Linear Algebra
Preparatory Course 2013

Alexandre N. Kohlhas*

*Many thanks goes to Heather Battey and Donald Robertson.
Please send comments and corrections to ak604@cam.ac.uk.

Contents

1 Basics

1.1 Basic definitions

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

$$A := (a_{ij}) = \begin{pmatrix} a_{11} & \cdot & \cdot & \cdot & a_{1p} \\ a_{21} & \cdot & \cdot & \cdot & a_{2p} \\ \cdot & \cdot & & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot \\ a_{n1} & \cdot & \cdot & \cdot & a_{np} \end{pmatrix}.$$

A matrix of order $1 \times p$ is called a row vector and a matrix of order $n \times 1$ is called a column vector.

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

Definition 3. Given an $n \times p$ matrix $A = (a_{ij})$, its transpose $A' := (a_{ji})$ (often also denoted A^T) is the $p \times n$ matrix obtained by interchanging the rows and columns of A. I.e., the rows of A' are the columns of A (and the rows of A are the columns of A').[1]

Definition 4. An $n \times p$ matrix is called a square matrix if $n = p$.

Definition 5. Let A be a square matrix, then A is symmetric if $A' = A$.

Definition 6. Let A be a square matrix, then A is diagonal if $a_{ij} = 0$ for all $i \neq j$ [i.e., if all off-diagonal elements are zero].

Example 1 (**Identity matrix**). *The identity matrix of order n is the diagonal matrix of order n with ones along the main diagonal (and zeros elsewhere), i.e.:*

$$I_n := \begin{pmatrix} 1 & 0 & \cdot & \cdot & 0 \\ 0 & 1 & & & \\ \cdot & & \cdot & & \\ \cdot & & & \cdot & \\ 0 & \cdot & \cdot & \cdot & 1 \end{pmatrix}.$$

[1] Note that by definition: $(A')' = A$.

The notational convention in what follows (and in many texts that use matrix algebra) will be to use capital letters such as A and B to denote matrices; bold (often subscripted) lower-case letters such as $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{b}_j$ to denote vectors; and lower-case subscripted letters such as a_k, b_k and b_{jk} to denote their elements (here the k^{th} element of $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{b}_j$), respectively.

1.2 Basic operations

1.2.1 Scalar multiplication, vector addition, and linear combinations

If we multiply a vector by a scalar, $c \in \mathbb{R}$, then we multiply each element of that vector by that scalar. Hence, for an n-dimensional vector $\mathbf{v}$:

$$c\mathbf{v} = \begin{pmatrix} cv_1 \\ cv_2 \\ \vdots \\ cv_n \end{pmatrix} = \mathbf{v}c.$$

The result of such an operation is to stretch ($|c| > 1$) or contract ($|c| < 1$) the original vector away from or towards the origin. Note that if c is negative, we also send the original vector in the opposite direction (flip the vector across the x and y axes).

We can add and subtract two vectors of the *same dimension*. Since subtracting a $n \times 1$ vector $\mathbf{w}$ from a $n \times 1$ vector $\mathbf{v}$ is the same as adding minus one times $\mathbf{w}$ to $\mathbf{v}$, we focus exclusively on vector addition. Adding $\mathbf{w}$ to $\mathbf{v}$ amounts to adding the i^{th} element of $\mathbf{w}$ to the i^{th} element of $\mathbf{v}$ for all $i \in \{1, \dots, n\}$:

$$\mathbf{v} + \mathbf{w} = \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \vdots \\ v_n + w_n \end{pmatrix}.$$

Note that the order of addition makes no difference to the vector sum. The following diagram provides a visual illustration of vector addition and vector subtraction.

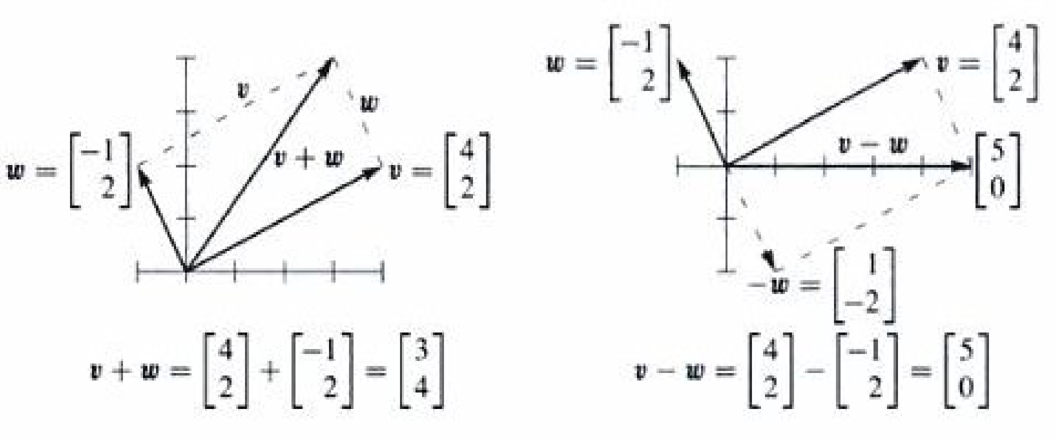


Figure 1: sum of vectors (left); difference of vectors (right)

By adding and subtracting the two vectors $\mathbf{v}$ and $\mathbf{w}$, we are in fact defining a special kind of linear combination of $\mathbf{v}$ and $\mathbf{w}$.

Definition 7. A linear combination of vectors is the sum of a scalar multiple of each vector.

Example 2. *A linear combination of p vectors $\mathbf{v}_1, \ldots, \mathbf{v}_p$ may therefore be written as:*

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

where $(c_s)_{s=1}^{p}$ are scalars taking their values anywhere on the real line, $c_s \in \mathbb{R}\ \forall s$.

In light of Example 2, adding $\mathbf{v}$ and $\mathbf{w}$ is equivalent to taking the linear combination: $c\mathbf{v} + d\mathbf{w}$, where $c = d = 1$. Other special linear combinations are:

- $1\mathbf{v} + (-1)\mathbf{w}$ = difference of vectors, as depicted in Figure 1.
- $0\mathbf{v} + 0\mathbf{w} = \underline{0}$ [2]
- $c\mathbf{v} + 0\mathbf{w}$ = a stretch or contraction of $\mathbf{v}$.

Section 2 considers a particular application where linear combinations of vectors are used.

Exercise 1. Let $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{z}$ be vectors of the same order. Show that

[2] A $n \times 1$ zero vector, 0_n , is an $n \times 1$ vector containing only zeros.

(a) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

(b) $(\mathbf{u} + \mathbf{v}) + \mathbf{z} = \mathbf{v} + (\mathbf{u} + \mathbf{z})$

Exercise 2. For vectors $\mathbf{u}$ and $\mathbf{v}$ of the same order and scalars c and d. Show that

(a) $(c + d)(\mathbf{u} + \mathbf{v}) = c\mathbf{v} + c\mathbf{u} + d\mathbf{v} + d\mathbf{u}$

(b) The zero vector is uniquely determined by the condition that $c\mathbf{0} = \mathbf{0}$ for all finite scalars c.

1.2.2 Matrix addition

Similarly to vector addition, if A and B are two matrices of the same dimension, then:

$$A + B = \begin{pmatrix} a_{11} + b_{11} & \cdot & \cdot & \cdot & a_{1p} + b_{1p} \\ a_{21} + b_{21} & \cdot & \cdot & \cdot & a_{2p} + b_{2p} \\ \cdot & \cdot & & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot \\ a_{n1} + b_{n1} & \cdot & \cdot & \cdot & a_{np} + b_{np} \end{pmatrix}.$$

Note that in our discussions of vector and matrix addition, we required that the relevant objects were of the same dimension. In these cases, we say that $\mathbf{v}$ and $\mathbf{w}$, and A and B, respectively, are *conformable* to addition. This is a particular example of the concept of *conformability*.

Definition 8. If an operation is valid for matrices A and B, then A and B are said to be conformable to that operation.

Specifically, conformability will apply to matrix addition and matrix multiplication.

Example 3. *Let A and B be matrices of order $n \times p$ and $m \times k$, respectively. If $n = m$ and $p = k$, then A and B are conformable to addition.*

Laws of matrix addition: For two conformable matrices A and B:

(i) $A + B = B + A$ (commutative)

(ii) $c(A + B) = cA + cB$ (distributive)

(iii) $A + (B + C) = (A + B) + C$ (associative)

Exercise 3. Show that $(A + B)' = A' + B'$ for any two conformable matrices A and B.

2 Linear equations

Consider a system in which we have three linear equations and three unknowns:

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

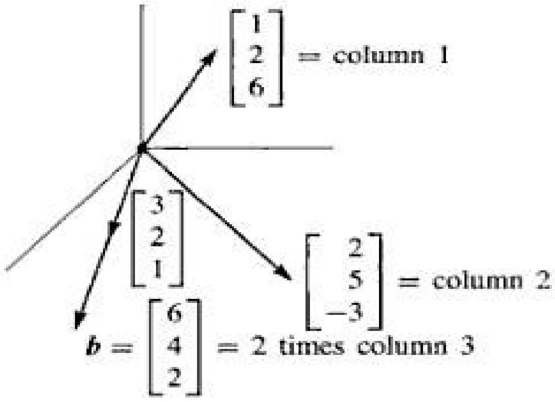
This system of equations can compactly be written as, using matrix notation, $A\mathbf{x} = \mathbf{b}$, where[3]:

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

We wish to find the x, y and z that solve the above system of simultaneous equations. In due course, we will discuss conditions under which the existence of a unique solution for x, y and z is guaranteed. But for the particular system above, we take for granted that the solutions do indeed exist.

There are two ways to visualize the above system. One way is to plot the three equations in the x, y, z-space, with the rows of A providing the coordinates. Visually, each equation in the system defines a plane in the x, y, z-space. Two planes meet in a line, while three planes meet at a point. The coordinates of this point provide the values of x, y and z that solve our system. An alternative way to visualize the problem is to consider the three constituent column vectors of A (i.e. the columns of A) and find the linear combination of the column vectors that produces the vector $\mathbf{b}$.

[3] Note, this way of writing the system involves multiplying a 3×3 matrix by a 3×1 vector, forming another 3×1 vector on the right hand side. In general, for any coefficient, b_i: $\sum a_{ij} x_j = b_i$ (see more below).



In this example, we immediately see that $(x, y, z) = (0, 0, 2)$ produces the required solution:

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

Exercise 4. Write the following system in matrix form and draw the corresponding *column picture*. By simple inspection, identify the **x** vector that solves the system:

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

.

Exercise 5. Write the following system in matrix form and draw the corresponding *column picture*. By simple inspection, identify the **x** vector that solves the system:

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

.

2.1 Linear dependence and rank

The existence and uniqueness of a solution to a system of linear equations may be ascertained using a quantity called the *rank*, whose definition relies on the following:

Definition 9. The vectors in the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ are said to be linearly independent if none of them can be expressed as a linear combination of the other vectors in the set. I.e. there exists no set of scalars $c_1, \ldots, c_p$ – where at least one of them is non-zero – such that:

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

Conversely, if scalars *do* exist such that $\mathbf{0} = c_1\mathbf{v}_1 + \cdots + c_1\mathbf{v}_p$ holds, then $\mathbf{v}_1, \ldots, \mathbf{v}_p$ are said to be linearly dependent.

Exercise 6. Let $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ be linearly independent vectors of order n, show that $(\mathbf{x}+\mathbf{y})$, $(\mathbf{x}+\mathbf{z})$, and $(\mathbf{y}+\mathbf{z})$ are also linearly independent.

Definition 10. The column rank of a matrix A is the maximum number of linearly independent columns of A. Likewise, the row rank is the maximum number of linearly independent rows.

In fact, the column rank and row rank are always equal [see e.g. Magnus and Neudecker (1999)], so we simply refer to the *rank* and use $r(A)$ to denote the rank of A. Clearly, in light of Definition 10, the rank of an $n \times p$ matrix is at most $\min(n, p)$. A matrix that has a rank equal to $\min(n, p)$ is said to have full rank; otherwise, the matrix is rank deficient.

Example 4.

$$r\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} = 1 \quad \textit{whilst} \quad r\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} = 2$$

Example 5.

$$(i)\ r\begin{pmatrix} 1 & 2 & 4 \\ 2 & 4 & 8 \\ 3 & 6 & 12 \end{pmatrix} = 1 \quad (ii)\ r\begin{pmatrix} 1 & 2 & 0 \\ 2 & 4 & 7 \\ 3 & 6 & 2 \end{pmatrix} = 2 \quad (iii)\ r\begin{pmatrix} 1 & 2 & 0 \\ 2 & 3 & 7 \\ 3 & 3 & 2 \end{pmatrix} = 3 \quad (iv)\ r\begin{pmatrix} 2 & 0 \\ 3 & 7 \\ 3 & 2 \end{pmatrix} = 2$$

(i) and (ii) are rank deficient, whilst (iii) and (iv) are full rank. We will now use the concept of rank to determine whether a system of equations has a solution. Consider the following worked example:

Worked example 1.

$$\begin{array}{ccc} A\mathbf{x} & = & \mathbf{b} \\ \begin{pmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} & = & \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix} \end{array}$$

Does $A\mathbf{x} = \mathbf{b}$ have a solution for every possible $\mathbf{b}$ (existence)? The answer is no, but there are some vectors $\mathbf{b}$ that we can solve for; for instance:

- $\mathbf{b} = [0\ \ 0\ \ 0\ \ 0]' \Rightarrow \mathbf{x} = [0\ \ 0\ \ 0]'$
- $\mathbf{b} = [1\ \ 2\ \ 3\ \ 4]' \Rightarrow \mathbf{x} = [1\ \ 0\ \ 0]'$
- $\mathbf{b} = [1\ \ 1\ \ 1\ \ 1]' \Rightarrow \mathbf{x} = [0\ \ 1\ \ 0]'$
- $\mathbf{b} = [2\ \ 3\ \ 4\ \ 5]' \Rightarrow \mathbf{x} = [0\ \ 0\ \ 1]'$.

We immediately see that whenever $\mathbf{b}$ is a linear combination of the columns of A, the system admits at least one solution. Formally, this requires:

$$r\left(\begin{array}{ccc|c} 1 & 1 & 2 & b_1 \\ 2 & 1 & 3 & b_2 \\ 3 & 1 & 4 & b_3 \\ 4 & 1 & 5 & b_4 \end{array}\right) = r\begin{pmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{pmatrix}$$

As the example shows, the condition for a system of linear equations to have a solution is that the rank of the coefficient matrix is equal to the rank of the *augmented matrix*, where the augmented matrix is just the coefficient matrix with the $\mathbf{b}$ vector attached as an extra column.

We now address the question of *uniqueness* of the solution and show that, for a solution to be unique, the coefficient matrix A must be of full column rank; otherwise, there are infinitely many solutions.

Worked example 2. *Consider the previous system, but now look at a particular $\mathbf{b}$ vector:*

$$\begin{array}{ccc} & A\mathbf{x} = \mathbf{b} & \\ \begin{pmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} & = & \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \end{array}.$$

Does $A\mathbf{x} = \mathbf{0}$ have a solution? It has many, for instance:

- $\mathbf{x} = [1\ \ 1\ \ -1]'$
- $\mathbf{x} = [2\ \ 2\ \ -2]'$

In fact, it has infinitely many solutions, all of the form $\mathbf{x} = [c \;\; c \;\; -c]' = c[1 \;\; 1 \;\; -1]', c \in \mathbb{R}$. *Similarly, although* $\mathbf{x} = [1 \;\; 0 \;\; 0]'$ *solves:*

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

so does any $\mathbf{x}$ *of the form:*

$$\begin{pmatrix} 1+c \\ 0+c \\ 0-c \end{pmatrix},$$

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

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

Now consider a matrix A of order $n \times p$, a vector $\mathbf{x}$ of order $p \times 1$ and let $\mathbf{b}$ be a vector of order $n \times 1$. The implications of the four possible scenarios regarding the rank of the matrix for the number of solutions to the linear system are here stated without proof:

1. $r(A) = n = p$ (A is square and has full rank) $A\mathbf{x} = \mathbf{b}$ has one (unique) solution
2. $r(A) = n < p$ (A is short and wide) $A\mathbf{x} = \mathbf{b}$ has ∞ solutions
3. $r(A) = p < n$ (A is tall and thin) $A\mathbf{x} = \mathbf{b}$ has 0 or 1 solution
4. $r(A) < n$ and $r(A) < p$ (A is not full rank) $A\mathbf{x} = \mathbf{b}$ has 0 or ∞ solutions.

Exercise 7. Let A be an $n \times p$ matrix. Suppose you know that there exists an $n \times 1$ vector $\mathbf{a}$, such that if $\mathbf{a}$ is added as an additional column to A, the rank of A increases by 1. I.e., $r(A : \mathbf{a}) = r(A) + 1$. Show that this implies that the rows of A are linearly dependent. Hint: prove by contradiction.

Exercise 8. *[Optional]* Show that it is not true in general that $r(AB) = r(BA)$ for two square matrices A and B.

3 Matrix multiplication

Let A and B be matrices of order $n \times p$ and $k \times m$, respectively. If $p = k$, A and B are conformable to matrix multiplication. In this case, we may define:

$$C := AB,$$

where C is a matrix of order $n \times m$. The elements of C are given by:

$$(C_{ij}) := AB := \begin{pmatrix} \sum_{s=1}^{k} a_{1s}b_{s1} & \cdot & \cdot & \cdot & \sum_{s=1}^{k} a_{1s}b_{sm} \\ \sum_{s=1}^{k} a_{2s}b_{s1} & \cdot & \cdot & \cdot & \sum_{s=1}^{k} a_{2s}b_{sm} \\ \cdot & \cdot & & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot \\ \sum_{s=1}^{k} a_{ns}b_{s1} & \cdot & \cdot & \cdot & \sum_{s=1}^{k} a_{ns}b_{sm} \end{pmatrix}.$$

The entry in row i and column j of $C := \mathbf{AB}$ is therefore given by the i^{th} row of A times the j^{th} column of B; i.e.:

$$(AB)_{ij} := A_i' B_j := \sum_{s=1}^{k} a_{is}b_{sj}.$$

Note that, in general, and in the above example, $AB \neq BA$. Indeed, BA is not even defined unless $n = m$ (in which case it is of dimension $k \times p$). Matrix multiplication does, however, obey the following laws (matrices can be square or rectangular, as long as they are conformable to the relevant operation).

Laws of matrix multiplication:

(i) $C(A + B) = CA + CB$

(ii) $(A + B)C = AC + BC$

(iii) $A(BC) = (AB)C$ (parentheses are not required)

(iii) means that you can multiply either BC first or AB first.[4] To see a situation where (iii) is applicable, let $A = B = C =$ a square matrix, then A^2 times A is equal to A times A^2. By induction: $A^pA = AA^p$. Thus, matrix powers, A^p, follow the same rules as numbers:

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

Exercise 9. 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?

$$BA \qquad AB \qquad ABC \qquad DBA \qquad A(B + C)$$

[4] The proof is a little awkward, so we will not cover it here.

Exercise 10. Let

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

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

Exercise 11. (a) Show that $(AB)' = B'A'$

(b) Show that $(ABC)' = C'B'A'$. Hint: let $D := BC$ and use (a).

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

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

Exercise 12. For square matrices A and B, which of the following statements are true, and why?

(a) $(A + B)^2 = (B + A)^2$

(b) $(A + B)^2 = A^2 + 2AB + B^2$

(c) $(A + B)^2 = A(A + B) + B(A + B)$

(d) $(A + B)^2 = (A + B)(B + A)$

(e) $(A + B)^2 = A^2 + AB + BA + B^2$

Exercise 13. For square matrices A and B, which of the following statements are true, and why?

(a) $(A - B)^2 = (B - A)^2$

(b) $(A - B)^2 = A^2 - B^2$

(c) $(A - B)^2 = A^2 - 2AB + B^2$

(d) $(A - B)^2 = A(A - B) - B(A - B)$

(e) $(A - B)^2 = A^2 - AB - BA + B^2$

In addition to the above, there are three other, equivalent, ways to view matrix multiplication. Let again A be a matrix of order $n \times p$ and let B be a matrix of order $p \times m$. Multiplying A and B in the following ways yield the same result as the above:

1. Matrix A times columns of B: Each column of AB is a combination of the columns of B, so:

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

where $\mathbf{b}_j = (b_{1j}, b_{2j}, \ldots, b_{pj})'$ is the j^{th} column of B.

2. Rows of A times matrix B: Each row of AB is a combination of the rows of A, so:

$$AB = \begin{bmatrix} (\text{---} \; \mathbf{a}_1^* \; \text{---})B \\ \vdots \\ (\text{---} \; \mathbf{a}_n^* \; \text{---})B \end{bmatrix},$$

where $\mathbf{a}_i^* := (a_{i1}, a_{i2}, \ldots, a_{ip})$ is the i^{th} row of A.

3. Block matrix multiplication: Each block of AB is the sum of block-rows of A times block-columns of B; i.e.:

$$AB = \left(\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right)\left(\begin{array}{c|c} B_{11} & B_{12} \\ \hline B_{21} & B_{22} \end{array}\right) = \begin{pmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{pmatrix}.$$

Example 6. *As an important special case of 3 above, let the blocks of A be its p columns and the blocks of B be its p rows, then AB is the sum of the columns of A times the rows of B:*

$$\begin{aligned} AB &= \begin{pmatrix} | & & | \\ \mathbf{a}_1 & \ldots & \mathbf{a}_p \\ | & & | \end{pmatrix}\begin{pmatrix} \text{---} & \mathbf{b}_1^* & \text{---} \\ & \vdots & \\ \text{---} & \mathbf{b}_p^* & \text{---} \end{pmatrix} \\ &= \begin{pmatrix} | \\ \mathbf{a}_1 \\ | \end{pmatrix}\begin{pmatrix} \text{---} & \mathbf{b}_1^* & \text{---} \end{pmatrix} + \ldots + \begin{pmatrix} | \\ \mathbf{a}_p \\ | \end{pmatrix}\begin{pmatrix} \text{---} & \mathbf{b}_p^* & \text{---} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{a_1} & \ldots & \mathbf{a_p} \end{pmatrix}\begin{pmatrix} b_1^* \\ \vdots \\ b_p^* \end{pmatrix}, \end{aligned}$$

where $\mathbf{b}_i^* = (b_{i1}, b_{i2}, \ldots, b_{im})$ is the i^{th} row of B. Besides the above example, block matrices arise naturally in many circumstances. In particular, matrices with identity matrices as their blocks are frequently encountered; in this case, it can be much easier to work with the blocks than the entire matrix.

Exercise 14. What rows or columns or matrices do you multiply to find

(a) the third column of AB?

(b) the first row of AB?

(c) the entry in row three, column 4 of AB?

(d) the entry in row 1, column 1 of CDE?

Exercise 15. An econometrician collects data on the number of years of education and the marital status of n different individuals. He organizes the data for the i^{th} individual as a 2×1 vector $\mathbf{x}_i$ and arranges these n 2-dimensional vectors into a matrix X of order $n \times 2$. Show that $X'X = \sum_i \mathbf{x}_i \mathbf{x}_i'$.

Exercise 16. The econometrician also collects data on income of the same n individuals. Let y be the $n \times 1$ vector with i^{th} element y_i. Show that $X'y = \sum_i \mathbf{x}_i y_i$

Exercise 17. Suppose that for every $i = 1, \ldots, n$, the econometrician applies a weight $1/\sigma_i$ to the i^{th} observation, and arranges these new n 2-dimensional observations into a matrix Z of order $n \times 2$, such that the ij^{th} element of Z is x_{ij}/σ_i. Let Ω be the diagonal matrix with ii^{th} element $\sigma_i \sigma_i = \sigma_i^2$. Show that

$$Z'Z = \sum_i (\mathbf{x}_i/\sigma_i)(\mathbf{x}_i/\sigma_i)' = \sum_i \mathbf{x}_i \sigma_i^{-2} \mathbf{x}_i' = X'\Omega^{-1}X$$

and

$$Z'y = \sum_i (\mathbf{x}_i/\sigma_i) y_i/\sigma_i = \sum_i \mathbf{x}_i \sigma_i^{-2} y_i = X'\Omega^{-1}\mathbf{y}$$

Exercise 18. Multiply A (order 3×3) and I_3 using columns of A times rows of I_3.

Exercise 19. Multiply AB using *columns times rows*:

$$AB = \begin{pmatrix} 1 & 0 \\ 2 & 4 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 3 & 3 & 0 \\ 1 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \begin{pmatrix} 3 & 3 & 0 \end{pmatrix} + \underline{\qquad\qquad} = \underline{\qquad\qquad}$$

3.1 Matrices as operators

Matrices can also be used as (linear) operators, which can be made to act through matrix multiplication. You input a vector $\mathbf{x}$ or matrix B and the output is another vector $A\mathbf{x}$ or matrix AB.

Example 7. *Consider the matrix*

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

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

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad \Rightarrow \quad \mathbf{b} = A\mathbf{x} = \begin{pmatrix} x_2 \\ x_1 \end{pmatrix}.$$

I.e., it permutes the elements of $\mathbf{x}$.

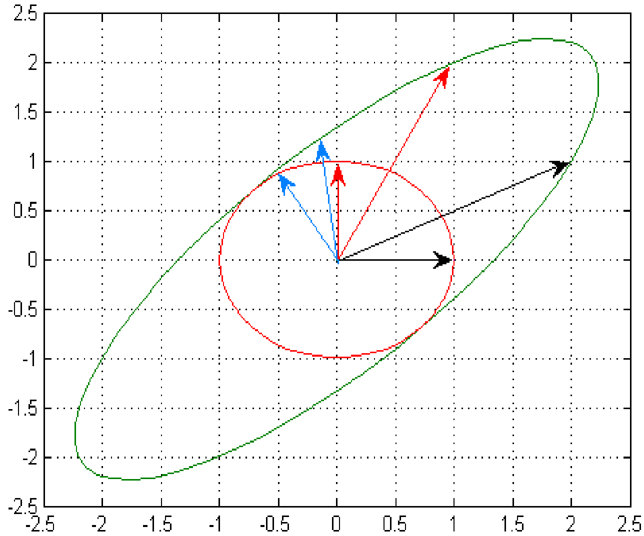
Definition 11. A permutation matrix is a matrix obtained by reordering the rows of the identity matrix.

Exercise 20. Construct all the permutation matrices of order $p = 3$. How many permutation matrices of order $p = 4$ are there?

Example 8. *Below, we plot 50 points on a circle and use the matrix:*

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

to map them to 50 points on an ellipse. Here A is an operator matrix and B is a 2×50 *matrix of coordinates on the circle.* $C = AB$ *produces a* 2×50 *matrix of coordinates on an ellipse.*



Exercise 21. Suppose you are given 50 points on the surface of a sphere. What is the dimension of the matrix you would require in order to map those points to 50 points on the surface of a 3-dimensional ellipsoid?

Exercise 22. You are given the matrix:

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

Construct a matrix A such that:

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

4 Matrix calculus

In this section, we will first define some important matrices frequently encountered in function optimization as well as more advanced matrix calculus.

Let $\mathbf{f}$ be an $m \times 1$ vector function of a $p \times 1$ vector $\mathbf{x}$. The derivative of $\mathbf{f}$, called the Jacobian matrix, is the $m \times p$ matrix:

$$D\mathbf{f}(\mathbf{x}) := \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}'} = \left(\frac{\partial f_i(\mathbf{x})}{\partial x_j} \right),$$

where $i = 1, \ldots, m$ and $j = 1, \ldots, p$.

Consider now a scalar function, $\varphi(\mathbf{x})$, of a $p \times 1$ vector $\mathbf{x}$. Differentiating a scalar valued function with respect to a vector produces another vector, often called the gradient; 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 is a matrix.

Definition 12. Let $\varphi(\mathbf{x})$ be a scalar valued function of the $p \times 1$ vector $\mathbf{x}$. The Hessian matrix is the $p \times p$ matrix of second derivatives:

$$H(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 \varphi(\mathbf{x})}{(\partial x_1)^2} & \cdots & \frac{\partial^2 \varphi(\mathbf{x})}{\partial x_1 \partial x_p} \\ \frac{\partial^2 \varphi(\mathbf{x})}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 \varphi(\mathbf{x})}{\partial x_2 \partial x_p} \\ \vdots & \ddots & \\ \frac{\partial^2 \varphi(\mathbf{x})}{\partial x_p \partial x_1} & \cdots & \frac{\partial^2 \varphi(\mathbf{x})}{(\partial x_p)^2} \end{pmatrix}.$$

To see how differentials work with vector functions, consider first the well-known *one dimensional case*. We have that:

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

defines the derivative of φ at x. Re-writing gives:

$$\varphi(x+u) = \varphi(x) + \varphi'(x)u + r_x(u),$$

where $r_x(u)/u \to 0$ as $u \to 0$. Define the first differential of φ at x (with increment u) as $d\varphi(x;u) = \varphi'(x)u$. For example, for $\varphi(x) = x^2$, $d\varphi(x;u) = 2xu$. Setting aside rigorous justification for the double use of the symbol "d", we will write dx for u. Hence $d\varphi(x) = \varphi'(x)dx$.

In the *vector case*, we similarly have that:

$$\mathbf{f}(\mathbf{x}+\mathbf{u}) = \mathbf{f}(\mathbf{x}) + (D\mathbf{f}(\mathbf{x}))\mathbf{u} + \mathbf{r}_{\mathbf{x}}(\mathbf{u})$$

as well as the first differential being defined as: $d\mathbf{f}(\mathbf{x};\mathbf{u}) = (D(\mathbf{f})(\mathbf{x}))\mathbf{u}$. In addition, assume that the following identification result (that the first derivative can be identified from the first differential) is true[5]:

$$d\mathbf{f}(\mathbf{x}) = A(\mathbf{x})d\mathbf{x} \Leftrightarrow D(\mathbf{f}(\mathbf{x})) = A(\mathbf{x}).$$

We will also sometime require an analogue of the chain rule for differentials of a composite function, $h(x) = g(f(x))$. The chain rule tells us that:

$$Dh(x) = Dg(f(x))Df(x).$$

The equivalent result for differentials (*Cauchy's invariance rule*) is:

$$dh(x;u) = dg(f(x); df(x;u)).$$

For example, if $\varphi(x) = \sin(x^2)$ [$g(y) = \sin y$ and $f(x) = x^2$] then $D\varphi(x) = (\cos x^2)(2x)$. The differential given by:

$$d\varphi = (\cos x^2)dx^2 = (\cos x^2)(2xdx).$$

4.1 Basic rules of matrix calculus

1. $d(\alpha\mathbf{f}) = \alpha d\mathbf{f}$
2. For any two vector functions of the same order $d(\mathbf{f}+\mathbf{h}) = d\mathbf{f} + d\mathbf{h}$
3. For any two vector functions of the same order $d(\mathbf{f}-\mathbf{h}) = d\mathbf{f} - d\mathbf{h}$

[5] It is, but we will not prove it.

4. $d(tr\mathbf{f}) = tr(d\mathbf{f})$ [6]

5. For any two conformable vector functions $d(\mathbf{fg}) = (d\mathbf{f})\mathbf{g} + \mathbf{f}(d\mathbf{g})$

All of the above rules also apply to $m \times q$ matrix-valued functions F of a $n \times p$ matrix A.

In the following, we list a set specific results that should prove useful in your future studies:

(i) Let $\varphi(\mathbf{x}) := \mathbf{a}'\mathbf{x}$. Then $D\varphi = \mathbf{a}'$.

(ii) Let $\varphi(\mathbf{x}) := \mathbf{x}'A\mathbf{x}$. Then applying rule (5) using $\mathbf{f}(\mathbf{x}) = \mathbf{x}'$ and $\mathbf{g}(\mathbf{x}) = A\mathbf{x}$, we get:

$$d\varphi(\mathbf{x}) = (d\mathbf{x})'A\mathbf{x} + \mathbf{x}'Ad\mathbf{x} = \mathbf{x}'(A + A')d\mathbf{x},$$

where the final equality arises from the observation that $(d\mathbf{x})'A\mathbf{x} = ((d\mathbf{x})'A\mathbf{x})'$, so $D\varphi = \mathbf{x}'(A+A')$. Note that if A is symmetric, this becomes $D\varphi = 2\mathbf{x}'A$.

Example 9. *Let*

$$A = \begin{pmatrix} 1 & 3 \\ 3 & 4 \end{pmatrix},$$

such that $\mathbf{x}'A\mathbf{x} = x_1^2 + 4x_2^2 + 6x_1x_2$ (note A is symmetric). Then:

$$\begin{aligned} \frac{\partial \mathbf{x}'A\mathbf{x}}{\partial \mathbf{x}'} &= (2x_1 + 6x_2 \quad 6x_1 + 8x_2) \\ (x_1\ x_2) &\begin{pmatrix} 2 & 6 \\ 6 & 8 \end{pmatrix} = 2\mathbf{x}'A. \end{aligned} \tag{4.1}$$

Equivalently, you can write:

$$\begin{aligned} \frac{\partial \mathbf{x}'A\mathbf{x}}{\partial \mathbf{x}} &= \begin{pmatrix} 2x_1 + 6x_2 \\ 6x_1 + 8x_2 \end{pmatrix} \\ &= \begin{pmatrix} 2 & 6 \\ 6 & 8 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 2A\mathbf{x}. \end{aligned} \tag{4.2}$$

(iii) Let A be a matrix valued function of a scalar, θ, and let $F(A) := A^{-1}$. Since $A^{-1}A = I$, by rule (5) we have:

$$d(A^{-1}A) = (dA^{-1})A + A^{-1}dA = dI = 0_p \quad \text{(a zero matrix of order p)}.$$

Hence, solving for dA^{-1} gives:

$$dA^{-1}(\theta) = -A^{-1}(\theta)dA(\theta)\ A^{-1}(\theta).$$

Thus:

$$\frac{\partial F(A(\theta))}{\partial \theta} = -A^{-1}\frac{\partial A(\theta)}{\partial \theta}A^{-1}.$$

[6] For the definition of the trace operator, see section below.

The above result (iii) is particularly useful in statistics, where covariance matrices often depend on scalar parameters. You will learn about such things in due course. For a thorough treatment of matrix calculus with statistical applications, see Magnus and Neudecker (1999), Chapter 13.

5 Square matrices

In this section, we will only be concerned with matrices whose column dimension is the same as the row dimension. The term *matrix of order p* therefore refers to a square matrix with p rows and p columns.

5.1 Definite matrices

Consider a square matrix, A, of order p. The term $\mathbf{x}'A\mathbf{x} > 0$ is referred to as the quadratic form. The matrix A is said to be:

- positive definite if $\mathbf{x}'A\mathbf{x} > 0$ for any non-null $p \times 1$ vector $\mathbf{x}$.
- positive semi definite or non-negative definite if $\mathbf{x}'A\mathbf{x} \geq 0$ for any non-null $p \times 1$ vector $\mathbf{x}$.
- negative definite if $\mathbf{x}'A\mathbf{x} < 0$ for any non-null $p \times 1$ vector $\mathbf{x}$.
- negative semi definite if $\mathbf{x}'A\mathbf{x} \leq 0$ for any non-null $p \times 1$ vector $\mathbf{x}$.

5.2 Determinants

A determinant is a single number associated with every square matrix. For square matrices with real entries, a geometrical interpretation can be given to the value of the determinant: the absolute value of the determinant is the scale factor by which the area spanned by the columns of the matrix is multiplied under the associated linear transformation, while its sign indicates whether the transformation preserves orientation. Although the determinant has several uses, here we will focus on just a few, which will be discussed in Section 5.2 and Section 5.4. In the subsequent, we will first analyze some of the properties of determinants and then discuss how to calculate them. The determinant of a square matrix A will be denoted $|A|$ or $\det(A)$.

5.2.1 Properties of determinants

Let A and B be square matrices of the same order. The determinant satisfies, inter alia, the following properties:

1. $\det(I) = 1$

2. The determinant of A changes sign when any two rows of A are exchanged.

3. The determinant is a function of each row separately. More specifically:

$$\begin{vmatrix} ta & tb \\ c & d \end{vmatrix} = t \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

   and

$$\begin{vmatrix} a+a' & b+b' \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & d \end{vmatrix}.$$

   Note that it is NOT true that $\det(A+B) = \det(A)+\det(B)$. Property (3) has the important consequence that, for a $p \times p$ matrix A:

$$\det(cA) = c^p \det(A)$$

4. If two rows of A are equal, then $\det(A) = 0$.

5. Subtracting a multiple of one row of A from another leaves $\det(A)$ unchanged.

6. If A has a row of zeros then $\det(A) = 0$

7. $\det(AB) = \det(A)\det(B)$

8. $\det(A') = \det(A)$

Definition 13. A singular matrix is a square matrix whose determinant is zero.

From 5-6 it follows that if a matrix has full rank, then $\det(A) \neq 0$. In fact, the statement also holds the other way around (full rank iff. $\det(A) \neq 0$).

Exercise 23. Prove property (4) using property (2). Hint: force a contradiction by supposing A had two equal rows and $det(A) = a \neq 0$.

Exercise 24. Prove property (5) using properties (3) and (4). Hint: Use both parts of property (3) to separate the relevant determinant into two components and apply property (4).

Exercise 25. Prove property (6) using property (3).

We will not derive properties (7) and (8), but some implications of these are: By (7) and (1), we have:

$$\det(A^{-1})\det(A) = \det(A^{-1}A) = \det(I) = 1,$$

so we immediately have that:

$$\det(A^{-1}) = \frac{1}{\det(A)}.$$

Property (7) also says that:

$$\det(A^2) = \det(A)\det(A) = \det(A)^2.$$

By property (8), all the rules that apply to rows, apply equally to columns.

Exercise 26. Consider the matrices:

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

Using the properties of determinants, show that $AB = 0$ 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.

5.2.2 Computing the determinant

For a (2×2) matrix:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

the determinant has the simple form $ad - bc$.

Exercise 27. Use properties (1), (2), (3), (6) and (8) above to show that:

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

Exercise 28. Use the same method (hence properties (1), (2), (3), (6) and (8) above) to show that:

$$\det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} - a_{13}a_{22}a_{31} + a_{13}a_{21}a_{32}$$

The determinant of a general $n \times n$ matrix can be calculated as:

$$\det(A) = \sum_{j=1}^{p} a_{i,j}(-1)^{i+j} \det(A_{ij}),$$

where A_{ij} is the $(p-1 \times p-1)$ matrix obtained by deleting the i^{th} row and j^{th} column from A.[Use Exercise 28 to verify this!]

5.3 Matrix invertability and inverses

In Section 3 above, we discussed matrices A that operate on another matrix (or vector) B to produce a new matrix, C, through the equation $AB = C$. Whatever operation is performed by A can be undone by its matrix inverse A^{-1}, provided such a matrix exists.

Definition 14. A square matrix A is invertible if there exists a matrix A^{-1} such that:

$$A^{-1}A = I \quad \text{and} \quad AA^{-1} = I$$

Therefore, if $AB = C$, then:

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

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. Although rectangular matrices do have inverses, we will not discuss them (the interested reader should look up Moore-Penrose inverses).

Theorem 1. *If an inverse exists, then it is unique.*

Exercise 29. Prove Theorem 1. Hint: force a contradiction by introducing matrices B and $C \neq B$ that both satisfy the definition of the inverse, i.e.:

$$AB = BA = I \qquad \text{and} \qquad AC = CA = I$$

Theorem 2. *If there exists a non-null vector* $\mathbf{x}$ *such that* $A\mathbf{x} = \mathbf{0}$*, then* A *is not invertible.*

Proof. Suppose not. Then there exists a non-null vector $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{0}$ and A is invertible. However, by Definition 14, $A^{-1}A\mathbf{x} = \mathbf{x} = \mathbf{0}$, but $\mathbf{x}$ is non-null, hence a contradiction. □

Thus, non-invertible matrices take some non-zero vector into zero, and there is no A^{-1} that can recover that vector.

Theorem 3. *If A is a square matrix of order p has an inverse, then the system of equations:*

$$\sum_{j=1}^{p} a_{ij}x_j = b_i \qquad (1 \le i \le p)$$

has a unique solution, for each choice of b_i.

Exercise 30. Use Theorem 1 to prove Theorem 2.

Exercise 31. Construct a matrix A that multiplies the vector $(3,\ -1)'$ to produce the zero vector $(0,\ 0)'$. What do you notice about the matrix A? Compute its determinant.

In Exercise 31 above, we found that a non-invertible matrix A has a determinant which is equal to zero. It turns out that this is a general property of singular matrices. In particular, we have:

Theorem 4. *The square matrix A is invertible if and only if it is non-singular (i.e. if $\det(A) \neq 0$).*

For a proof, see Chapter 5.1 in Abadir and Magnus (2005), but notice that $\det(A)$ appears as a scaling in Equation 5.1, which is consistent with the above theorem.

Theorem 5. *The square matrix A of order p is invertible if and only if $r(A) = p$.*

Exercise 32. Consider the matrix:

$$\begin{pmatrix} 1 & 2 & 0 \\ 2 & 4 & 7 \\ 3 & 6 & 2 \end{pmatrix},$$

and take note of its rank. Show, using Theorem 4, and the properties of determinants, that the matrix in the above display is non-invertible.

Theorem 6. *The product AB has an inverse if and only if A and B both have an inverse (and are of the same dimension).*

Exercise 33. Suppose A and B are matrices of the same order and both have an inverse. Show that:

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

Exercise 34. For any nonsingular square matrix A, show that:

(a) $(cA)^{-1} = (1/c)A^{-1}$ $(c \neq 0)$

(b) $(A^{-1})' = (A')^{-1}$

(c) $(A^{-1})^{-1} = A$.

5.3.1 Computing the inverse

In general, the formula for the inverse of a square matrix A is given by:

$$A^{-1} = \frac{1}{\det(A)} C', \tag{5.1}$$

where C is the matrix of *cofactors* of A.

Definition 15. The matrix of cofactors or cofactor matrix of A is the matrix whose $(ij)^{th}$ element is the one obtained by multiplying $(-1)^{i+j}$ by the ij^{th} minor of A, denoted M_{ij}. I.e., $C = (C_{ij})$ where $C_{ij} := (-1)^{i+j} M_{ij}$

Definition 16. The ij^{th} minor of A is the determinant of the matrix that remains when the i^{th} row and j^{th} column of A is removed [see also formula on how to compute a determinant].

For instance, the elements in the (1,1) and (2,3) positions of C are obtained as $C_{11} = M_{11} = \det(A_{11})$ and $C_{23} = (-1)M_{23} = (-1)\det(A_{23})$, where the 2×2 matrices A_{11} and A_{23} are given by

$$A_{11} := \begin{pmatrix} \cancel{a_{11}} & \cancel{a_{12}} & \cancel{a_{13}} \\ \cancel{a_{21}} & a_{22} & a_{23} \\ \cancel{a_{31}} & a_{32} & a_{33} \end{pmatrix} \quad \text{and} \quad A_{23} := \begin{pmatrix} a_{11} & a_{12} & \cancel{a_{13}} \\ \cancel{a_{21}} & \cancel{a_{22}} & \cancel{a_{23}} \\ a_{31} & a_{32} & \cancel{a_{33}} \end{pmatrix}.$$

Exercise 35. Let A and B be invertible square matrices of the same order and let also $C := AB$ be invertible. Find an expression for A^{-1} in terms of C^{-1} and B.

Exercise 36. Show that a matrix with a column of zeros cannot have an inverse.

Exercise 37. Use Equation 5.1 to calculate the inverse of:

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

5.3.2 Special inverses

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

Denote by E_{ij} the identity matrix with rows i and j interchanged; by $E_i(\gamma)$ the identity matrix whose i^{th} row is multiplied by $\gamma \neq 0$; and by $E_i(\gamma|j)$ the identity matrix where γ times row j is added to a row i $(i \neq j)$. For instance, with $p = 3$:

$$E_{13} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \qquad E_2(7) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad E_3(5|2) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 5 & 1 \end{pmatrix}$$

The following three results are stated without proof.

(i) $E_{ij}^{-1} = E_{ij}$

(ii) $E_i^{-1}(\gamma) = E_i(\gamma^{-1})$

(iii) $E_i^{-1}(\gamma|j) = E_i(-\gamma|j)$

Exercise 38. Suppose A is invertible and you exchange its first two rows to obtain B. Explain why the new matrix is invertible. How would you find B^{-1} from A^{-1}.

5.4 Trace

Definition 17. The trace of a matrix of order $(p \times p)$ is the sum of the elements along its main diagonal:

$$tr(A) := \sum_{i=1}^{p} a_{ii}.$$

The trace is a linear operator, i.e.:

$$tr(A + B) = tr(A) + tr(B) \quad \text{and} \quad tr(cA) = ctr(A) \ ,$$

both of which may be shown by direct calculation. The following two properties can also easily be shown:

1. $tr(A') = tr(A)$.
2. $tr(AB) = tr(BA)$.

In property (2), AB and BA must be square, though need not be of the same order.

Exercise 39. Referring to your computations in Exercise 10 above:

(a) Check that

$$tr(A'B) = tr(BA') = tr(AB') = tr(B'A).$$

(b) Prove that properties (1) and (2) of the trace hold for arbitrary $n \times p$ matrices A and B.

5.5 Eigenvectors and eigenvalues

In Section 3.1, we briefly discussed matrices as linear operators, and likened matrices to functions which take a vector $\mathbf{x}$ as their input and produce as their output another vector $A\mathbf{x}$. In this section, we will be concerned with vectors $\mathbf{x}$ that come out of the matrix multiplication simply as scaled version of the original vector.

Definition 18. The eigenvectors of a square matrix A and their corresponding eigenvalues are the special vectors $\mathbf{x}$ and corresponding numbers λ that satisfy:

$$A\mathbf{x} = \lambda\mathbf{x}, \tag{5.2}$$

where λ can be any real (or imaginary) number. Since $\lambda = -1$ is a permissible value, we also allow for scaled versions of the original vector that are in the exact opposite direction. Note, $\lambda = 0$ is allowed, so $\mathbf{x} = \mathbf{0}$ always satisfies Equation 5.2.

Example 10. *Consider the permutation matrix in Example 7. It is not difficult to spot two vectors* $\mathbf{x}$ *that would yield* $A\mathbf{x} = \lambda\mathbf{x}$*:*

$$\mathbf{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \Rightarrow \quad A\mathbf{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \Rightarrow A\mathbf{x} = \mathbf{x} \;\; (\lambda = 1)$$

$$\mathbf{x} = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \quad \Rightarrow \quad A\mathbf{x} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \Rightarrow A\mathbf{x} = -\mathbf{x} \;\; (\lambda = -1).$$

A square matrix A of order p will have p eigenvalues. To find them, the following results can be useful:

$$tr(A) = \sum_{s=1}^{p} \lambda_s \quad \text{and} \quad \det A = \prod_{s=1}^{p} \lambda_s.$$

More generally, we may solve for the eigenvector-eigenvalue pairs of A by re-writing 5.2 as:

$$(A - \lambda I_p)\mathbf{x} = 0. \tag{5.3}$$

We now know that $(A - \lambda I_p)$ must be singular; otherwise, the only $\mathbf{x}$ satisfying 5.2 would be the zero vector. The singularity of $(A - \lambda I_p)$ implies that the $\det(A - \lambda I_p) = 0$. Therefore, we can solve for the p values of λ that yield zero determinant for $(A - \lambda I)$ and then solve Equation 5.2 for $\mathbf{x}$. Note that there is a whole line of eigenvectors satisfying Equation 5.2; the usual procedure is to write down the most natural set of eigenvalues, or the ones satisfying $\mathbf{x}'\mathbf{x} = 1$.

Exercise 40. Find the two eigenvalues of the matrix :

$$B = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$$

by finding the characteristic roots of an equation of the form :

$$\alpha\lambda^2 + \beta\lambda + \gamma = 0.$$

Deduce the two eigenvectors of B by inspection.

Exercise 41. Let A be the permutation matrix in Example 10, and write the matrix B of Exercise 40 as $B = (A + 3I)$. Algebraically deduce that the eigenvalues of A are three less than the eigenvalues of B and that the eigenvectors are unchanged.

Exercise 42. Using $A\mathbf{x} = \lambda\mathbf{x}$, show that:

(a) λ^2 is an eigenvalue of A^2

(b) λ^{-1} is an eigenvalue of A^{-1}

(c) $\lambda + 1$ is an eigenvalue of $A + I$.

Exercise 43. Let

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

(a) Show that $\det(A - \lambda I) = \lambda^2 - tr(A)\lambda + \det(A)$.

(b) Find an expression for the two eigenvalues of A in terms of a, b, c and d. Show that the characteristic roots satisfy:

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

as well as $\lambda_1\lambda_2 = \det(A)$ and $\lambda_1 + \lambda_2 = tr(A)$.

6 Diagonalisation and powers of A

Eigenvectors and eigenvalues are extremely useful objects with many applications in statistics and other fields. We will consider just one of these. Let S be the matrix whose columns are the eigenvectors of A.

Suppose A possesses p linearly independent eigenvectors such that S is invertible. By Definition 18:

$$AS = A\begin{pmatrix} | & & | \\ \mathbf{x}_1 & \dots & \mathbf{x}_p \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ \lambda_1\mathbf{x}_1 & \dots & \lambda_p\mathbf{x}_p \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ \mathbf{x}_1 & \dots & \mathbf{x}_p \\ | & & | \end{pmatrix}\begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \cdot & & \vdots \\ \vdots & & \cdot & \\ 0 & \dots & & \lambda_p \end{pmatrix} := S\Lambda.$$

This shows that $S^{-1}AS$ diagonalises A, in the sense that it produces the diagonal matrix Λ which has the eigenvalues $\lambda_1, \dots, \lambda_p$ along its main diagonal. Post-multiplying $AS = S\Lambda$ by S^{-1} gives us the eigenvalue decomposition:

$$A = S\Lambda S^{-1}\,. \tag{6.1}$$

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

Both directions of the proof follow trivially using the representation in 6.1:

$$A^k = \underbrace{S\Lambda S^{-1}S\Lambda S^{-1}\cdots S\Lambda S^{-1}}_{k \text{ factors}} = S\Lambda^k S^{-1}.$$

This result is important for analysing the convergence properties of certain kinds of dependent series of the form:

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

where recursive substitution gives:

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

for some initial vector $\mathbf{u}_0$. As a useful example of this, consider Worked Example 3.

Worked example 3. *The first 8 numbers of the Fibonacci sequence are $0, 1, 1, 2, 3, 5, 8, 13, \dots$ where the k^{th} 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 potentially useful question to ask is "how fast are the Fibonacci numbers growing?" The answer to this question lies in the eigenvalues. Let:*

$$u_k = \begin{pmatrix} F_{k+1} \\ F_k \end{pmatrix}.$$

Then the Fibonacci sequence satisfies :

$$\mathbf{u}_{k+1} = A\mathbf{u}_k \quad \textit{with} \quad A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

We have:

$$\det(A - \lambda I) = \det \begin{pmatrix} 1-\lambda & 1 \\ 1 & -\lambda \end{pmatrix} = \lambda^2 - \lambda - 1.$$

Setting to zero and solving gives:

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

[check that $\lambda_1 + \lambda_2 = tr(A) = 1$ and $\lambda_1 \lambda_2 = \det(A) = -1$]. Since the largest eigenvalue is the one controlling the growth, the Fibonacci numbers are growing at the rate $(1 + \sqrt{5})/2$.

Exercise 44. Assuming you can write the initial vector $\mathbf{u}_0$ as a linear combination of the eigenvectors, find an expression for the k^{th} element of an arbitrary sequence of the form $\mathbf{u}_{k+1} = A\mathbf{u}_k$.

Exercise 45. Using the expression you found in Exercise 44, find the 100^{th} element of the Fibonacci sequence in Worked Example 3.

6.1 Symmetric and positive definite matrices

Symmetric matrices are special in that their eigenvector matrix, S, is orthogonal whilst positive (semi) definite matrices are special in that their eigenvalues are positive (non-negative) (Exercise 40).

Definition 19. An orthogonal matrix is a square matrix with *orthonormal columns*. I.e., for an orthogonal matrix Q of order p we have, letting $\mathbf{q}_j$ denote an arbitrary column of Q, $\mathbf{q}_j'\mathbf{q}_j = 1$ for $j = 1, \ldots, p$ and $\mathbf{q}_i'\mathbf{q}_j = 0$ for all $i \neq j$.

Exercise 46. Show that a positive (semi) definite matrix has positive (non-negative) eigenvalues.

Note that since $Q'Q = I_p$, orthogonal matrices have the useful property that $Q' = Q^{-1}$, implying that the eigendecomposition of a symmetric matrix, A, becomes the spectral decomposition:

$$A = S\Lambda S^{-1} = Q\Lambda Q'.$$

Exercise 47. Show that if a matrix is symmetric and has positive (non-negative) eigenvalues, then it must be positive (semi) definite.

References

A Summary of key results

A.1 General $p \times q$ matrix A

(i) $A = (a_{ij}) \Rightarrow A^T = (a_{ji})$.

(ii) $cA = (ca_{ij})$ if c a scalar.

(iii) $A + B = (a_{ij} + b_{ij})$ if B is the same dimension as A.

(vi) $AB = \left(\sum_{k=1}^{q} a_{ik}b_{kj}\right)$ if B has q rows.

(vii) $AB \neq BA$.

(viii) $(AB)^T = B^T A^T$.

(ix) $r(A)$ is the rank of A, i.e. the maximum numebr of independent columns.

(x) If $r(A) = q$, then A is full column rank.

A.2 Square A

(i) $tr(A) = \sum_{i=1}^{p} a_{ii}$ is the trace of A.

(ii) $A = diag(a_{11}, \dots, a_{pp})$ means that ij^{th} entries of A are zero $\forall i \neq j$.

(iii) $tr(BC) = tr(CB)$ if BC (and necessarily CB) are square.

(vi) $tr(A^T) = tr(A)$.

(vii) $\det(AB) = \det(A)\det(B)$ for scalar conformable A and B.

(viii) $\det(A^T) = \det(A)$.

(ix) $\det(cA) = c^p \det(A)$ for scalar c.

(x) $\det(A^{-1}) = 1/\det(A)$.

(xi) $(AB)^{-1} = B^{-1}A^{-1}$ if A and B are conformable and invertable (have non-zero determinant).

(xii) $tr(A) = \sum_{i=1}^{p} \lambda_i$ where $(\lambda_i)_{i=1,\dots,p}$ are the eigenvalues of A.

(xiii) If $A^T A = I_p$, A is called orthogonal (has orthonormal columns). Then $A^{-1} = A^T$.

A.3 Non-Negative Definite A

(i) $\mathbf{x}^T A\mathbf{x} \geq 0$ for $\mathbf{x}$ not equal to the zero vector.

(ii) $B^T AB \geq 0$ for all B.

A.4 (Symmetric) Positive Definite A

(i) $\mathbf{x}^T A\mathbf{x} > 0$ for x not equal to the zero vector.

(ii) $B^T AB > 0$ for all B of full column rank.

(iii) $A^{-1} > 0$

(iv) $rank(A) = p$

(v) $\Lambda = diag(\lambda_1(A), \ldots, \lambda_p(A))$ positive definite.

(vi) A is orthogonally diagonalisable, i.e. $A = Q\Lambda Q^T$, where Q is the matrix with the (orthonormal) eigenvectors of A as its columns and Λ has eigenvalues of A along the diagonal.

(vi) $rank(A) = \sum_{j=1}^{p} \mathbb{I}\{\lambda_j > 0\}$

A.5 Elementary Matrices

Denote by E_{ij} the identity matrix with rows i and j interchanged; by $E_i(\gamma)$ the identity matrix whose i^{th} row is multiplied by $\gamma \neq 0$; by $E_i(\gamma|j)$ the identity matrix where γ times row j is added to a row i $(i \neq j)$. E.g. with $p = 3$,

$$E_{13} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \qquad E_2(7) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad E_3(5|2) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 5 & 1 \end{pmatrix}$$

(i) $E_{ij}^{-1} = E_{ij}$

(ii) $E_i^{-1}(\gamma) = E_i(\gamma^{-1})$

(iii) $E_i^{-1}(\gamma|j) = E_i(-\gamma|j)$

A.6 Matrix Differentiation

(i) $\frac{\partial}{\partial \mathbf{x}} A\mathbf{x} = A^T$.

(ii) $\frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T A\mathbf{y} = A\mathbf{y}$.

(iii) $\frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T A\mathbf{x} = (A + A^T)\mathbf{x}$ ($= 2A\mathbf{x}$ if A symmetric).

(vi) $\frac{\partial}{\partial A} \mathbf{x}^T A\mathbf{x} = \mathbf{x}\mathbf{x}^T$

(vii) Let A be a function of a scaler θ. Then $\frac{\partial}{\partial \theta} A^{-1} = -A^{-1} \frac{\partial A}{\partial \theta} A^{-1}$