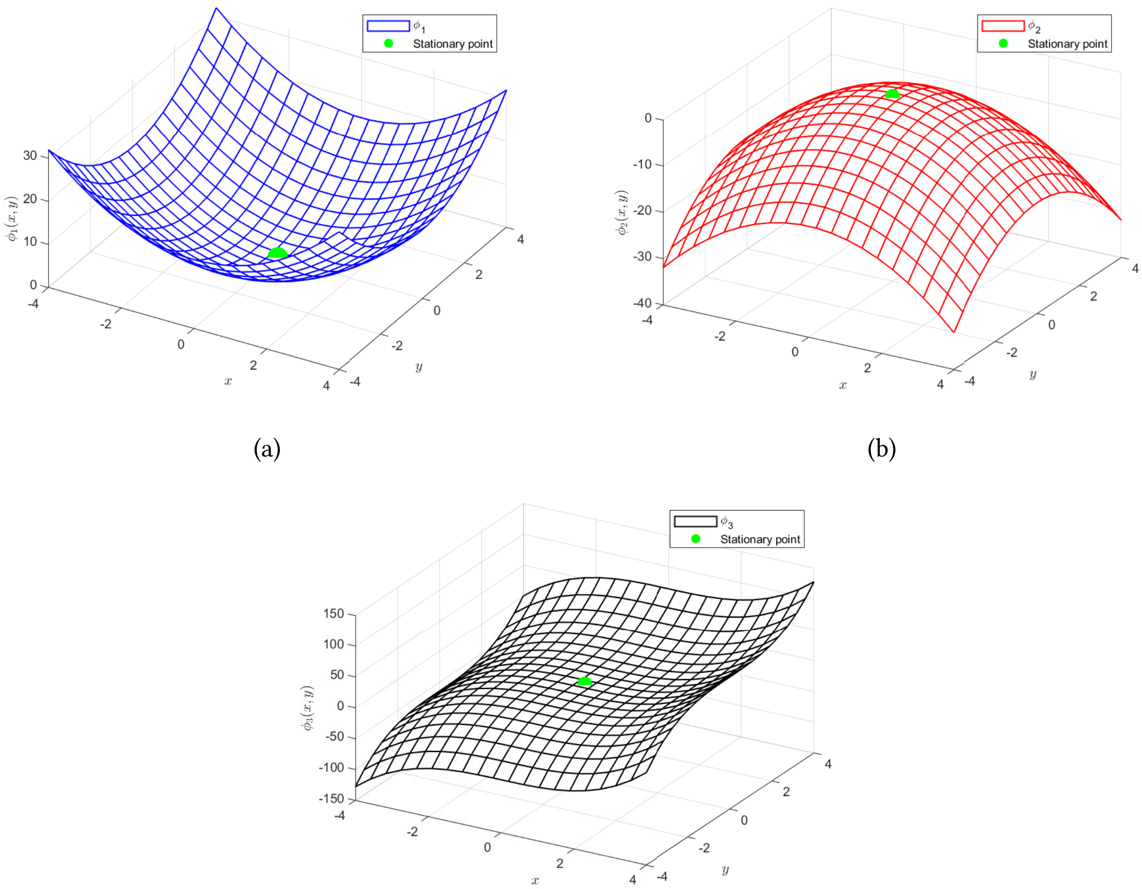
(iii) Finally, $H_3(\boldsymbol{x})$ evaluated at the stationary point gives $H_3(\boldsymbol{x}) = \mathbf{O}$, with $\det(H_3(\boldsymbol{x})) = 0$. This infers

$$\det(\boldsymbol{H}_3(\boldsymbol{x}) - \lambda \boldsymbol{I}_2) = \lambda^2$$

which gives $\lambda = 0$ (twice). As $\lambda = 0$, $\boldsymbol{H}_3(\boldsymbol{x})$ is neither positive definite nor negative definite at the stationary point, telling us that the function $\varphi_3(\boldsymbol{x})$ is reaches a point of inflection;

6.1.3 Application: cost minimisation (*)

We may use the tools of matrix calculus along with the concepts of positive definiteness to analyse the solution to a cost minimisation problem. Production levels, $\boldsymbol{x}$, are inputs to a firms cost function, $C(\boldsymbol{x})$. The firm also faces a linear production constraint $E(\boldsymbol{x})$. For the solution to the cost-minimisation



(a) (b)

(c)

Figure 14: Functions associated with Example 6.4. (a) $\varphi_1(\boldsymbol{x})$; (b) $\varphi_2(\boldsymbol{x})$; (c) $\varphi_3(\boldsymbol{x})$.

problem to be well-defined we need to ensure the cost function is concave. It turns out that the function will be concave if we have a positive definite Hessian matrix.

Example 6.5. Assume a firm has a cost function of with two arguments, $C(x, y)$, and must minimise this cost function, subject to the production requirement, $E(x, y) \geq 30$. Where

$$C(x, y) = 4x^2 + 4y^2 - 2xy - 40x - 140y + 1600,$$

$$E(x, y) = x + y.$$

Step 1: Jacobian 求偏导数

Before solving the constrained optimisation problem, we firstly inspect the cost function and solve the unconstrained cost minimisation problem. We may calculate the derivative, second derivative and cross-derivative of each element as

$$\frac{\partial C(x,y)}{\partial x} = 8x - 2y - 40, \quad \frac{\partial C(x,y)}{\partial y} = 8y - 2x - 140, \quad \frac{\partial^2 C(x,y)}{\partial x^2} = \frac{\partial^2 C(x,y)}{\partial y^2} = 8 \quad \text{and} \quad \frac{\partial^2 C(x,y)}{\partial xy} = -2.$$

This allows us to find the Jacobian of the problem as

$$DC(x, y) = \begin{bmatrix} 8x - 2y - 40, & 8y - 2x - 140 \end{bmatrix}$$

Step2: 让偏导数=0，为极大值极小值or infection point，求出 使偏导数等于0的x 和 y

Stationary points arise when $DC(x, y) = 0$, in this case we have a single stationary point at $x = 10$, $y = 20$. Whether this point represents a maximum, minimum or saddle point is yet to be determined. The cost function will be concave if it is associated with a positive definite Hessian matrix, and this point will then be the global minimum. The derivatives found earlier may be combined to show that the Hessian of the cost function may be written as

$$\boldsymbol{H}(x, y) = \begin{bmatrix} \frac{\partial^2 C(x,y)}{\partial x^2} & \frac{\partial^2 C(x,y)}{\partial x \partial y} \\ \frac{\partial^2 C(x,y)}{\partial y \partial x} & \frac{\partial^2 C(x,y)}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 8 & -2 \\ -2 & 8 \end{bmatrix}.$$

Step3:用 Hessian判断curvture， positive definite negative definite or infection

By inspection Lemma 2 applies, as the matrix is symmetric and therefore positive definiteness may be determined from an inspection of the eigenvalues of the matrix. To show that the Hessian matrix is positive definite, we calculate the eigenvalues and show these to be positive.

$$\det(\boldsymbol{H}(x, y) - \lambda \boldsymbol{I}_2) = \begin{vmatrix} 8 - \lambda & -2 \\ -2 & 8 - \lambda \end{vmatrix} = (8 - \lambda)^2 - 4 = 0, \quad \text{giving} \quad \lambda = 10 \quad \text{and} \quad \lambda = 6.$$

The cost function therefore has a unique minimum at its stationary point. In this problem the constraint plays little role, as the global minimum is attainable while respecting the constraint. Figure 15 shows a graphical representation of this minimisation problem.

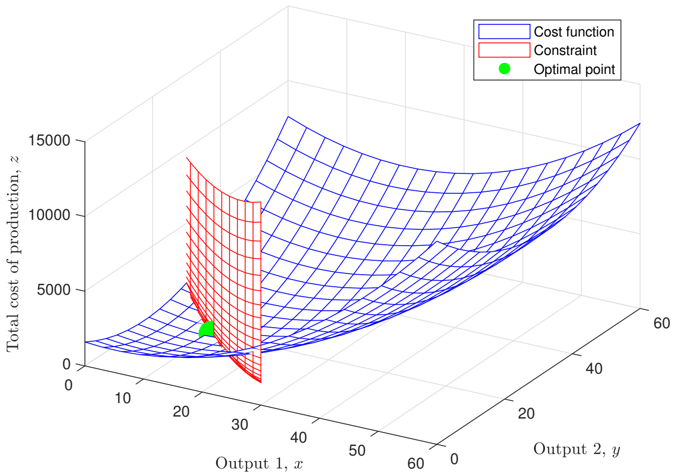


Figure 15: Cost function and constrained minimisation solution.

7 References

Abadir, K. M. and J. R. Magnus (2005). Matrix Algebra. In K. M. Abadir, J. R. Magnus, and P. C. B. Phillips (Eds.), *Econometric Exercises*, Chapter 1. New York: Cambridge University Press.

Binmore, K. and J. Davies (2002). *Calculus: Concepts and Methods.* Cambridge University Press.

James, G., D. Witten, T. Hastie, and R. Tibshirani (2013). *An Introduction to Statistical Learning.* New York: Springer (Science and Business Media).

Magnus, J. R. and H. Neudecker (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics* (Second ed.). Chichester/New York: John Wiley and Sons.

Strang, G. (2009). *Introduction to Linear Algebra* (Fourth ed.). Cambridge, Massachusetts: Wellesley.

A Summary of key results

For conformable matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$:

(1) $\boldsymbol{A} = [(\boldsymbol{A})_{ij}] \Leftrightarrow \boldsymbol{A}' = [(\boldsymbol{A})_{ji}]$;

(2) $c\boldsymbol{A} = [c(\boldsymbol{A})_{ij}]$;

(3) $(\boldsymbol{A}+\boldsymbol{B})_{ij} = (\boldsymbol{A})_{ij} + (\boldsymbol{B})_{ij}$;

(4) $\boldsymbol{A}+\boldsymbol{B} = \boldsymbol{B}+\boldsymbol{A}$;

(5) $c(\boldsymbol{A}+\boldsymbol{B}) = c\boldsymbol{A}+c\boldsymbol{B}$, for any $c \in \mathbf{R}$;

(6) $\boldsymbol{A}+(\boldsymbol{B}+\boldsymbol{C}) = (\boldsymbol{A}+\boldsymbol{B})+\boldsymbol{C}$;

(7) $\boldsymbol{a} \cdot \boldsymbol{b} = \boldsymbol{b} \cdot \boldsymbol{a}$;

(8) $\boldsymbol{a} \cdot (\boldsymbol{b}+\boldsymbol{c}) = \boldsymbol{a} \cdot \boldsymbol{b} + \boldsymbol{a} \cdot \boldsymbol{c}$;

(9) $(c_1\boldsymbol{a}) \cdot (c_2\boldsymbol{b}) = c_1c_2(\boldsymbol{a} \cdot \boldsymbol{b})$, for any$c_1, c_2 \in \mathbf{R}$;

(10) $(\boldsymbol{AB})_{ij} = \boldsymbol{a}^i \cdot \boldsymbol{b}_j$;

(11) $\boldsymbol{AB} \neq \boldsymbol{BA}$;

(12) $\boldsymbol{A}(\boldsymbol{B}+\boldsymbol{C}) = \boldsymbol{AB}+\boldsymbol{AC}$;

(13) $(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{C} = \boldsymbol{AC}+\boldsymbol{BC}$;

(14) $\boldsymbol{A}(\boldsymbol{BC}) = (\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{ABC}$;

(15) $c(\boldsymbol{AB}) = (c\boldsymbol{A})\boldsymbol{B} = \boldsymbol{A}(c\boldsymbol{B})$;

(16) $(\boldsymbol{AB})' = \boldsymbol{B}'\boldsymbol{A}'$;

(17) the rank of a matrix, $\operatorname{rk} \boldsymbol{A}$, is the maximum number of linearly independent rows or columns of $\boldsymbol{A}$;

(18) $0 \leq \operatorname{rk} \boldsymbol{A} \leq \min\{n, p\}$;

(19) $\operatorname{rk} \boldsymbol{A} = \operatorname{rk} \boldsymbol{A}'$;

(20) $\operatorname{rk} \boldsymbol{A} = 0$ if and only if $\boldsymbol{A} = \boldsymbol{O}_{n\times p}$;

(21) $\operatorname{rk} \boldsymbol{I}_n = n$;

(22) $\operatorname{rk} c\boldsymbol{A} = \operatorname{rk} \boldsymbol{A}$, for any $c \in \mathbf{R}_0$;

(23) $\operatorname{rk} \boldsymbol{A}+\boldsymbol{B} \leq \operatorname{rk} \boldsymbol{A} + \operatorname{rk} \boldsymbol{B}$;

(24) $\operatorname{rk} \boldsymbol{A}-\boldsymbol{B} \geq |\operatorname{rk} \boldsymbol{A} - \operatorname{rk} \boldsymbol{B}|$;

(25) $\operatorname{rk} \boldsymbol{AB} \leq \min\{\operatorname{rk} \boldsymbol{A}, \operatorname{rk} \boldsymbol{B}\}$;

(26) $\operatorname{rk} \boldsymbol{A} = \sum_{i=1}^{n} \mathbf{1}\{\lambda_i \neq 0\}$;

(27) $\det \boldsymbol{I}_n = 1$;

(28) $\det \boldsymbol{A}' = \det \boldsymbol{A}$;

(29) $\det \boldsymbol{A}^{-1} = (\det \boldsymbol{A})^{-1}$;

(30) $\det \boldsymbol{AB} = \det \boldsymbol{A} \det \boldsymbol{B}$;

(31) $\det(c\boldsymbol{A}) = c^n \det \boldsymbol{A}$, for any $c \in \mathbf{R}$;

(32) for a lower or upper triangular matrix, including a diagonal matrix, $\boldsymbol{A}$

$$\det \boldsymbol{A} = \prod_{i=1}^{n} a_{ii};$$

(33) the determinant changes signs when two rows of $\boldsymbol{A}$ are interchanged;

(34) subtracting a multiple of one row of $\boldsymbol{A}$ from another leaves $\det \boldsymbol{A}$ unchanged;

(35) if $\boldsymbol{A}$ has a row of zeros then $\det \boldsymbol{A} = 0$;

(36) the determinant is a function of each row separately, i.e.

$$\begin{vmatrix} ca_{11} & ca_{12} & \dots & ca_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} = c \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}$$

and

$$\begin{vmatrix} a_{11} + a'_{11} & a_{12} + a'_{12} & \dots & a_{1n} + a'_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} + \begin{vmatrix} a'_{11} & a'_{12} & \dots & a'_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix};$$

(37) $\operatorname{tr} \boldsymbol{A} = \sum_{i=1}^{n} a_{ii}$;

(38) $\operatorname{tr} \boldsymbol{A} + \boldsymbol{B} = \operatorname{tr} \boldsymbol{A} + \operatorname{tr} \boldsymbol{B}$;

(39) $\operatorname{tr}(c\boldsymbol{A}) = c \operatorname{tr} \boldsymbol{A}$, for any $c \in \mathbf{R}$;

(40) if $\boldsymbol{A} \in \mathbf{R}^{n \times p}$ and $\boldsymbol{B} \in \mathbf{R}^{p \times n}$, then

$$\operatorname{tr} \boldsymbol{AB} = \operatorname{tr} \boldsymbol{BA};$$

(41) $\operatorname{tr}(\boldsymbol{A}') = \operatorname{tr} \boldsymbol{A}$;

(42) $\operatorname{tr} c = c$, for $c \in \mathbf{R}$;

(43) a square matrix is positive definite if $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0$ for all $\boldsymbol{x} \in \mathbf{R}^n \backslash \{\mathbf{0}\}$;

(44) $\det \boldsymbol{A} = \prod_{i=1}^{n} \lambda_i$;

(45) $\operatorname{tr} \boldsymbol{A} = \sum_{i=1}^{n} \lambda_i$;

(46) for a symmetric positive definite matrix:

   (i) $\det \boldsymbol{A} > 0$ such that it is also non-singular;

   (ii) $\operatorname{tr} \boldsymbol{A} > 0$;

   (iii) if $\boldsymbol{A} \geq \boldsymbol{O}$ and $\boldsymbol{B} \geq \boldsymbol{O}$, then $\boldsymbol{A} + \boldsymbol{B}$ is at least positive semi-definite;

   (iv) $\boldsymbol{A} > \boldsymbol{O}$ if and only if $\boldsymbol{A}^{-1} > \boldsymbol{O}$.

A.1 Vector and matrix differentiation

(1) $\mathrm{d}\boldsymbol{f}' = (\mathrm{d}\boldsymbol{f})'$;

(2) $\mathrm{d}(c\boldsymbol{f}) = c\,\mathrm{d}\boldsymbol{f}$, for any $c \in \mathbf{R}$;

(3) $\mathrm{d}(\boldsymbol{f} \pm \boldsymbol{g}) = \mathrm{d}\boldsymbol{f} \pm \mathrm{d}\boldsymbol{g}$;

(4) $\mathrm{d}(\operatorname{tr} \boldsymbol{f}) = \operatorname{tr} \mathrm{d}\boldsymbol{f}$;

(5) $\mathrm{d}\boldsymbol{f}\boldsymbol{g} = (\mathrm{d}\boldsymbol{f})\boldsymbol{g} + \boldsymbol{f}(\mathrm{d}\boldsymbol{g})$;

(6) $D\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}$;

(7) $D\boldsymbol{x}'\boldsymbol{A}\boldsymbol{a} = \boldsymbol{a}'\boldsymbol{A}$;

(8) $D\boldsymbol{a}'\boldsymbol{A}\boldsymbol{x} = \boldsymbol{a}'\boldsymbol{A}$;

(9) $D\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}'(\boldsymbol{A} + \boldsymbol{A}')$.

B Additional Material (All Non-Examinable)

This section of the appendix collects together additional material which has been taught on the pre-course in previous years. It is meant solely as a reference and does not form part of the course this year.

B.1 Preliminary concepts

B.1.1 Hadamard and Kronecker product (*)

The way to define matrix multiplication as we have done before is not the only way. Below, we present two more ways to do so.

Definition 27 (Hadamard product)**.** For two $n \times p$ matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ the Hadamard product $\boldsymbol{A} \odot \boldsymbol{B}$ is obtained by element-wise multiplication,

$$(\boldsymbol{A} \odot \boldsymbol{B})_{ij} := (\boldsymbol{A})_{ij}(\boldsymbol{B})_{ij}.$$

The Hadamard product yields a matrix of the same dimensions as the individual factors. By the definition above, it is not defined for two matrices of different dimensions. This particular matrix product satisfies the properties below.

Properties 11 (Hadamard product)**.** For three conformable matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$:

(i) $\boldsymbol{A} \odot \boldsymbol{B} = \boldsymbol{B} \odot \boldsymbol{A}$;

(ii) $\boldsymbol{A} \odot (\boldsymbol{B} \odot \boldsymbol{C}) = (\boldsymbol{A} \odot \boldsymbol{B}) \odot \boldsymbol{C}$;

(iii) $\boldsymbol{A} \odot (\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{A} \odot \boldsymbol{B} + \boldsymbol{A} \odot \boldsymbol{C}$;

(iv) the identity matrix under the Hadamard product $\boldsymbol{I}_H$ is a matrix of the same dimension of all ones

$$\boldsymbol{A}\boldsymbol{I}_H = \boldsymbol{A} = \boldsymbol{I}_H\boldsymbol{A},$$

and every matrix without any zero-valued elements has an inverse with elements $(\boldsymbol{A}_H^{-1})_{ij} = 1/(\boldsymbol{A})_{ij}$ such that

$$\boldsymbol{A} \odot \boldsymbol{A}_H^{-1} = \boldsymbol{I}_H = \boldsymbol{A}_H^{-1} \odot \boldsymbol{A}.$$

(v) $\operatorname{rk} \boldsymbol{A} \odot \boldsymbol{B} \leq \operatorname{rk} \boldsymbol{A} \operatorname{rk} \boldsymbol{B}$.

Example B.1. The Hadamard product between the matrices

$$\boldsymbol{A} = \begin{bmatrix} 18 & -5 \\ 13 & -10 \end{bmatrix} \quad \text{and} \quad \boldsymbol{B} = \begin{bmatrix} -1 & -9 \\ 2 & -17 \end{bmatrix},$$

is

$$\boldsymbol{A} \odot \boldsymbol{B} = \begin{bmatrix} -18 & 45 \\ 26 & 170 \end{bmatrix}.$$

Finally, we introduce the Kronecker product. This product is defined for any two matrices of arbitrary dimensions.

Definition 28 (Kronecker product)**.** The Kronecker product between an $n \times p$ matrix $\boldsymbol{A}$ and a $k \times m$ matrix $\boldsymbol{B}$ is the $nk \times pm$ matrix

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11}\boldsymbol{B} & \dots & a_{1p}\boldsymbol{B} \\ \vdots & \ddots & \vdots \\ a_{n1}\boldsymbol{B} & \dots & a_{np}\boldsymbol{B} \end{bmatrix}.$$

The Kronecker product has many interesting applications and properties, some of which are listed below.

Properties 12 (Kronecker product)**.** For four matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$:

(i) $c(\boldsymbol{A} \otimes \boldsymbol{B}) = (c\boldsymbol{A}) \otimes \boldsymbol{B} = \boldsymbol{A} \otimes (c\boldsymbol{B})$, for any $c \in \mathbf{R}$;

(ii) $(\boldsymbol{A} \otimes \boldsymbol{B})' = \boldsymbol{A}' \otimes \boldsymbol{B}'$;

(iii) $(\boldsymbol{A} \otimes \boldsymbol{B}) \otimes \boldsymbol{C} = \boldsymbol{A} \otimes (\boldsymbol{B} \otimes \boldsymbol{C})$;

(iv) $(\boldsymbol{A} + \boldsymbol{B}) \otimes \boldsymbol{C} = \boldsymbol{A} \otimes \boldsymbol{C} + \boldsymbol{B} \otimes \boldsymbol{C}$;

(v) $\boldsymbol{A} \otimes (\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{A} \otimes \boldsymbol{B} + \boldsymbol{A} \otimes \boldsymbol{C}$;

(vi) $(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = \boldsymbol{AC} \otimes \boldsymbol{BD}$;

(vii) $\operatorname{rk} \boldsymbol{A} \otimes \boldsymbol{B} = \operatorname{rk} \boldsymbol{A} \operatorname{rk} \boldsymbol{B}$;

(viii) for $\boldsymbol{A} \in \mathbf{R}^{n \times n}$ and $\boldsymbol{B} \in \mathbf{R}^{m \times m}$:

  (a) $\operatorname{tr} \boldsymbol{A} \otimes \boldsymbol{B} = \operatorname{tr} \boldsymbol{B} \otimes \boldsymbol{A} = \operatorname{tr} \boldsymbol{A} \operatorname{tr} \boldsymbol{B}$;

  (b) $\det \boldsymbol{A} \otimes \boldsymbol{B} = \det \boldsymbol{B} \otimes \boldsymbol{A} = (\det \boldsymbol{A})^n (\det \boldsymbol{B})^m$.

The Kronecker product will be useful for us in Appendix B.3.1.

B.2 Square matrices

B.2.1 Determinant by Laplacian expansion (*)

Another way to compute the determinant is given by the Laplacian expansion. In order to define the determinant in this way, we consider a set $(1, \dots, n)$ and a permutation $\sigma = (j_1, \dots, j_n)$ thereof. Recall that a permutation is any rearrangement of the original set. For a set of size n, there are $n!$ permutations.

We call a transposition the interchange of two elements of $(1, \dots, n)$ and denote the total number of transpositions needed to obtain σ from $(1, \dots, n)$ by $N(\sigma)$.

Example B.2. Consider the set $(1, 2, 3)$ which has $3! = 6$ permutations (check this). Two of these permutations are $\sigma_1 = (1, 3, 2)$ and $\sigma_2 = (3, 1, 2)$. To get σ_1, we interchanged 2 and 3 such that we needed only one transposition. Hence, $N(\sigma_1) = 1$. For σ_2, we needed two transpositions

$$(1, 2, 3) \mapsto (3, 2, 1) \mapsto (3, 1, 2),$$

and thus $N(\sigma_2) = 2$.

Now, let the map $\mathrm{sgn} : \mathbf{N} \to \{-1, 1\}$ be

$$\mathrm{sgn}(\sigma) = \begin{cases} -1 & \text{if } N(\sigma) \text{ is odd} \\ 1 & \text{if } N(\sigma) \text{ is even} \end{cases}.$$

Example B.3 (Example B.2 continued)**.** Given the above definition, $\mathrm{sgn}(\sigma_1) = -1$ and $\mathrm{sgn}(\sigma_2) = 1$.

Definition 29 (Determinant by Laplacian expansion)**.** The determinant of a square $n \times n$ matrix $\boldsymbol{A}$ defined by the *Laplacian expansion* is

$$\det \boldsymbol{A} := \sum \mathrm{sgn}(\sigma) \prod_{i=1}^{n} a_{i j_i}, \tag{9}$$

where the sum is taken over all $n!$ permutations, σ, of $(1, \dots, n)$.

Example B.4. Consider the 2×2 matrix

$$\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

There are two permutations of $(1, 2)$, i.e. $(2, 1)$ and $(1, 2)$ itself. We have that $\mathrm{sgn}(1, 2) = 1$ and $\mathrm{sgn}(2, 1) = -1$. Plugging this into (9)

$$\begin{aligned} \det \boldsymbol{A} &= 1 \times a_{11} a_{22} + (-1) \times a_{12} a_{21} \\ &= ad - bc. \end{aligned}$$

Example B.5. Consider the 3×3 matrix

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

The set $(1, 2, 3)$ has 6 permutations:

$$\begin{aligned}&(1,2,3), \quad (1,3,2),\\ &(2,1,3), \quad (2,3,1),\\ &(3,1,2), \quad (3,2,1).\end{aligned}$$

The determinant of a 3×3 is therefore

$$\begin{aligned}\det \boldsymbol{A} = &\operatorname{sgn}(1,2,3)a_{11}a_{22}a_{33} + \operatorname{sgn}(1,3,2)a_{11}a_{23}a_{32} + \operatorname{sgn}(2,1,3)a_{12}a_{21}a_{33} + \operatorname{sgn}(2,3,1)a_{12}a_{23}a_{31}\\ &+ \operatorname{sgn}(3,1,2)a_{13}a_{21}a_{32} + \operatorname{sgn}(3,2,1)a_{13}a_{22}a_{31}\\ =&a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}.\end{aligned}$$

The Laplacian expansion gives you exactly the same as expanding in cofactors such that these two definitions for the determinant are equivalent.

B.3 Vector and matrix calculus

We have all the tools to introduce vector and matrix calculus. To gain some intuition, recall the definition of the derivative in the one-dimensional case

$$\lim_{u \to 0} \frac{\varphi(x+u) - \varphi(x)}{u} = \varphi'(x).$$

Re-writing gives,

$$\varphi(x+u) = \varphi(x) + \varphi'(x)u + r_x(u) \tag{10}$$

where $r_x(u)$ is the remainder term such that $r_x(u)/u \to 0$ as $u \to 0$.

Define the first differential of φ at x (with increment u) as $\mathrm{d}\varphi(x; u) = \varphi'(x)u$. Note, that we do not require u to be infinitesimally small for the differential to be well defined. As the example below shows, $\mathrm{d}\varphi(x; u)$ is the linear part (in u) of the increment $\varphi(x+u) - \varphi(x)$. Setting aside rigorous justification for the double use of the symbol "d", we will write $\mathrm{d}x$ for u, hence $\mathrm{d}\varphi(x) = \varphi'(x)\,\mathrm{d}x$.

Example B.6. For $\varphi(x) = x^2$, we have $\varphi(x+u) - \varphi(x) = 2xu + u^2$. Then, it follows that $\mathrm{d}\varphi(x; u) = 2xu$ with $r_x(u) = u^2$ such that $r_x(u)/u = u \to 0$ as $u \to 0$.

The expansion in (10) caries over into the vector case

$$\boldsymbol{f}(\boldsymbol{x}+\boldsymbol{u}) = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{A}(\boldsymbol{x})\boldsymbol{u} + \boldsymbol{r}_{\boldsymbol{x}}(\boldsymbol{u}). \tag{11}$$

If $\boldsymbol{A}(\boldsymbol{x})$ depends only on $\boldsymbol{x}$ but not on $\boldsymbol{u}$ and $\boldsymbol{r}_{\boldsymbol{x}}(\boldsymbol{u})/\|\boldsymbol{u}\| \to \boldsymbol{0}$ as $\boldsymbol{u} \to \boldsymbol{0}$, then $\boldsymbol{f}$ is differentiable at $\boldsymbol{x}$. From this expansion you may see that $\boldsymbol{A}(\boldsymbol{x})$ is an operator acting on a vector $\boldsymbol{u} \in \mathbf{R}^n$. Combine this with the fact that $\boldsymbol{f} : \mathbf{R}^n \to \mathbf{R}^m$ yields an m-vector-valued output, then there can be no ambiguity that $\boldsymbol{A}(\boldsymbol{u})$

needs to be an $m \times n$ matrix.

Remark. In this exposition, we have concentrated on vector-valued functions, $\boldsymbol{f}$, with vectors as input. The results carry over for scalar functions or scalar inputs. In those cases, set $m = 1$ or $n = 1$ or both. Hence, differentiating a scalar-valued function with respect to a vector produces a *row* vector and differentiating a vector-valued function with respect to a vector produces a matrix. The obvious consequence of this is that the second derivative of a scalar-valued function with respect to a vector is a matrix, e.g. recall the Hessian matrix in definition 26.

The problem with the above expansion is that it does not always provide us with an easy recipe to find $\boldsymbol{A}(\boldsymbol{x})$. Let the first differential be

$$\mathrm{d}\boldsymbol{f}(\boldsymbol{x};\boldsymbol{u}) = \boldsymbol{A}(\boldsymbol{x})\boldsymbol{u},$$

and take the following identification result as given

$$\mathrm{d}\boldsymbol{f}(\boldsymbol{x}) = A(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \Leftrightarrow D\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}(\boldsymbol{x}). \tag{12}$$

This result states that if the function is differentiable at $\boldsymbol{x}$, we can identify the first derivative as the Jacobian matrix from definition 25. Hence, depending on the situation we can either find $\boldsymbol{A}(\boldsymbol{u})$ directly or compute $D\boldsymbol{f}(\boldsymbol{x})$.

Unlike the one-dimensional case, it is often easier to work with differentials rather than derivatives in the multi-dimensional case; this is especially true when we consider matrices further down below. Therefore, the following properties are stated in terms of differentials and not derivatives.

Properties 13 (Vector-valued differentials)**.** Let $\boldsymbol{f}$ and $\boldsymbol{g}$ be two conformable vector functions:

(i) $\mathrm{d}\boldsymbol{f}' = (\mathrm{d}\boldsymbol{f})'$;

(ii) $\mathrm{d}(c\boldsymbol{f}) = c\,\mathrm{d}\boldsymbol{f}$, for any $c \in \mathbf{R}$;

(iii) $\mathrm{d}(\boldsymbol{f} \pm \boldsymbol{g}) = \mathrm{d}\boldsymbol{f} \pm \mathrm{d}\boldsymbol{g}$;

(iv) $\mathrm{d}\boldsymbol{f}\boldsymbol{g} = (\mathrm{d}\boldsymbol{f})\boldsymbol{g} + \boldsymbol{f}(\mathrm{d}\boldsymbol{g})$.

Example B.7. Consider the scalar-valued function $\varphi(\boldsymbol{x}) = \boldsymbol{a}'\boldsymbol{x}$, where $\boldsymbol{a}$ is a vector of coefficients. Then,

$$\boldsymbol{a}'(\boldsymbol{x} + \boldsymbol{u}) - \boldsymbol{a}'\boldsymbol{x} = \boldsymbol{a}'\boldsymbol{u}.$$

In line with (11), set $\boldsymbol{A}(\boldsymbol{x}) = \boldsymbol{a}'$ and $\boldsymbol{r}_{\boldsymbol{x}}(\boldsymbol{u}) = 0$. Hence, $\mathrm{d}\boldsymbol{a}'\boldsymbol{x} = \boldsymbol{a}'\,\mathrm{d}\boldsymbol{x}$. It follows that $D\varphi(\boldsymbol{x}) = \boldsymbol{a}'$, a row vector as expected.

Example B.8. Consider the scalar-valued function $\varphi(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$, where $\boldsymbol{A}$ is a matrix of coefficients.

First, notice that with similar reasoning as in example B.7, we can show that $\mathrm{d}\boldsymbol{Ax} = \boldsymbol{A}\,\mathrm{d}\boldsymbol{x}$. Furthermore, $(\mathrm{d}\boldsymbol{x}')\boldsymbol{Ax}$ is a scalar so it is equal to its transpose $\boldsymbol{x}'\boldsymbol{A}'(\mathrm{d}\boldsymbol{x}')'$. Using this and properties (i) and (v) above,

$$\begin{aligned}\mathrm{d}\varphi(\boldsymbol{x}) &= (\mathrm{d}\boldsymbol{x}')\boldsymbol{Ax} + \boldsymbol{x}'\,\mathrm{d}\boldsymbol{Ax} \\ &= \boldsymbol{x}'\boldsymbol{A}'(\mathrm{d}\boldsymbol{x}')' + \boldsymbol{x}'\boldsymbol{A}\,\mathrm{d}\boldsymbol{x} \\ &= \boldsymbol{x}'\boldsymbol{A}'\,\mathrm{d}\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{A}\,\mathrm{d}\boldsymbol{x} \\ &= \boldsymbol{x}'(\boldsymbol{A}+\boldsymbol{A}')\,\mathrm{d}\boldsymbol{x}.\end{aligned}$$

Hence, $D\varphi(\boldsymbol{x}) = \boldsymbol{x}'(\boldsymbol{A}+\boldsymbol{A}')$. Of course, if $\boldsymbol{A}$ is symmetric this reduces to: $D\varphi(\boldsymbol{x}) = 2\boldsymbol{x}'\boldsymbol{A}$.

We illustrate using the derivative directly according to (12) which sets the following examples apart from the previous ones.

Example B.9. Consider two vector-valued functions $\boldsymbol{f}(\boldsymbol{x}) = (1, x_2, x_1^2)'$ and $\boldsymbol{g}(\boldsymbol{x}) = (x_1x_2\exp\{x_3^2\}, \log x_4)'$, then the derivatives are

$$D\boldsymbol{f} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 2x_1 & 0 \end{bmatrix},$$

and

$$D\boldsymbol{g} = \begin{bmatrix} x_2\exp\{x_3^2\} & x_1\exp\{x_3^2\} & 2x_1x_2x_3\exp\{x_3^2\} & 0 \\ 0 & 0 & 0 & 1/x_4 \end{bmatrix}.$$

Exercise B.1. Consider the scalar function $\varphi(x,y) = \sin(x)\sin(y)$. Find all its stationary points and determine whether these are local minima, maxima or saddle points by computing the Hessian matrix. *Hint: what does the definiteness of the Hessian matrix tell you about the concavity or convexity of φ?*

Example B.10. Consider some given $n \times 1$ vector $\boldsymbol{y}$, $n \times m$ matrix $\boldsymbol{X}$, and $m \times 1$ vector $\boldsymbol{\beta}$. Find the expression for $\boldsymbol{\beta}$ (in terms of $\boldsymbol{y}$ and $\boldsymbol{X}$) that satisfies the following condition

$$\frac{\partial}{\partial\boldsymbol{\beta}}\Big[(\boldsymbol{y}-\boldsymbol{X\beta})'(\boldsymbol{y}-\boldsymbol{X\beta})\Big] = \boldsymbol{0}.$$

Firstly let us expand the expression and consider the dimensions of our problem.

$$(\boldsymbol{y}-\boldsymbol{X\beta})'(\boldsymbol{y}-\boldsymbol{X\beta}) = (-\boldsymbol{\beta}'\boldsymbol{X}'+\boldsymbol{y}')(\boldsymbol{y}-\boldsymbol{X\beta}) = -\underbrace{\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y}}_{1\times1} + \underbrace{\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X\beta}}_{1\times1} + \underbrace{\boldsymbol{y}'\boldsymbol{y}}_{1\times1} - \underbrace{\boldsymbol{y}'\boldsymbol{X\beta}}_{1\times1},$$

such that it becomes clear immediately that we are being asked to take the derivative of a scalar with respect to a $m \times 1$ vector. We already know the resultant matrix will therefore be of size $m \times 1$ (notice here we differentiate with respect to $\boldsymbol{\beta}$, rather than $\boldsymbol{\beta}'$ as would be the standard case according to our definition of the Jacobian matrix).

Next, differentiate each term individually to give

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\beta}}\Big[(\boldsymbol{y}-X\boldsymbol{\beta})'(\boldsymbol{y}-X\boldsymbol{\beta})\Big] &= \underset{m\times 1}{-X'\boldsymbol{y}} + \underset{m\times 1}{2X'X\boldsymbol{\beta}} - \underset{m\times 1}{(\boldsymbol{y}'X)'},\\ &= -2X'\boldsymbol{y} + 2X'X\boldsymbol{\beta} = -2X'(\boldsymbol{y}-X\boldsymbol{\beta}) = 0,\\ &\rightarrow \boldsymbol{\beta} = (X'X)^{-1}X'\boldsymbol{y}.\end{aligned}$$

Exercise B.2. In statistics, the problem of ridge regression is formulated as follows

$$\hat{\boldsymbol{\beta}}_R(\lambda) := \underset{\beta\in\mathbf{R}^k}{\arg\min}\ (\boldsymbol{y}-X\boldsymbol{\beta})'(\boldsymbol{y}-X\boldsymbol{\beta}) + \lambda\|\boldsymbol{\beta}\|^2,$$

where $\lambda \in \mathbf{R}$ is a parameter that needs to be chosen beforehand. Notice, that this is a least-squares problem but where $\boldsymbol{\beta}$'s which are "too large" are penalised. By computing the first-order conditions, find $\hat{\boldsymbol{\beta}}_R(\lambda)$.

Exercise B.3 (PCA continued)**.** Show that the solution to (5) in Section 4.4.1 is indeed an eigenvector of $\tilde{X}'\tilde{X}$. Argue that the optimal solution is the eigenvector associated with the largest eigenvalue.

B.3.1 Derivative with respect to a matrix (*)

So far we have only discussed how to differentiate with respect to a scalar or a vector, but not with respect to a matrix. However, before we do so we must introduce the concept of vectorisation.

Definition 30. The vectorisation of a matrix $\boldsymbol{A}$, denoted by vec $\boldsymbol{A}$, reorders the $n \times p$ matrix by stacking the columns on top of each other into an np-vector. Formally,

$$\operatorname{vec} \boldsymbol{A} := (a_{11}, \ldots, a_{n1}, a_{12}, \ldots, a_{n2}, \ldots, a_{1p}, \ldots, a_{np})'.$$

Example B.11. The vectorisation of the matrix $\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is

$$\operatorname{vec} \boldsymbol{A} = \begin{bmatrix} a \\ c \\ b \\ d \end{bmatrix}.$$

Vectorise the matrix-valued function and the matrix with respect to which we are differentiating, then the Jacobian matrix of $\boldsymbol{F}$ at $\boldsymbol{X}$ is identified by

$$D\boldsymbol{F}(\boldsymbol{X}) = \frac{\mathrm{d}\operatorname{vec} \boldsymbol{F}(\boldsymbol{X})}{\mathrm{d}(\operatorname{vec} \boldsymbol{X})'}, \tag{13}$$

which for $\boldsymbol{F}$ a $n \times p$ matrix-valued function and $\boldsymbol{X}$ a $m \times q$ matrix yields a matrix of order $np \times mq$.

This is not to only way to represent the matrix derivative. However, Magnus and Neudecker (1999) argue that this is the only *natural* generalisation of the Jacobian matrix from vectors to matrices. It is possible to go via a definition using partial derivatives which orders the derivatives as such

$$\frac{\partial \boldsymbol{F}(\boldsymbol{X})}{\partial \boldsymbol{X}} = \begin{bmatrix} \frac{\partial f_{11}(\boldsymbol{X})}{\partial \boldsymbol{X}} & \cdots & \frac{\partial f_{1p}(\boldsymbol{X})}{\partial \boldsymbol{X}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_{n1}(\boldsymbol{X})}{\partial \boldsymbol{X}} & \cdots & \frac{\partial f_{np}(\boldsymbol{X})}{\partial \boldsymbol{X}} \end{bmatrix},$$

with

$$\frac{\partial f_{ij}(\boldsymbol{X})}{\partial \boldsymbol{X}} = \begin{bmatrix} \frac{\partial f_{ij}(\boldsymbol{X})}{\partial x_{11}} & \cdots & \frac{\partial f_{ij}(\boldsymbol{X})}{\partial x_{1q}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_{ij}(\boldsymbol{X})}{\partial x_{m1}} & \cdots & \frac{\partial f_{ij}(\boldsymbol{X})}{\partial x_{mq}} \end{bmatrix}.$$

The matrix $\frac{\partial \boldsymbol{F}(\boldsymbol{X})}{\partial \boldsymbol{X}}$ is of order $mn \times pq$ and contains the same $mnpq$ partial derivatives as DF but in a different order. Magnus and Neudecker (1999, p. 195) state that this matrix is not the Jacobian matrix, and thus its determinant is meaningless which is a problem for applications, nor does this definition provide us with a practical chain rule. In the remainder of the discussion, we will therefore stick with (13).

Below, we present some properties of vectorisation.

Properties 14 (Vectorisation)**.** For three conformable matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$:

(i) $\operatorname{vec}(\boldsymbol{A} + \boldsymbol{B}) = \operatorname{vec} \boldsymbol{A} + \operatorname{vec} \boldsymbol{B}$;

(ii) $\operatorname{vec}(\boldsymbol{A} \odot \boldsymbol{B}) = \operatorname{vec} \boldsymbol{A} \odot \operatorname{vec} \boldsymbol{B}$;

(iii) $\operatorname{vec} \boldsymbol{ABC} = (\boldsymbol{C}' \otimes \boldsymbol{A}) \operatorname{vec} \boldsymbol{B}$;

(iv) $\operatorname{vec} \boldsymbol{AB} = (\boldsymbol{B}' \otimes \boldsymbol{I}_n) \operatorname{vec} \boldsymbol{A} = (\boldsymbol{I}_m \otimes A) \operatorname{vec} \boldsymbol{B}$;

(v) $\operatorname{vec}(\mathrm{d}\boldsymbol{F}(\boldsymbol{X})) = \mathrm{d} \operatorname{vec} \boldsymbol{F}(\boldsymbol{X})$.

By vectorising, we can analyse matrix derivatives using exactly the same apparatus already developed above. Let us look at some examples.

Example B.12. Consider the following matrix function $\boldsymbol{F}(\boldsymbol{X}) = \boldsymbol{x}\boldsymbol{x}'$, i.e. $\boldsymbol{F}$ maps a vector into its outer product. It should be obvious that $\mathrm{d}\boldsymbol{x}\boldsymbol{x}' = (\mathrm{d}\boldsymbol{x})\boldsymbol{x}' + \boldsymbol{x}(\mathrm{d}\boldsymbol{x})'$. Combining this with property (v) we get

$$\begin{aligned} \mathrm{d}(\operatorname{vec} \boldsymbol{x}\boldsymbol{x}') &= \operatorname{vec}(\mathrm{d}\boldsymbol{x}\boldsymbol{x}') \\ &= \operatorname{vec}[(\mathrm{d}\boldsymbol{x})\boldsymbol{x}' + \boldsymbol{x}(\mathrm{d}\boldsymbol{x})'], \end{aligned}$$

splitting up the sum by property (i), and rewriting the multiplications within the vectorisation using property (iv) yields

$$
\begin{aligned}
&= \operatorname{vec}[(\mathrm{d}\boldsymbol{x})\boldsymbol{x}'] + \operatorname{vec}[\boldsymbol{x}(\mathrm{d}\boldsymbol{x})'] \\
&= (\boldsymbol{x} \otimes \boldsymbol{I}) \operatorname{vec}(\mathrm{d}\boldsymbol{x}) + (\boldsymbol{I} \otimes \boldsymbol{x}) \operatorname{vec}((\mathrm{d}\boldsymbol{x})') \\
&= [(\boldsymbol{x} \otimes \boldsymbol{I}) + (\boldsymbol{I} \otimes \boldsymbol{x})]\, \mathrm{d}\boldsymbol{x},
\end{aligned}
$$

where the final equality follows upon realising that by definition $\operatorname{vec}((\mathrm{d}\boldsymbol{x})') = \operatorname{vec}(\mathrm{d}\boldsymbol{x}) = \mathrm{d}\boldsymbol{x}$. Hence, by (13)

$$
DF(X) = \boldsymbol{x} \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \boldsymbol{x}.
$$

Example B.13. Consider the scalar function $\varphi(X) = \boldsymbol{a}'X\boldsymbol{b}$, where $\boldsymbol{a}$ and $\boldsymbol{b}$ are conformable coefficient vectors. First, see that $\mathrm{d}(\boldsymbol{a}'X\boldsymbol{b}) = \boldsymbol{a}'(\mathrm{d}X)\boldsymbol{b}$. Then, using the properties of vectorisation

$$
\begin{aligned}
\mathrm{d}(\operatorname{vec} \boldsymbol{a}'X\boldsymbol{b}) &= \operatorname{vec}(\mathrm{d}\boldsymbol{a}'X\boldsymbol{b}) \\
&= \operatorname{vec} \boldsymbol{a}'(\mathrm{d}X)\boldsymbol{b} \\
&= (\boldsymbol{b}' \otimes \boldsymbol{a}') \operatorname{vec} \mathrm{d}X \\
&= (\boldsymbol{b}' \otimes \boldsymbol{a}') \operatorname{dvec} X.
\end{aligned}
$$

Therefore,

$$
D\varphi(X) = \boldsymbol{b}' \otimes \boldsymbol{a}',
$$

which is a row vector, or in partial derivative notation

$$
\frac{\partial \varphi(X)}{\partial X} = \boldsymbol{a}\boldsymbol{b}'.
$$

Example B.14. Let $F(\boldsymbol{A}) = \boldsymbol{A}^{-1}$ be a matrix-valued function. We have $\boldsymbol{O}_n = \mathrm{d}\boldsymbol{I}_n = \mathrm{d}\boldsymbol{A}^{-1}\boldsymbol{A}$, such that

$$
\begin{aligned}
\boldsymbol{O}_n = \mathrm{d}\boldsymbol{I}_n &= \mathrm{d}\boldsymbol{A}^{-1}\boldsymbol{A} \\
&= (\mathrm{d}\boldsymbol{A}^{-1})\boldsymbol{A} + \boldsymbol{A}^{-1}\, \mathrm{d}\boldsymbol{A},
\end{aligned}
$$

and thus,

$$
\mathrm{d}\boldsymbol{A}^{-1} = -\boldsymbol{A}^{-1}(\mathrm{d}\boldsymbol{A})\boldsymbol{A}^{-1}.
$$

By the properties of vectorisation, it follows that

$$
DF(\boldsymbol{A}) = -(\boldsymbol{A}')^{-1} \otimes \boldsymbol{A}^{-1}.
$$

This result is interesting, because we can consider $\boldsymbol{A}(\theta)$ to be matrix function of a scalar parameter $\theta \in \mathbf{R}$ such that $\boldsymbol{F}(\theta) = \boldsymbol{A}^{-1}(\theta)$. This is useful in statistics as (inverses of) covariance matrices often depend on scalar parameters. The result above combined with a chain rule Magnus and Neudecker (1999, Theorem 12 p. 108) gives

$$DA^{-1}(\theta) = -[(A')^{-1}(\theta) \otimes A^{-1}(\theta)]DA(\theta).$$

Exercise B.4. Show that for $\varphi(X) = a'XX'a$, we have $DF(X) = 2(a'X) \otimes a'$.

Exercise B.5. Show for a square matrix X, $F(X) = X^n$ and $n \in \mathrm{N}_0$, that

$$DF(X) = \sum_{p=1}^{n} (X')^{n-p} \otimes X^{p-1}.$$