Matrix Methods – Statistics Honours
A Recap of Linear Algebra and Matrix Fundamentals

Mr Stefan S. Britz
Department of Statistical Sciences
University of Cape Town

1 Nomenclature and Revision of Basics

1.1 Nomenclature

The following notation will be used throughout the course:

Concept	Symbol	Typed text	Handwritten
Scalar	a	Lower case	Lower case
Matrix	$\boldsymbol{A}\colon n \times p$	Bold upper case	Upper case
Column vector	$\boldsymbol{a}\colon n \times 1$	Bold lower case	Lower case underlined
Row vector	$\boldsymbol{a}'\colon 1 \times p$	Bold lower case with transpose	Lower case underlined with transpose
ij^{th} element of matrix A	a_{ij}		
j^{th} column of matrix A	$\boldsymbol{a}_j\colon n \times 1$		
j^{th} column of matrix A, written as a row vector	$\boldsymbol{a}_j'\colon n \times 1$		
i^{th} row of matrix A	$\boldsymbol{a}_{(i)}'\colon 1 \times p$		
i^{th} row of matrix A, written as a column vector	$\boldsymbol{a}_{(i)}\colon 1 \times p$		

The matrix $\boldsymbol{A}$ is sometimes also written as $\{a_{ij}\}$.

1.2 Special matrices

The following are some special matrices:

Square matrix

$\boldsymbol{A}\colon n \times n$, the number of rows = number of columns.

Diagonal matrix

$\boldsymbol{A}: n \times n$ where $a_{ij} = 0 \; \forall \; i \neq j$. Sometimes written $\boldsymbol{A} = diag(d_i)$ where $d_i = a_{ii}$.

Symmetric matrix

$\boldsymbol{A}: n \times n$ where $a_{ij} = a_{ji}$.

Upper (Lower) triangular matrix

$\boldsymbol{A}: n \times n$ where $a_{ij} = 0$ for $i > j (j > i)$.

Identity matrix

$\boldsymbol{I}: n \times n = diag(\mathbf{1})$ where $\mathbf{1}$ is a vector of ones.

1.3 Operations

The following are common simple operations on the matrix $\boldsymbol{A}$:

Transpose

If $\boldsymbol{A} = \{a_{ij}\}$, then its transpose is $\boldsymbol{A}' = \boldsymbol{A}^T = \{a_{ji}\}$. Note that for symmetric matrices $\boldsymbol{A} = \boldsymbol{A}'$.

Addition

$\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B}$ is the matrix $\{a_{ij} + b_{ij}\}$.

Scalar multiplication

$\boldsymbol{B} = c\boldsymbol{A}$ is the matrix $\{ca_{ij}\}$.

Trace

The trace of the square matrix $\boldsymbol{A}: n \times n$ is defined as the sum of the diagonal:

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

Partitioning

It is often convenient to look at rectangular subsets of the matrix elements, which are themselves matrices (or vectors/scalars) of lower dimensions. For example, $\boldsymbol{A}: r \times c$ could be split up, or "partitioned" as follows:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11}: p \times q & \boldsymbol{A}_{12}: p \times (c-q) \\ \boldsymbol{A}_{21}: (r-p) \times q & \boldsymbol{A}_{22}: (r-p) \times (c-q) \end{bmatrix}$$

where, for example, $\boldsymbol{A}_{11}$ is the matrix consisting of the first p rows and the first q columns of $\boldsymbol{A}$ only.

1.4 Linear algebra

The concept of a matrix is intimately linked with the linear algebraic concept of the linear transformation of one vector to another. Let $\boldsymbol{x}: n \times 1$ and $\boldsymbol{y}: m \times 1$ be n- and m-dimensional vectors respectively, such that each element of $\boldsymbol{y}$ is related to $\boldsymbol{x}$ through the linear relationship:

$$y_i = \sum_{j=1}^{n} a_{ij} x_j, \quad i = 1, \ldots, m. \tag{1}$$

The convention is to represent this transformation by the vector-matrix equation $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ where $\boldsymbol{A}$ is the matrix $\{a_{ij}\}$. This immediately defines the concept of multiplication of a matrix by a vector: each element of the result must be as given by 1. This immediately also defines matrix multiplication, since any matrix can be viewed as a collection of column (or row) vectors.

For an arbitrary matrix $\boldsymbol{B}$, let $\boldsymbol{b}_j$ be the vector consisting of the elements of the j^{th} column of $\boldsymbol{B}$. Therefore, $\boldsymbol{B}$ can be viewed as partitioned in the form $\boldsymbol{B} = \begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \cdots & \boldsymbol{b}_p \end{bmatrix}$. The only consistent definition for the product of an $m \times n$ matrix $\boldsymbol{A}$ and an $n \times p$ matrix $\boldsymbol{B}$ would then be $\boldsymbol{AB} = \begin{bmatrix} \boldsymbol{A}\boldsymbol{b}_1 & \boldsymbol{A}\boldsymbol{b}_2 & \cdots & \boldsymbol{A}\boldsymbol{b}_p \end{bmatrix}$. This yields the standard definition of $\boldsymbol{C} = \boldsymbol{AB}$ as the $m \times p$ matrix with (i, j)-th element given by:

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}. \tag{2}$$

Note that the multiplication rules for matrices defined in terms of their elements apply also to manipulation on the component matrices of partitioned matrices.

For example,

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

where of course the partitioning must be such that all the individual matrix multiplications are allowed or, in other words, are conformable.

Side note

A square matrix can be multiplied by itself as often as desired to give $\boldsymbol{A}^k = \boldsymbol{A}\boldsymbol{A}\ldots\boldsymbol{A}$ (k times). This has a particularly important application in the study of Markov chains. Suppose that the state of some system, e.g. stock in inventory, no-claim bonus status of an insurance policy, or condition of machinery, changes over time according to a Markov process. I.e. the occurrence of a particular state at a given point in time is a random event depending only on the immediately preceding state.

Now define the matrix $\boldsymbol{P} = \{p_{ij}\}$ where $p_{ij} = \Pr[\text{current state} = j | \text{previous state} = i]$, i.e. the probability of transitioning from state i to j in one step. The corresponding transition probabilities from states at the beginning to states at the end of k such steps are given by the elements of $\boldsymbol{P}^k$.

Exercise 1.1

Show that if $\boldsymbol{D} = diag(d_i)$, then $\boldsymbol{DA} = \{d_i a_{ij}\}$; note that this implies that $\boldsymbol{IA} = \boldsymbol{A}$, where $\boldsymbol{I}$ is the identity matrix.

Exercise 1.2

Show that

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

(b) $tr(\boldsymbol{AB}) = tr(\boldsymbol{BA})$

1.5 Commutativity, associativity and distributive laws

The following laws apply to matrix operations:

Associativity

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

Distributivity

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

and

$$(\boldsymbol{B}+\boldsymbol{C})\boldsymbol{A} = \boldsymbol{BA}+\boldsymbol{CA}$$

Commutativity

Matrices are commutative only in addition:

$$\boldsymbol{A}+\boldsymbol{B} = \boldsymbol{B}+\boldsymbol{A}$$

Therefore, in general:

$$\boldsymbol{AB} \neq \boldsymbol{BA}$$

1.6 Other types of matrix multiplication

The standard definition of matrix multiplication as used above derives from the concept of the matrix as a linear mapping from one vector space into another. Two other "products" derive from different considerations.

Kronecker product

The direct or Kronecker product of any two matrices (there are no size matching restrictions), say $\boldsymbol{A}\colon p \times q$ and $\boldsymbol{B}\colon m \times n$, is defined as the $pm \times qn$ matrix

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \dots & a_{1q}\boldsymbol{B} \\ a_{12}\boldsymbol{B} & a_{22}\boldsymbol{B} & \dots & a_{2q}\boldsymbol{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1}\boldsymbol{B} & a_{p2}\boldsymbol{B} & \dots & a_{pq}\boldsymbol{B} \end{bmatrix}$$

Note that $(\boldsymbol{A} \otimes \boldsymbol{B})' = \boldsymbol{A}' \otimes \boldsymbol{B}'$.

Hadamard product

The element-wise product or Hadamard product of two matrices of the same size, $\boldsymbol{A} \circ \boldsymbol{B}$, is the matrix with elements $\{a_{ij}b_{ij}\}$.

1.7 Properties of some special matrices

Symmetric matrices

(a) If $\boldsymbol{A}$ and $\boldsymbol{B}$ are both symmetric, then $(\boldsymbol{AB})' = \boldsymbol{B}'\boldsymbol{A}' = \boldsymbol{BA} \neq \boldsymbol{AB}$ in general. Therefore, $\boldsymbol{AB}$ is not (necessarily) symmetric.

(b) For any matrix $\boldsymbol{A}$, both $\boldsymbol{A}'\boldsymbol{A}$ and $\boldsymbol{AA}'$ are symmetric. This also applies to the $n \times n$ matrix $\boldsymbol{xx}'$ where $\boldsymbol{x}$ is an n-vector.

(c) $tr(\boldsymbol{A}'\boldsymbol{A})$ is the sum of the squares of the elements of $\boldsymbol{A}$.

Exercise 1.3

Show that for a square matrix $\boldsymbol{A}: n \times n$, $tr(\boldsymbol{A}'\boldsymbol{A}) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2$.

Matrices of ones

Define $\mathbf{1}$ as the vector consisting only of 1's and $\boldsymbol{J} = \mathbf{11}'$ as the square matrix consisting only of 1's.

Exercise 1.4

Show that if $\boldsymbol{x}$ and $\mathbf{1}$ have length n, then

(a) $\mathbf{1}'\mathbf{1} = n$

(b) $\mathbf{1}'\boldsymbol{x} = \sum_{i=1}^{n} x_i$

(c) $\left(\boldsymbol{I} - \frac{1}{n}\boldsymbol{J}\right)\boldsymbol{x}$ is the vector with elements $x_i - \bar{x}$ where $\bar{x}$ is the mean of the elements of $\boldsymbol{x}$. The matrix $\boldsymbol{I} - \frac{1}{n}\boldsymbol{J}$ is termed the centring matrix.

Idempotency

A matrix $\boldsymbol{K}$ such that $\boldsymbol{K}^2 = \boldsymbol{K}$ is said to be idempotent.

Exercise 1.5

Show that the centring matrix is idempotent. (Applying centring to an already centred vector does nothing further to it!)

Orthogonal vectors

Vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are orthogonal if $\boldsymbol{x}'\boldsymbol{y} = \boldsymbol{y}'\boldsymbol{x} = 0$.

Norm

The norm or length of $\boldsymbol{x}$ is defined by $||\boldsymbol{x}|| = \sqrt{\boldsymbol{x}'\boldsymbol{x}}$. The vector $\boldsymbol{u} = \dfrac{\boldsymbol{x}}{||\boldsymbol{x}||}$ has a norm of 1 and is said to be normalised.

Orthonormal vectors

If $\boldsymbol{x}$ and $\boldsymbol{y}$ are orthogonal and normalised, they are an orthonormal pair of vectors. Clearly, if the columns of the matrix $\boldsymbol{P}$ are an orthonormal set of vectors, then $\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{I}$. Similarly, if the rows form an orthonormal set then $\boldsymbol{P}\boldsymbol{P}' = \boldsymbol{I}$.

Orthogonal matrix

If the matrix $\boldsymbol{P}$ is square and the rows and columns are mutually orthonormal, $\boldsymbol{P}$ is an orthogonal matrix.

Bilinear and quadratic forms

Observe that

$$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{y} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i y_j$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are $n \times 1$ vectors and $\boldsymbol{A}$ an $n \times n$ matrix. This is termed a bilinear form and, if $\boldsymbol{x} = \boldsymbol{y}$, a quadratic form.

Positive definiteness

A symmetric matrix $\boldsymbol{A}$ is:

- Positive definite if $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0$ for any $\boldsymbol{x} \neq 0$.
- Positive semi-definite if $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \geq 0$ and $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = 0$ for at least one $\boldsymbol{x} \neq 0$.
- Negative definite if $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} < 0$ for any $\boldsymbol{x} \neq 0$.

Examples:

(a) $\boldsymbol{J}\colon 2 \times 2$ yields $\boldsymbol{x}'\boldsymbol{J}\boldsymbol{x} = (x_1 + x_2)^2$ which is non-negative, but zero whenever $x_1 = -x_2$. Thus $\boldsymbol{J}$ is positive semi-definite.

(b) The matrix $\boldsymbol{A} = \begin{bmatrix} 2 & 1 \\ -2 & -1 \end{bmatrix}$ yields $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = (x_1 - x_2)(2x_1 + x_2)$ and hence $\boldsymbol{A}$ is indefinite.

Exercise 1.6

Show that

$$\boldsymbol{x}'\left(\boldsymbol{I} - \frac{1}{n}\boldsymbol{J}\right)\left(\boldsymbol{I} - \frac{1}{n}\boldsymbol{J}\right)\boldsymbol{x} = \boldsymbol{x}'\left(\boldsymbol{I} - \frac{1}{n}\boldsymbol{J}\right)\boldsymbol{x} = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

2 Determinants

2.1 Definitions

Note that other, more complex, definitions of determinants exist than what will be covered here, but they imply the definitions we will use. We will start with the base definition for a 2×2 matrix and illustrate a simple and useful geometric interpretation thereof, to give some sense of understanding to its meaning as opposed to just an arbitrary formula:

$$det(\boldsymbol{A}) = |\boldsymbol{A}| = a_{11}a_{22} - a_{12}a_{21} \tag{3}$$

Suppose that $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, where both $\boldsymbol{y}$ and $\boldsymbol{x}$ are 2-dimensional vectors. Consider the unit square in $\boldsymbol{x}$-space defined by the points $(0,0); (0,1); (1,0)$ and $(1,1)$. The matrix $\boldsymbol{A}$ linearly transforms ("pulls") the coordinates in $\boldsymbol{x}$-space to a new set in $\boldsymbol{y}$-space:

$$y_1 = a_{11}x_1 + a_{12}x_2$$
$$y_2 = a_{21}x_1 + a_{22}x_2$$

It follows that all the points $\boldsymbol{x}$ inside the unit square transform to points $\boldsymbol{y}$ inside a parallelogram, as illustrated in Figure 1. In this example, we specifically used $a_{11} > a_{21} > 0$ and $a_{22} > a_{12} > 0$.

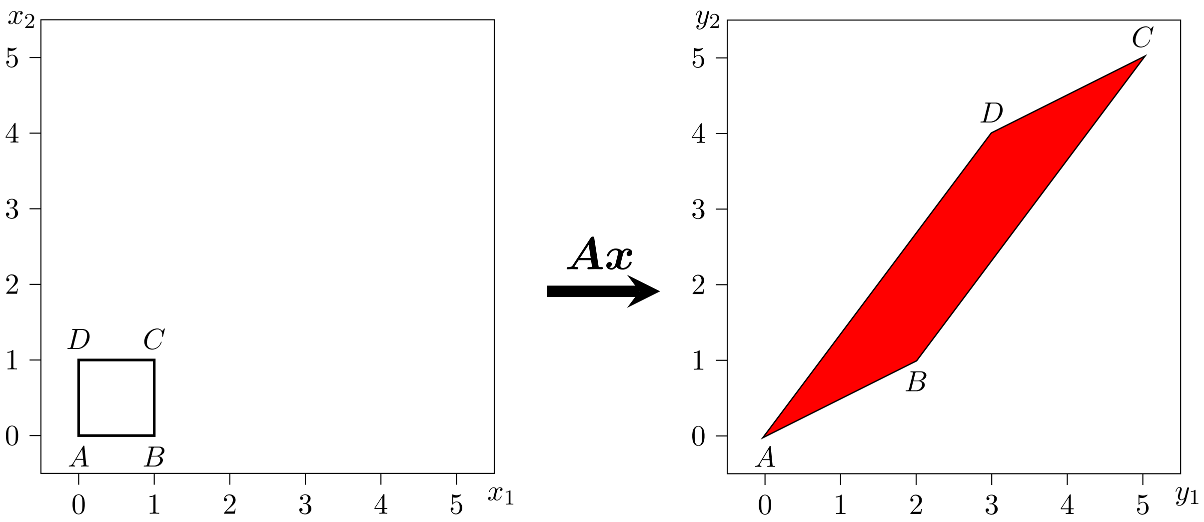
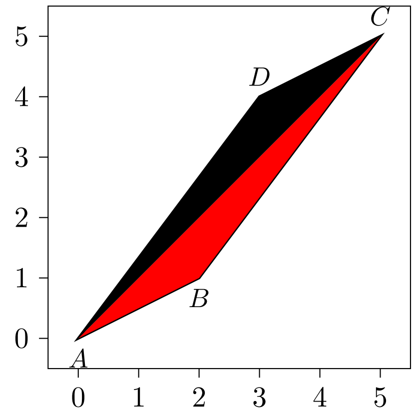


Figure 1: The original coordinates forming a square (on the left) are transformed to a new set of coordinates by matrix multiplication, yielding the parallelogram on the right.

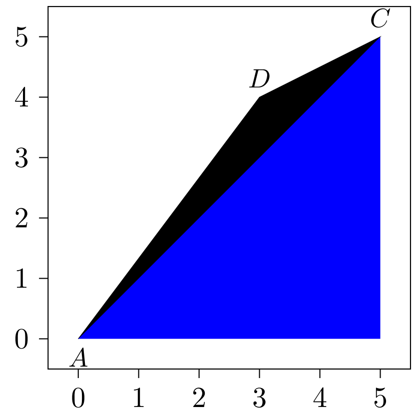
Showing the transformation of each of the corner points explicitly, where $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix}$:

	Initial coordinates	Transformed coordinates
A	$\boldsymbol{x}' = [0\ 0]$	$\boldsymbol{y}' = [0\ 0]$
B	$\boldsymbol{x}' = [1\ 0]$	$\boldsymbol{y}' = [a_{11}\ a_{21}]$
C	$\boldsymbol{x}' = [1\ 1]$	$\boldsymbol{y}' = [a_{11} + a_{12}\ a_{21} + a_{22}]$
D	$\boldsymbol{x}' = [0\ 1]$	$\boldsymbol{y}' = [a_{12}\ a_{22}]$

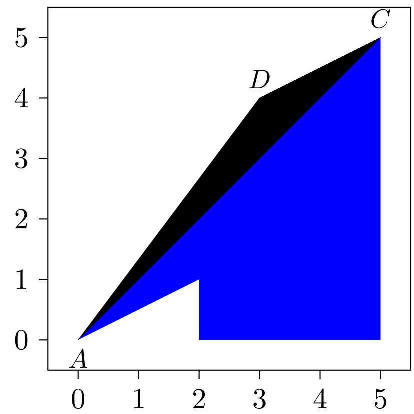
We are now interested in finding the area of the red parallelogram, which we will do as follows:



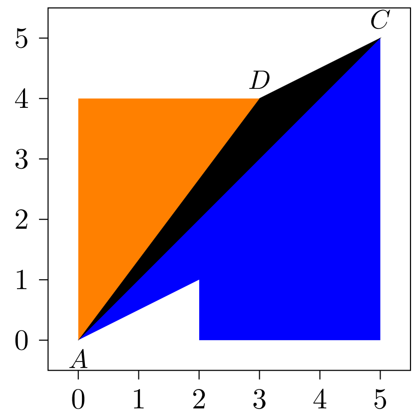
The area of the parallelogram is twice the area of the red triangle



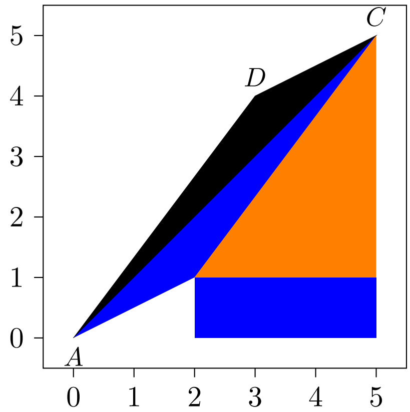
Start with the area of the blue triangle



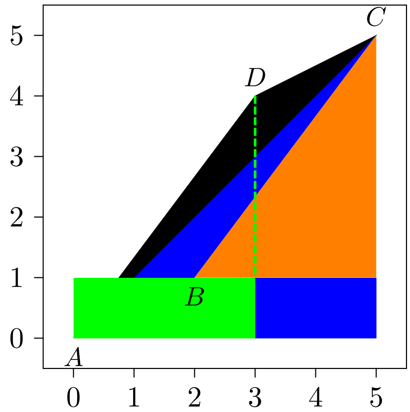
Remove the white triangle



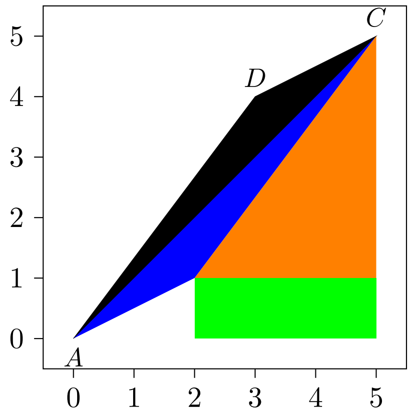
Consider the size of the orange triangle



Rotate and shift the orange triangle as such



Consider the size of the green rectangle



And shift the green rectangle as such

Now we can calculate the area of the original red triangle as a function of the areas of the separate shapes we introduced, each of which can be expressed in terms of the entries of the matrix $\boldsymbol{A}$.

$$\begin{aligned}\text{Red triangle area} &= \text{blue} - \text{white} - \text{orange} - \text{green}\\ &= \frac{1}{2}(a_{11}+a_{12})(a_{21}+a_{22}) - \frac{1}{2}a_{11}a_{21} - \frac{1}{2}a_{12}a_{22} - a_{12}a_{21}\end{aligned}$$

Therefore,

$$\begin{aligned}\text{Parallelogram area} &= 2 \times \text{Red triangle area}\\ &= (a_{11}+a_{12})(a_{21}+a_{22}) - a_{11}a_{21} - a_{12}a_{22} - 2a_{12}a_{21}\\ &= a_{11}a_{21} + a_{12}a_{21} + a_{11}a_{22} + a_{12}a_{22} - a_{11}a_{21} - a_{12}a_{22} - 2a_{12}a_{21}\\ &= a_{11}a_{22} - a_{12}a_{21}\end{aligned}$$

As stated in Equation 3, this area is indeed the determinant of the 2×2 matrix $\boldsymbol{A}$:

$$|\boldsymbol{A}| = \begin{vmatrix} a_{11} & a_{11} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

This result is in fact more general than we have shown: any region having unit area in $\boldsymbol{x}$-space is mapped onto a region in $\boldsymbol{y}$-space having area equal to the magnitude of $|\boldsymbol{A}|$. Therefore, a matrix's determinant is the factor by which an area (hypervolume) is scaled in the linear transformation produced by multiplying vectors with the matrix.

Our example was chosen such that $|\boldsymbol{A}| > 0$. When it is negative, the area of the parallelogram is $-|\boldsymbol{A}|$, indicating a flip/rotation into another quadrant as well. Note that this extends to higher dimensions, i.e. when $n > 2$, in which case the determinant describes the magnitude of increase in volume or hypervolume.

Calculating matrix determinants by hand

For $n > 2$, we define the determinant of $\boldsymbol{A}: n \times n$ by the recursive formula:

$$|\boldsymbol{A}| = \sum_{j=1}^{n} a_{ij}(-1)^{i+j}\left|\boldsymbol{M}_{(ij)}\right|$$

for any arbitrary $i = 1, \ldots, n$, where the $(n-1) \times (n-1)$ matrix $\boldsymbol{M}_{(ij)}$ is obtained from $\boldsymbol{A}$ by deleting row i and column j. The determinant $\left|\boldsymbol{M}_{(ij)}\right|$ is termed the **minor** and $(-1)^{i+j}\left|\boldsymbol{M}_{(ij)}\right|$ the **cofactor** of a_{ij} in the matrix $\boldsymbol{A}$.

Exercise 2.1

Check that you obtain the same formula for $|\boldsymbol{A}|$ when $n = 3$, for each of $i = 1, 2, 3$ respectively.

2.2 Some properties of determinants

1. $|\boldsymbol{A}| = |\boldsymbol{A}'|$
2. Swapping two rows (or two columns) of a matrix $\boldsymbol{A}$ changes the sign of the determinant.
3. The addition of a multiple of one row (or column) of $\boldsymbol{A}$ to another row (or column) leaves the determinant unchanged. This is a useful property, as any matrix $\boldsymbol{A}$ can be converted by a sequence of such operations into an upper or lower triangular matrix, the determinant of which is simple to evaluate, as the following example illustrates:

$$\begin{vmatrix} 1 & -1 & 0 \\ 2 & 1 & 2 \\ 4 & 4 & 9 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 2 & 3 & 2 \\ 4 & 8 & 9 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 2 & 3 & 0 \\ 4 & 8 & \frac{11}{3} \end{vmatrix} = 1 \times 3 \times \frac{11}{3} = 11$$

4. $|\boldsymbol{AB}| = |\boldsymbol{A}||\boldsymbol{B}| = |\boldsymbol{BA}|$
5. If $\boldsymbol{A}$ is orthogonal, then $|\boldsymbol{A}|^2 = |\boldsymbol{A}||\boldsymbol{A}'| = |\boldsymbol{AA}'| = |\boldsymbol{I}| = 1$, such that $|\boldsymbol{A}| = \pm 1$

2.3 Elementary operators

Elementary operators are matrices obtained from making one alteration to the identity matrix. Multiplying a matrix with these elementary operators applies specific changes to the matrix, with the effect differing for pre- and post-multiplication. We will consider three different types of elementary matrices.

$\boldsymbol{P}_{(ij,\lambda)}$

First consider $\boldsymbol{P}_{(ij,\lambda)}$, which is identical to $\boldsymbol{I}$, except that $p_{ij} = \lambda,\ i \neq j$.

For pre-multiplication, $\boldsymbol{P}_{(ij,\lambda)}\boldsymbol{A}$ yields a matrix which is identical to $\boldsymbol{A}$ except that the i^{th} row is replaced by λ times the j^{th} row added to the i^{th} row of $\boldsymbol{A}$. For post-multiplication, $\boldsymbol{A}\boldsymbol{P}_{(ij,\lambda)}$ yields a matrix which is identical to $\boldsymbol{A}$ except that the j^{th} column is replaced by λ times the i^{th} column added to the j^{th} column of $\boldsymbol{A}$.

Note that $\left|\boldsymbol{P}_{(ij,\lambda)}\right| = 1$, such that $\left|\boldsymbol{P}_{(ij,\lambda)}\boldsymbol{A}\right| = \left|\boldsymbol{A}\boldsymbol{P}_{(ij,\lambda)}\right| = |\boldsymbol{A}|$.

$\boldsymbol{E}_{(ij)}$

Next consider $\boldsymbol{E}_{(ij)}$, which is $\boldsymbol{I}$, but with the i^{th} and j^{th} rows interchanged.

Pre- (or post-) multiplication of $\boldsymbol{A}$ by $\boldsymbol{E}_{(ij)}$ has the effect of swopping the i^{th} and j^{th} rows (or columns) of $\boldsymbol{A}$.

Note that $\left|\boldsymbol{E}_{(ij)}\right| = -1$, such that $\left|\boldsymbol{E}_{(ij)}\boldsymbol{A}\right| = \left|\boldsymbol{A}\boldsymbol{E}_{(ij)}\right| = -|\boldsymbol{A}|$.

$\boldsymbol{R}_{(i,\lambda)}$

Finally, consider $\boldsymbol{R}_{(i,\lambda)}$, which is $\boldsymbol{I}$, but with the i^{th} diagonal element replaced by λ.

Now pre- (or post-) multiplication of $\boldsymbol{A}$ by $\boldsymbol{R}_{(i,\lambda)}$ has the effect of multiplying the i^{th} row (or column) by λ.

Note that $\left|\boldsymbol{R}_{(i,\lambda)}\right| = \lambda$, such that $\left|\boldsymbol{R}_{(i,\lambda)}\boldsymbol{A}\right| = \left|\boldsymbol{A}\boldsymbol{R}_{(i,\lambda)}\right| = \lambda|\boldsymbol{A}|$.

Exercise 2.2

Confirm the multiplication effects stated for the elementary operator matrices $\boldsymbol{P}_{(ij,\lambda)}$, $\boldsymbol{E}_{(ij)}$ and $\boldsymbol{R}_{(i,\lambda)}$.

3 Inverses

3.1 Notation

For any square matrix $\boldsymbol{A}$, suppose that there exists another matrix $\boldsymbol{A}^{-1}$ of the same size such that $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$. We term $\boldsymbol{A}^{-1}$ the inverse of $\boldsymbol{A}$. For any linear transformation defined by $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, it follows that $\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$, i.e. $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$, and thus solving linear equations and inverting square matrices are equivalent operations. You should be familiar with this fact and with simple numerical methods for finding the inverse of a matrix.

Exercise 3.1

Show that $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$.

3.2 Evaluation using elementary operators

Suppose we can find a sequence of elementary operator matrices $\boldsymbol{E}_{(1)}, \boldsymbol{E}_{(2)}, \ldots, \boldsymbol{E}_{(K)}$ such that

$$\boldsymbol{E}_{(K)}\boldsymbol{E}_{(K-1)}\cdots\boldsymbol{E}_{(2)}\boldsymbol{E}_{(1)}\boldsymbol{A} = \boldsymbol{I}$$

In other words, if we apply a sequence of elementary row operations to $\boldsymbol{A}$, then the result is the identity matrix. This implies by definition that $\boldsymbol{E}_{(K)}\boldsymbol{E}_{(K-1)}\cdots\boldsymbol{E}_{(2)}\boldsymbol{E}_{(1)} = \boldsymbol{A}^{-1}$. It is not, however, necessary to identify the matrices $\boldsymbol{E}_{(\boldsymbol{k})}$ explicitly, nor to do the relevant matrix multiplications, since clearly (again by definition): $\boldsymbol{E}_{(K)}\boldsymbol{E}_{(K-1)}\cdots\boldsymbol{E}_{(2)}\boldsymbol{E}_{(1)}\boldsymbol{I} = \boldsymbol{A}^{-1}$. Therefore, if we apply the same order of row operations to the identity matrix, this will yield $\boldsymbol{A}^{-1}$.

This is the basis of the method for inverting matrices which is often taught in elementary mathematics courses. Note that we could have done the same thing using only column operations (post-multiplying by the elementary operator matrices), but that it will not work if we mix row and column operations.

Exercise 3.2

Invert the matrix $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix}$ by applying the same sequence of elementary row operations to $\boldsymbol{A}$ and to $\boldsymbol{I}$ in such a way that $\boldsymbol{A}$ is transformed to $\boldsymbol{I}$. Note that this method is not generally the most stable numerical method for obtaining inverses of large matrices in software.

3.3 Adjugate matrix

Another way of determining the inverse (although not much less tedious) is to use the adjugate matrix. The adjugate of $\boldsymbol{A}$, or $adj(\boldsymbol{A})$, is the transpose of the matrix in which each element of $\boldsymbol{A}$ is replaced by its cofactor. We then determine the inverse as

$$\boldsymbol{A}^{-1} = \frac{adj(\boldsymbol{A})}{|\boldsymbol{A}|}$$

Although this may not be a particularly convenient method of determining inverses, it does imply an important property of matrices: The inverse of the matrix $\boldsymbol{A}$ exists if and only if $|\boldsymbol{A}| \neq 0$. If $|\boldsymbol{A}| = 0$, then we say that the matrix is **singular**.

Exercise 3.3

Evaluate the adjugate and determinant of the matrix in Exercise 3.2 and hence determine its inverse.

Exercise 3.4

(a) $[a\boldsymbol{I} + b\boldsymbol{J}]^{-1} = \frac{1}{a}\left[\boldsymbol{I} - \frac{b}{a+nb}\boldsymbol{J}\right]$ where n is the size of $\boldsymbol{I}$ and $\boldsymbol{J}$, and where $a \neq 0$ and $a + nb \neq 0$.

(b) $\left(\boldsymbol{I} + \boldsymbol{A}^{-1}\right)^{-1} = \left[(\boldsymbol{A} + \boldsymbol{I})\,\boldsymbol{A}^{-1}\right]^{-1} = \boldsymbol{A}(\boldsymbol{A} + \boldsymbol{I})^{-1}$

(c) $(\boldsymbol{A} + \boldsymbol{B}\boldsymbol{B}')^{-1}\boldsymbol{B} = \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{I} + \boldsymbol{B}'\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}$

(d) $(\boldsymbol{A}^{-1} + \boldsymbol{B}^{-1})^{-1} = \boldsymbol{A}(\boldsymbol{A} + \boldsymbol{B})^{-1}\boldsymbol{B} = \boldsymbol{B}(\boldsymbol{A} + \boldsymbol{B})^{-1}\boldsymbol{A}$

(e) $\boldsymbol{A} - \boldsymbol{A}(\boldsymbol{A} + \boldsymbol{B})^{-1}\boldsymbol{A} = \boldsymbol{B} - \boldsymbol{B}(\boldsymbol{A} + \boldsymbol{B})^{-1}\boldsymbol{B}$

(f) $\boldsymbol{A}^{-1} + \boldsymbol{B}^{-1} = \boldsymbol{A}^{-1}(\boldsymbol{A} + \boldsymbol{B})\boldsymbol{B}^{-1}$

(g) $(\boldsymbol{I} + \boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{I} - \boldsymbol{A}(\boldsymbol{I} + \boldsymbol{B}\boldsymbol{A})^{-1}$

(h) $(\boldsymbol{I} + \boldsymbol{A}\boldsymbol{B})^{-1}\boldsymbol{A} = \boldsymbol{A}(\boldsymbol{I} + \boldsymbol{B}\boldsymbol{A})^{-1}$

3.4 Completing the square

In the univariate case, completing the square is a technique used to rewrite a quadratic expression $ax^2 + bx + c$ in a more convenient form, where the variable is contained in a squared term and the appropriate constants added. The completed square form can be written as:

$$ax^2 + bx + c = \left(x + \frac{b}{2a}\right)^2 + \left(c - \frac{b^2}{4a}\right)$$

We can now extend this idea to the multivariate case. Let $\boldsymbol{A}\colon p \times p$ be a non-singular matrix, $\boldsymbol{x}$ and $\boldsymbol{b}$ be p-dimensional vectors, and c a scalar. To complete the square, we aim to express the multivariate expression $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} - 2\boldsymbol{b}'\boldsymbol{x} + c$ as a quadratic form plus a term of constants, as follows:

$$(\boldsymbol{x} - \boldsymbol{d})'\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{d}) + e,$$

with appropriate vector $\boldsymbol{d}$ and constant e. We will only state the result, without derivation:

$$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} - 2\boldsymbol{b}'\boldsymbol{x} + c = \left(\boldsymbol{x} - \boldsymbol{A}^{-1}\boldsymbol{b}\right)'\boldsymbol{A}\left(\boldsymbol{x} - \boldsymbol{A}^{-1}\boldsymbol{b}\right) - \boldsymbol{b}'\boldsymbol{A}^{-1}\boldsymbol{b} + c$$

This result can easily be verified by starting with the right-hand side and multiplying out the quadratic form.

$$\begin{aligned}
\left(\boldsymbol{x} - \boldsymbol{A}^{-1}\boldsymbol{b}\right)'\boldsymbol{A}\left(\boldsymbol{x} - \boldsymbol{A}^{-1}\boldsymbol{b}\right) - \boldsymbol{b}'\boldsymbol{A}^{-1}\boldsymbol{b} + c &= \left(\boldsymbol{x}' - \boldsymbol{b}'\boldsymbol{A}^{-1}\right)\boldsymbol{A}\left(\boldsymbol{x} - \boldsymbol{A}^{-1}\boldsymbol{b}\right) - \boldsymbol{b}'\boldsymbol{A}^{-1}\boldsymbol{b} + c \\
&= \left(\boldsymbol{x}'\boldsymbol{A} - \boldsymbol{b}'\right)\left(\boldsymbol{x} - \boldsymbol{A}^{-1}\boldsymbol{b}\right) - \boldsymbol{b}'\boldsymbol{A}^{-1}\boldsymbol{b} + c \\
&= \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}'\boldsymbol{x} - \boldsymbol{x}'\boldsymbol{b} + \boldsymbol{b}'\boldsymbol{A}^{-1}\boldsymbol{b} - \boldsymbol{b}'\boldsymbol{A}^{-1}\boldsymbol{b} + c \\
&= \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} - 2\boldsymbol{b}'\boldsymbol{x} + c
\end{aligned}$$

Note that the choice of using the term $2\boldsymbol{b}'\boldsymbol{x}$ instead of $\boldsymbol{b}'\boldsymbol{x}$ is purely to simplify the completed square form; the vector of constants $\boldsymbol{b}$ can be arbitrarily scaled.

4 The Rank of a Matrix

4.1 Linearly independent vectors

Column vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ are linearly independent if no vector $\boldsymbol{a} \neq \mathbf{0}$ exists such that

$$\sum_{i=1}^{n} a_i \boldsymbol{x}_i = \mathbf{0}.$$

If we define a matrix $\boldsymbol{X}$ with these vectors as columns, i.e. $\boldsymbol{X} = [\boldsymbol{x}_1 \ \boldsymbol{x}_2 \ \cdots \ \boldsymbol{x}_n]$, then the vectors are linearly independent if and only if there is no non-zero solution to the matrix-vector equation $\boldsymbol{X}\boldsymbol{a} = \mathbf{0}$.

We can always find the desired non-zero $\boldsymbol{a}$ if one or more of the $\boldsymbol{x}_i$ are zero vectors, and thus the question of linear independence only arises if $\boldsymbol{x}_i \neq \mathbf{0}$ for all i. In this case, if there exists a solution to $\boldsymbol{X}\boldsymbol{a} = \mathbf{0}$ with $\boldsymbol{a} \neq \mathbf{0}$, then at least two elements of $\boldsymbol{a}$ must be non-zero. Now suppose that $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ are linearly dependent, and that $a_n \neq 0$ (the latter assumption involving no loss of generality). Then we can write:

$$\boldsymbol{x}_n = -\frac{a_1}{a_n}\boldsymbol{x}_1 - \frac{a_2}{a_n}\boldsymbol{x}_2 - \cdots - \frac{a_{n-1}}{a_n}\boldsymbol{x}_{n-1}$$

with at least one of the coefficients on the RHS being non-zero, i.e. $\boldsymbol{x}_n$ is a linear combination of the other vectors and can be substituted out in any formula involving it.

If the remaining vectors are not linearly independent then we can repeat the process for the next vector in the set (say $\boldsymbol{x}_{n-1}$) without loss of generality. Eventually we can reach a point where r vectors, say $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r$, remain which are linearly independent, and all the other vectors are linear combinations of these. In effect, we have partitioned the matrix $\boldsymbol{X}$ into $\begin{bmatrix}\boldsymbol{X}_1 \colon m \times r & \boldsymbol{X}_2 \colon m \times (n-r)\end{bmatrix}$, such that each column of $\boldsymbol{X}_2$ is a linear combination of the columns forming $\boldsymbol{X}_1$. Thus we can express $\boldsymbol{X}_2$ in the form $\boldsymbol{X}_2 = \boldsymbol{X}_1\boldsymbol{B}$.

Example

Consider the matrix

$$\boldsymbol{X} = [\boldsymbol{x}_1 \ \boldsymbol{x}_2 \ \boldsymbol{x}_3 \ \boldsymbol{x}_4] = \begin{bmatrix} 1 & 3 & 2 & 1 \\ 4 & 2 & 3 & 9 \\ 1 & -1 & 0 & 3 \end{bmatrix}$$

It is easily confirmed that $3\boldsymbol{x}_1 - \boldsymbol{x}_3 - \boldsymbol{x}_4 = \mathbf{0}$, implying that the vectors are linearly dependent, with $\boldsymbol{a} = \begin{bmatrix}3 & 0 & -1 & -1\end{bmatrix}'$. Thus $\boldsymbol{x}_4 = 3\boldsymbol{x}_1 - \boldsymbol{x}_3$. For the remaining three vectors, we can confirm that $\boldsymbol{x}_1 + \boldsymbol{x}_2 - 2\boldsymbol{x}_3 = \mathbf{0}$, and thus these are again linearly dependent. Since the ordering of the labelling of the vectors is arbitrary, let us solve for $\boldsymbol{x}_1 = -\boldsymbol{x}_2 + 2\boldsymbol{x}_3$. Then, of course, $\boldsymbol{x}_4 = -3\boldsymbol{x}_2 + 5\boldsymbol{x}_3$. We can easily verify that the vectors $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$ are linearly independent.

Now the matrix $\boldsymbol{X}$ can be written as $\begin{bmatrix}\boldsymbol{X}_1 \colon 3 \times 2 & \boldsymbol{X}_2 \colon 3 \times 2\end{bmatrix}$ where $\boldsymbol{X}_1 = \begin{bmatrix}\boldsymbol{x}_2 & \boldsymbol{x}_3\end{bmatrix}$ and $\boldsymbol{X}_2 = \begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_4\end{bmatrix} = \boldsymbol{X}_1\boldsymbol{B}$, with $\boldsymbol{B} = \begin{bmatrix} -1 & -3 \\ 2 & 5 \end{bmatrix}$.

4.2 Matrix rank

If the columns of a square matrix $\boldsymbol{X}$ are linearly dependent, then it is possible to add multiples of the other columns to any one chosen column of $\boldsymbol{X}$ in such a way that this column consists entirely of 0's. The determinant of this resulting matrix is 0. However, as we saw previously such operations do not change the determinant of a matrix, and thus we conclude that $|\boldsymbol{X}| = 0$. This implies that $\boldsymbol{X}^{-1}$ does not exist, and there is no solution for $\boldsymbol{a}$ in the equation $\boldsymbol{X}\boldsymbol{a} = \boldsymbol{b}$ for any $\boldsymbol{b} \neq \boldsymbol{0}$. On the other hand, from the definition of linear dependence, there does exist a vector $\boldsymbol{a}$ such that $\boldsymbol{X}\boldsymbol{a} = \boldsymbol{0}$. This is in fact a crucial general property of square matrices:

There is either a non-trivial solution to $\boldsymbol{X}\boldsymbol{a} = \boldsymbol{0}$, or a solution to $\boldsymbol{X}\boldsymbol{a} = \boldsymbol{b} \neq \boldsymbol{0}$, but not both.

We also note the following important facts regarding any arbitrary $m \times n$ matrix $\boldsymbol{X}$:

1. The maximum number of columns of $\boldsymbol{X}$ that can be linearly independent is m (the number of rows of $\boldsymbol{X}$).
2. The maximum number of rows of $\boldsymbol{X}$ that can be linearly independent is n (the number of columns of $\boldsymbol{X}$).
3. The number of linearly independent rows of $\boldsymbol{X}$ = the number of linearly independent columns of $\boldsymbol{X}$.

Defining matrix rank

The rank of a matrix $\boldsymbol{X}$, written $rank(\boldsymbol{X})$, is the number of linearly independent columns (or rows) of $\boldsymbol{X}$.

For an $m \times n$ matrix $\boldsymbol{X}$, it is clear that $rank(\boldsymbol{X}) \leq \min(m, n)$. We say that $\boldsymbol{X}$ is of full row rank if $rank(\boldsymbol{X}) = m$, and of full column rank if $rank(\boldsymbol{X}) = n$. If a square matrix is of full row rank, then it is also of full column rank (and vice versa), and $\boldsymbol{X}$ is said to be of full rank.

Note that in the previous example, $rank(\boldsymbol{X}) = rank(\boldsymbol{X}_1) = 2$, and that $\boldsymbol{X}_1$ is of full column rank. Note also that a matrix $\boldsymbol{X}$ is of full rank if and only if it is invertible.

4.3 Factorisation of a matrix into matrices of full column and row ranks

It is often useful to factorize a matrix into matrices of full column and row ranks. Let the $p \times q$ matrix $\boldsymbol{A}$ have rank $r < \min(p, q)$, and suppose that the rows and columns of $\boldsymbol{A}$ have been ordered such the first r of each are linearly independent. We can therefore write $\boldsymbol{A}$ in the partitioned form:

$$ {}_p\boldsymbol{A}_q = \begin{bmatrix} {}_r\boldsymbol{X}_r & {}_r\boldsymbol{Y}_{q-r} \\ {}_{p-r}\boldsymbol{Z}_r & {}_{p-r}\boldsymbol{W}_{q-r} \end{bmatrix} $$

where $\boldsymbol{X}$ is of full rank.

For appropriate matrices $\boldsymbol{F}$ and $\boldsymbol{H}$, representing the relationships between the linearly dependent and independent rows and columns of $\boldsymbol{A}$ respectively, we can write the last $p - r$ rows of $\boldsymbol{A}$ as

$$ \begin{bmatrix} \boldsymbol{Z} & \boldsymbol{W} \end{bmatrix} = \boldsymbol{F} \begin{bmatrix} \boldsymbol{X} & \boldsymbol{Y} \end{bmatrix} $$

and the last $q - r$ columns of $\boldsymbol{A}$ as

$$\begin{bmatrix} \boldsymbol{Y} \\ \boldsymbol{W} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Z} \end{bmatrix} \boldsymbol{H}.$$

From this we see that $\boldsymbol{Z} = \boldsymbol{FX}$; $\boldsymbol{Y} = \boldsymbol{XH}$; and $\boldsymbol{W} = \boldsymbol{FY} = \boldsymbol{ZH} = \boldsymbol{FXH}$. Also note that, since $\boldsymbol{X}$ is of full rank, we can derive $\boldsymbol{F} = \boldsymbol{ZX}^{-1}$ and $\boldsymbol{H} = \boldsymbol{X}^{-1}\boldsymbol{Y}$, providing a manner of obtaining $\boldsymbol{F}$ and $\boldsymbol{H}$. Now the matrix $\boldsymbol{A}$ can be factorised as:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{X} & \boldsymbol{XH} \\ \boldsymbol{FX} & \boldsymbol{FXH} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{F} \end{bmatrix} \boldsymbol{X} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{H} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{F} \end{bmatrix} \begin{bmatrix} \boldsymbol{X} & \boldsymbol{XH} \end{bmatrix}.$$

This shows that any matrix of less than full rank can be factorised into the product of two matrices of full column rank and full row rank, respectively.

4.4 Canonical forms

From the above factorisation we can see that the matrix $\boldsymbol{A}$ can be transformed by elementary operations into a matrix consisting only of 0's, except that the first r diagonal elements are 1's. Recalling that elementary operations are equivalent to pre- or post-multiplication by the relevant operator matrices, we thus see that it is always possible to find matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$ such that:

$$\boldsymbol{PAQ} = \begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} = \boldsymbol{C}.$$

The matrix $\boldsymbol{C}$ is termed the equivalent canonical form of $\boldsymbol{A}$. If two matrices reduce to the same canonical form, then they are said to be equivalent. Note that $\boldsymbol{P}$ and $\boldsymbol{Q}$ are not unique in general, but that they are invertible (the elementary operations can always be "undone"). We can therefore also express the matrix as a function of its canonical form:

$$\boldsymbol{A} = \boldsymbol{P}^{-1}\boldsymbol{C}\boldsymbol{Q}^{-1}.$$

If $\boldsymbol{A}$ itself is non-singular, then it is of full rank and $\boldsymbol{C} = \boldsymbol{I}$, which gives a factorisation of $\boldsymbol{A} = \boldsymbol{P}^{-1}\boldsymbol{Q}^{-1}$.

For a symmetric matrix $\boldsymbol{A}$ it is possible to find a matrix $\boldsymbol{P}$ such that $\boldsymbol{PAP'} = \boldsymbol{C}$, i.e. to reach the canonical form by performing equivalent operations on rows and columns. If $\boldsymbol{A}$ is also non-singular, this implies that $\boldsymbol{A} = \boldsymbol{KK'}$ where $\boldsymbol{K} = \boldsymbol{P}^{-1}$.

Exercise 4.1

Verify the following useful rank theorems:

(a) $rank(\boldsymbol{AB}) \leq min\{rank(\boldsymbol{A}), rank(\boldsymbol{B})\}$

(b) If $\boldsymbol{A}$ is non-singular, then $rank(\boldsymbol{AB}) = rank(\boldsymbol{B})$

(c) If $\boldsymbol{AGA} = \boldsymbol{A}$, then $rank(\boldsymbol{GA}) = rank(\boldsymbol{A})$

(d) $rank(\boldsymbol{A} + \boldsymbol{B}) \leq rank([\boldsymbol{A}\ \boldsymbol{B}]) \leq rank(\boldsymbol{A}) + rank(\boldsymbol{B})$

4.5 Gram-Schmidt process

Sometimes it is useful or necessary to find orthogonal vectors in some vector space. Given a set of n linearly independent vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$, the Gram-Schmidt process provides an algorithm for forming a set of n orthogonal vectors $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n$ spanning the same vector space. Therefore, any vector $\boldsymbol{v} = a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_n\boldsymbol{x}_n$ can be written as $\boldsymbol{v} = b_1\boldsymbol{q}_1 + b_2\boldsymbol{q}_2 + \cdots + b_n\boldsymbol{q}_n$. Geometrically, this can be seen as the projection of the vectors onto an orthonormal basis.

The Gram-Schmidt process is as follows:

Step 1

$$\boldsymbol{q}_1 = \boldsymbol{x}_1$$

Step 2

$$\boldsymbol{q}_2 = \boldsymbol{x}_2 - \left(\frac{\boldsymbol{q}_1'\boldsymbol{x}_2}{\boldsymbol{q}_1'\boldsymbol{q}_1}\right)\boldsymbol{q}_1$$

Step 3

$$\boldsymbol{q}_3 = \boldsymbol{x}_3 - \left(\frac{\boldsymbol{q}_1'\boldsymbol{x}_3}{\boldsymbol{q}_1'\boldsymbol{q}_1}\right)\boldsymbol{q}_1 - \left(\frac{\boldsymbol{q}_2'\boldsymbol{x}_3}{\boldsymbol{q}_2'\boldsymbol{q}_2}\right)\boldsymbol{q}_2$$

$\vdots$

Step n

$$\boldsymbol{q}_n = \boldsymbol{x}_n - \left(\frac{\boldsymbol{q}_1'\boldsymbol{x}_n}{\boldsymbol{q}_1'\boldsymbol{q}_1}\right)\boldsymbol{q}_1 - \left(\frac{\boldsymbol{q}_2'\boldsymbol{x}_n}{\boldsymbol{q}_2'\boldsymbol{q}_2}\right)\boldsymbol{q}_2 - \left(\frac{\boldsymbol{q}_{n-1}'\boldsymbol{x}_n}{\boldsymbol{q}_{n-1}'\boldsymbol{q}_{n-1}}\right)\boldsymbol{q}_{n-1}$$

Exercise 4.2

(a) Compute $\boldsymbol{q}_1'\boldsymbol{q}_2$ above to show that they are orthogonal vectors.

(b) Consider the two vectors $\boldsymbol{x}_1 = \begin{bmatrix}1 & 0 & 2\end{bmatrix}'$ and $\boldsymbol{x}_2 = \begin{bmatrix}1 & -1 & 1\end{bmatrix}'$. Find a set of two orthogonal vectors spanning the same vector space as $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.

(c) Find a set of two orthonormal vectors spanning the same vector space as $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in (b).

5 Generalised inverses and linear equations

Suppose that the $m \times n$ matrix $\boldsymbol{A}$ is not invertible. It may be a square but singular matrix, or a non-square, i.e. rectangular matrix. Can we still say something about solutions to $\boldsymbol{Ax} = \boldsymbol{b}$, for fixed $\boldsymbol{b}$, where $\boldsymbol{x}$ and $\boldsymbol{b}$ are n- and m-vectors respectively?

5.1 Solutions to different types of sets of linear equations

The following cases are of interest:

1) $m < n$ and $rank(\boldsymbol{A}) = m$, i.e. $\boldsymbol{A}$ is of full row rank

We have m linear equations, so we will aim to solve for m elements of $\boldsymbol{x}$ by arbitrarily setting $n-m$ of its elements to 0. Specifically, suppose that we set the last $n-m$ elements to 0 (there is no loss of generality here). Now partition $\boldsymbol{A}$ as follows:

$$\boldsymbol{A} = [{}_m\boldsymbol{B}_m \quad {}_m\boldsymbol{N}_{n-m}]$$

where $\boldsymbol{B}$ forms the so-called basis for $\boldsymbol{A}$ and the remaining columns non-basic.

We can then write $\boldsymbol{Ax} = \boldsymbol{Bx}_B + \boldsymbol{Nx}_N$, where $\boldsymbol{x}_B$ and $\boldsymbol{x}_N$ represent the corresponding partitioning of $\boldsymbol{x}$. To find a solution, we can set $\boldsymbol{x}_N = \boldsymbol{0}$ and then solve $\boldsymbol{Bx}_B = \boldsymbol{b}$. Since $\boldsymbol{B}$ is a full-rank square matrix it can be inverted, yielding the solution $\boldsymbol{x}_B = \boldsymbol{B}^{-1}\boldsymbol{b}$.

This can succinctly be expressed as $\boldsymbol{x} = \boldsymbol{Gb}$, where $\boldsymbol{G}$ is the $n \times m$ matrix of which the first m rows constitute $\boldsymbol{B}^{-1}$ and the remaining rows are all 0's. Such a solution is called a basic solution, but it is, of course, not unique. We could set any $n-m$ elements of $\boldsymbol{x}$ to any value and solve for the remaining elements.

Exercise 5.1

Show that $\boldsymbol{AGA} = \boldsymbol{A}$ and that $\boldsymbol{GAG} = \boldsymbol{G}$ where $rank(\boldsymbol{A}) = m$.

2) $m > n$ and $rank(\boldsymbol{A}) = n$, i.e. $\boldsymbol{A}$ is of full column rank

Since we have more linear equations (m) than variables (n) to solve for, there will be no general solution unless some redundancy exists. We could, however, find a "closest" approximation to a solution in a least squares sense, by finding the vector $\boldsymbol{x}$ that minimises $(\boldsymbol{b}-\boldsymbol{Ax})'(\boldsymbol{b}-\boldsymbol{Ax})$. Setting the derivative of this expression with respect to $\boldsymbol{x}$ (i.e. with respect to each element of $\boldsymbol{x}$ in turn and expressing the results in vector form) equal to zero gives the condition:

$$2\boldsymbol{A}'(\boldsymbol{b} - \boldsymbol{Ax}) = \boldsymbol{0}.$$

Solving this equation for $\boldsymbol{x}$ yields

$$\begin{aligned}
\boldsymbol{A}'\boldsymbol{b} - \boldsymbol{A}'\boldsymbol{Ax} &= \boldsymbol{0} \\
\boldsymbol{A}'\boldsymbol{Ax} &= \boldsymbol{A}'\boldsymbol{b} \\
\boldsymbol{x} &= (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{b} \\
&= \boldsymbol{Gb}
\end{aligned}$$

Note that it can be shown that the $n \times n$ matrix $\boldsymbol{A}'\boldsymbol{A}$ has rank n and is therefore invertible. Therefore, the least squares solution is again of the form $\boldsymbol{G}\boldsymbol{b}$, where $\boldsymbol{G} = (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$ in this case.

Exercise 5.2

Show that $\boldsymbol{AGA} = \boldsymbol{A}$ and that $\boldsymbol{GAG} = \boldsymbol{G}$ where $rank(\boldsymbol{A}) = n$.

3) $m = n$, i.e. $\boldsymbol{A}$ is square, but $rank(\boldsymbol{A}) < m$

We know that a solution to $\boldsymbol{Ax} = \boldsymbol{b}$ exists for $\boldsymbol{b} \neq \boldsymbol{0}$ if $\boldsymbol{A}$ is of full rank, i.e. $rank(\boldsymbol{A}) = m = n$, namely $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$. When $rank(\boldsymbol{A}) < m$, a solution will only exist if the linear relationships between the rows of $\boldsymbol{A}$ are mirrored by equivalent relationships between the elements of $\boldsymbol{b}$. This is illustrated in the following example:

Let

$$\boldsymbol{Ax} = \begin{bmatrix} 2 & 3 & 1 \\ 1 & 1 & 2 \\ 3 & 5 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 14 \\ 6 \\ 22 \end{bmatrix}$$

Since $\boldsymbol{a}_{(3)} = 2\boldsymbol{a}_{(1)} - \boldsymbol{a}_{(2)}$, we know that $rank(\boldsymbol{A}) < 3$. However, on the right-hand side we also have $22 = 2(14) - 6$, mimicking the row relationships. Now we effectively have 2 equations and 3 unknowns, so solutions can certainly be found.

In general, if $rank([\boldsymbol{A} \quad \boldsymbol{b}]) = rank(\boldsymbol{A})$, then the aforementioned mirrored relationship is present. This ensures the existence of a solution to the equation $\boldsymbol{Ax} = \boldsymbol{b}$, which is then said to be consistent.

5.2 Generalised inverse definitions

Any $m \times n$ matrix $\boldsymbol{G}$ related to $\boldsymbol{A}$ through $\boldsymbol{AGA} = \boldsymbol{A}$ and/or $\boldsymbol{GAG} = \boldsymbol{G}$ plays an important role in finding solutions to $\boldsymbol{Ax} = \boldsymbol{b}$. Note, incidentally, that if $\boldsymbol{A}$ is non-singular, then $\boldsymbol{G} = \boldsymbol{A}^{-1}$ satisfies these properties. The following definitions are useful here:

Generalised inverse

If $\boldsymbol{AGA} = \boldsymbol{A}$ then $\boldsymbol{G}$ is a generalised inverse of $\boldsymbol{A}$, often written as $\boldsymbol{A}^-$.

Reflexive generalised inverse

If $\boldsymbol{AGA} = \boldsymbol{A}$ and $\boldsymbol{GAG} = \boldsymbol{G}$, then $\boldsymbol{G}$ is a generalised inverse of $\boldsymbol{A}$ and $\boldsymbol{A}$ is a generalised inverse of $\boldsymbol{G}$. We then say that $\boldsymbol{G}$ is a reflexive generalised inverse of $\boldsymbol{A}$.

Pseudo inverse (Moore-Penrose inverse)

If $\boldsymbol{G}$ is a reflexive generalized inverse of $\boldsymbol{A}$ and if $\boldsymbol{AG}$ and $\boldsymbol{GA}$ are symmetric, then $\boldsymbol{G}$ is the Moore-Penrose inverse of $\boldsymbol{A}$, also termed the pseudo inverse. This matrix is unique, but that is not true in general for generalised, or reflexive generalised inverses.

5.3 Constructing a generalised inverse

One method of constructing a generalized inverse starts by reducing $\boldsymbol{A}$ to the form:

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{D}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$$

where $r = rank(\boldsymbol{A})$ such that $\boldsymbol{D}_r$ is of full rank. This is done in the same way as when reducing a matrix to canonical form – i.e. by the application of elementary operations to the rows and columns of $\boldsymbol{A}$ – expect that $\boldsymbol{D}_r$ does not necessarily equal $\boldsymbol{I}_r$.

As we have previously seen, this can be expressed as $\boldsymbol{R} = \boldsymbol{PAQ}$, where $\boldsymbol{P}$ and $\boldsymbol{Q}$ are non-singular matrices constructed as the products of the relevant elementary operators. Then

$$\boldsymbol{G} = \boldsymbol{Q} \begin{bmatrix} \boldsymbol{D}_r^{-1} & \boldsymbol{X} \\ \boldsymbol{Y} & \boldsymbol{Z} \end{bmatrix} \boldsymbol{P}$$

is a generalised inverse of $\boldsymbol{A}$ for any arbitrary matrices (of the required sizes) $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$.

Exercise 5.3

a) Demonstrate that $\boldsymbol{G}$ as defined above is a generalised inverse of $\boldsymbol{A}$, i.e. $\boldsymbol{AGA} = \boldsymbol{A}$, by invoking the relationship $\boldsymbol{A} = \boldsymbol{P}^{-1}\boldsymbol{R}\boldsymbol{Q}^{-1}$.

b) Also show that $\boldsymbol{G}$ is a reflexive generalised inverse if and only if $\boldsymbol{Z} = \boldsymbol{Y}\boldsymbol{D}_r\boldsymbol{X}$.

A special case

There is one special case in which the generalised inverse is simplified, negating the need to determine the products of the elementary operators, $\boldsymbol{P}$ and $\boldsymbol{Q}$.

Let

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{B} & \boldsymbol{S} \\ \boldsymbol{T} & \boldsymbol{U} \end{bmatrix}$$

where $\boldsymbol{B}: r \times r$ is the non-singular sub-matrix of $\boldsymbol{A}$. If $\boldsymbol{U} = \boldsymbol{T}\boldsymbol{B}^{-1}\boldsymbol{S}$, then

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{B}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$$

is a generalised inverse of $\boldsymbol{A}$.

5.4 Solutions to Ax = b

Let us first briefly consider the case where $\boldsymbol{b} = \boldsymbol{0}$, i.e. the equation set given by $\boldsymbol{Ax} = \boldsymbol{0}$. Now, we know that $\boldsymbol{x} = \boldsymbol{0}$ is always a solution; the question is whether other (general) solutions exist. If $\boldsymbol{A}$ is non-singular, then the trivial solution is unique. In general, however, let us consider solutions of the form

$$\boldsymbol{x} = (\boldsymbol{GA} - \boldsymbol{I})\boldsymbol{z}$$

for any arbitrary vector $\boldsymbol{z}$, where $\boldsymbol{G}$ is a generalised inverse of $\boldsymbol{A}$. These are solutions to $\boldsymbol{Ax} = \boldsymbol{0}$, since:

$$\boldsymbol{Ax} = \boldsymbol{A}(\boldsymbol{GA} - \boldsymbol{I})\boldsymbol{z} = (\boldsymbol{AGA} - \boldsymbol{A})\boldsymbol{z} = \boldsymbol{0}$$

as required. If $\boldsymbol{A}$ is of less than full column rank, then this solution will in general be non-trivial.

Now let us return to $\boldsymbol{Ax} = \boldsymbol{b} \neq \boldsymbol{0}$, which we suppose is consistent. For any arbitrary matrix $\boldsymbol{G}$ of the appropriate size, multiplying by $\boldsymbol{AG}$ on both sides yields

$$\boldsymbol{AGAx} = \boldsymbol{AGb}.$$

If $\boldsymbol{G}$ is a generalised inverse of $\boldsymbol{A}$, then this becomes

$$\begin{aligned} \boldsymbol{Ax} &= \boldsymbol{AGb} \\ \boldsymbol{b} &= \boldsymbol{A}(\boldsymbol{Gb}) \end{aligned}$$

Thus if the equations are consistent, then a solution is given by $\boldsymbol{x} = \boldsymbol{Gb}$.

In fact, since $\boldsymbol{x} = (\boldsymbol{GA} - \boldsymbol{I})\boldsymbol{z}$ is a solution to $\boldsymbol{Ax} = \boldsymbol{0}$ for any arbitrary vector $\boldsymbol{z}$, and $\boldsymbol{G}$ is a generalised inverse of $\boldsymbol{A}$, it is easy to see that $\boldsymbol{Gb} + (\boldsymbol{GA} - \boldsymbol{I})\boldsymbol{z}$ is also a solution. Moreover, this characterises all solutions which can exist. The solution is unique if and only if $\boldsymbol{GA} = \boldsymbol{I}$.

Example

Consider again the earlier set of linear equations, where we have

$$\boldsymbol{Ax} = \begin{bmatrix} 2 & 3 & 1 \\ 1 & 1 & 2 \\ 3 & 5 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 14 \\ 6 \\ 22 \end{bmatrix}$$

We already know that some solution exists. To quickly find a (general) solution, we can note that this is an example of the special case mentioned above (showing this is left as an exercise), where $\boldsymbol{B}: 2 \times 2 = \begin{bmatrix} 2 & 3 \\ 1 & 1 \end{bmatrix}$. Therefore,

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{B}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} = \begin{bmatrix} -1 & 3 & 0 \\ 1 & -2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

A (non-unique) solution to $\boldsymbol{Ax} = \boldsymbol{b}$ is therefore given by $\boldsymbol{x} = \boldsymbol{Gb} = \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix}$.

Calculating the matrix $(\boldsymbol{GA} - \boldsymbol{I})$ yields $\begin{bmatrix} 0 & 0 & 5 \\ 0 & 0 & -3 \\ 0 & 0 & -1 \end{bmatrix}$, and by multiplying this with an arbitrary 3×1 vector $\boldsymbol{z}$, we obtain

$$\boldsymbol{x} = \begin{bmatrix} 5z_3 \\ -3z_3 \\ -z_3 \end{bmatrix}.$$

Any $\boldsymbol{x}$ of this form is a solution to $\boldsymbol{Ax} = \boldsymbol{0}$. We can also use this expression to determine the linear dependence structure in the columns of $\boldsymbol{A}$.

Finally, combining these results, we can define $\boldsymbol{x} = \boldsymbol{Gb} + (\boldsymbol{GA} - \boldsymbol{I})\boldsymbol{z} = \begin{bmatrix} 4 + 5z_3 \\ 2 - 3z_3 \\ -z_3 \end{bmatrix}$ as the exhaustive set of solutions to $\boldsymbol{Ax} = \boldsymbol{b}$. There are infinitely many such solutions, but there is only one degree of freedom: fixing any one element of $\boldsymbol{x}$ implies values for the other two elements.

6 Eigenvalues and eigenvectors

6.1 Definition

In Chapter 2 we saw that multiplying a matrix by a vector results in a linear transformation that (usually) changes the vector, such that it does not lie its span any more, i.e. it points in a different direction after the transformation. The question now arises: can we find vectors such that, for some matrix multiplication, the vector changes in such a way that it still lies on the same line? In other words, after this transformation it will either still point in the same direction or in the exact opposite direction. We can view this by saying that the effect of the matrix multiplication is that the vector is only multiplied by some constant.

Therefore, consider some $n \times n$ matrix $\boldsymbol{A}$ and a non-zero vector $\boldsymbol{x}$, and define the relationship

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x} \tag{4}$$

for some real number λ. If a vector $\boldsymbol{x}$ and scalar λ satisfies this equation, then we refer to them respectively as an eigenvector and corresponding eigenvalue of $\boldsymbol{A}$. Note that the scaling of $\boldsymbol{x}$ is arbitrary, since if the above is true, then $\boldsymbol{A}(k\boldsymbol{x}) = \lambda(k\boldsymbol{x})$. It is therefore common to express the eigenvector in normalised form.

6.2 Evaluating eigenvalues and eigenvectors

In order to find an eigenvector-eigenvalue pair, we need to solve Equation 4, which we can do by rewriting it as follows:

$$\begin{aligned}
\boldsymbol{A}\boldsymbol{x} &= \lambda \boldsymbol{x} \\
\boldsymbol{A}\boldsymbol{x} - \lambda \boldsymbol{x} &= \boldsymbol{0} \\
\boldsymbol{A}\boldsymbol{x} - \lambda \boldsymbol{I}\boldsymbol{x} &= \boldsymbol{0} \\
(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{x} &= \boldsymbol{0}
\end{aligned}$$

From Chapter 4 we know that a non-trivial solution will only exist if $(\boldsymbol{A} - \lambda \boldsymbol{I})$ is singular. Therefore, to find eigenvalues we can now determine for which values of λ the matrix on the left-hand side has a determinant of zero:

$$|\boldsymbol{A} - \lambda \boldsymbol{I}| = 0 \tag{5}$$

This equation, known as the characteristic equation, defines an n^{th}-order polynomial in λ, which we can use to solve for n values of λ (although, note that some values may repeat). For each solution we can then solve for $\boldsymbol{x}$ using Equation 4.

Example

Consider the following matrix, with $n = 3$:

$$\begin{bmatrix} -1 & -2 & -2 \\ 1 & 2 & 1 \\ -1 & -1 & 0 \end{bmatrix}$$

The characteristic equation simplifies to the following.

$$
\begin{aligned}
\begin{vmatrix} -1-\lambda & -2 & -2 \\ 1 & 2-\lambda & 1 \\ -1 & -1 & -\lambda \end{vmatrix} &= (-1-\lambda)\begin{vmatrix} 2-\lambda & 1 \\ -1 & -\lambda \end{vmatrix} - \begin{vmatrix} -2 & -2 \\ -1 & -\lambda \end{vmatrix} - \begin{vmatrix} -2 & -2 \\ 2-\lambda & 1 \end{vmatrix} \\
&= (-1-\lambda)(\lambda^2 - 2\lambda + 1) - (2\lambda - 2) - (-2 + 4 - 2\lambda) \\
&= -(\lambda+1)(\lambda-1)^2 = 0
\end{aligned}
$$

The eigenvalues are thus $\lambda = 1$ with a multiplicity of 2 and $\lambda = -1$.

For $\lambda = 1$, we have to find elements of $\boldsymbol{x}$ satisfying $\boldsymbol{A}\boldsymbol{x} = 1\boldsymbol{x}$:

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

These equations are equivalent to $(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}$, and since $\boldsymbol{A} - \lambda\boldsymbol{I}$ is by construction singular, there is at least one redundancy in this set of equations. In fact all three equations reduce to $x_1 + x_2 + x_3 = 0$, hence we effectively have one equation with three unknowns. One solution, chosen with $x_1 = 0$, yields an eigenvector $\boldsymbol{u}_1 = \begin{bmatrix} 0 & 1 & -1 \end{bmatrix}'$, where the scaling in $x_2 = -x_3$ is arbitrary. It is possible to find a further vector of the required form orthogonal to $\boldsymbol{u}_1$, as for example $\boldsymbol{u}_1 = \begin{bmatrix} -2 & 1 & 1 \end{bmatrix}'$. These two eigenvectors are linearly independent, but it can be seen that any further eigenvectors corresponding to $\lambda = 1$ would be linearly dependent on these two, which thus characterise the set of all eigenvectors corresponding to an eigenvalue of 1.

For $\lambda = -1$ we have

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

It is clear that for $x_1 = v$ the corresponding eigenvector must be of the form $\boldsymbol{u}_2 = \begin{bmatrix} v & -\frac{v}{2} & \frac{v}{2} \end{bmatrix}'$, where v is purely an arbitrary scaling – we could for example use $v = 2$ to give $\begin{bmatrix} 2 & -1 & 1 \end{bmatrix}'$ as a corresponding eigenvector.

In this example we found two *orthogonal* eigenvectors corresponding to the eigenvalue with multiplicity of 2. Although this need not generally be true, if unique eigenvectors exist they must be linearly independent (these form a subspace called the eigenspace corresponding to that eigenvalue). However, orthogonality is a stronger condition than linear independence, and therefore creating orthogonal vectors ensures linear independence.

Note that the polynomial solution for λ from the characteristic equation quickly becomes impracticable for larger n, and other algorithms for computing eigenvalues and eigenvectors have been developed and are used in various software and packages.

Exercise 6.1

Find the eigenvalues and normalised eigenvectors of the matrix $\begin{bmatrix} 0 & 1 & 2 \\ 2 & 3 & 0 \\ 0 & 4 & 5 \end{bmatrix}$.

Note that this matrix also has an eigenvalue of 1 with multiplicity of 2, but there corresponds the unique eigenvector (apart from scaling constants) $\begin{bmatrix}1 & -1 & 1\end{bmatrix}'$.

6.3 Properties

Property 1

If $\boldsymbol{Ax} = \lambda\boldsymbol{x}$, then $\boldsymbol{A}^2\boldsymbol{x} = \boldsymbol{AAx} = \boldsymbol{A}(\lambda\boldsymbol{x}) = \lambda\boldsymbol{Ax} = \lambda^2\boldsymbol{x}$, i.e. if $(\lambda, \boldsymbol{x})$ is an eigenvalue-eigenvector pair for $\boldsymbol{A}$, then $(\lambda^2, \boldsymbol{x})$ is an eigenvalue-eigenvector pair for $\boldsymbol{A}^2$. This can be expanded such that, in general, $\boldsymbol{A}^k\boldsymbol{x} = \lambda^k\boldsymbol{x}$

Property 2

If $\boldsymbol{A}$ is non-singular, all the eigenvalues will be non-zero. In this case, we can also extend the previous property such that $\left(\frac{1}{\lambda}, \boldsymbol{x}\right)$ is an eigenvalue-eigenvector pair for $\boldsymbol{A}^{-1}$.

$\lambda = 0$ can be an eigenvalue of $\boldsymbol{A}$, but this implies that $|\boldsymbol{A}| = 0$.

Property 3

The rank of $\boldsymbol{A}$ is the number of non-zero eigenvalues of $\boldsymbol{A}$ (counting repeated roots as many times as they occur).

Property 4

The sum of a matrix's eigenvalues equals its trace: $\displaystyle\sum_{i=1}^{n} \lambda_i = tr(\boldsymbol{A})$.

The product of a matrix's eigenvalues equals its determinant: $\displaystyle\prod_{i=1}^{n} \lambda_i = |\boldsymbol{A}|$.

Property 5

For each of the n eigenvalues of $\boldsymbol{A}$, $\lambda_1, \lambda_2, \dots, \lambda_n$ (some of which may be repeated), there are corresponding eigenvectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n$. Note that for repeated eigenvalues, we may or may not have repeated eigenvectors. By constructing the $n \times n$ matrix $\boldsymbol{U}$ having the eigenvectors as columns, we have that:

$$\boldsymbol{AU} = \boldsymbol{A}[\boldsymbol{u}_1 \quad \cdots \quad \boldsymbol{u}_n] = [\lambda_1\boldsymbol{u}_1 \quad \cdots \quad \lambda_n\boldsymbol{u}_n] = \boldsymbol{UD}$$

where $\boldsymbol{D} = diag(\lambda_i)$. If $\boldsymbol{U}$ is non-singular, then $\boldsymbol{U}^{-1}\boldsymbol{AU} = \boldsymbol{D}$ and $\boldsymbol{A} = \boldsymbol{UDU}^{-1}$.

Combining the results from properties 2 and 5, we have that:

$$\boldsymbol{A}^k = \boldsymbol{UD}^k\boldsymbol{U}^{-1}$$

and

$$\boldsymbol{A}^{-1} = \boldsymbol{UD}^{-1}\boldsymbol{U}^{-1}$$

where $\boldsymbol{D}^k = diag\left(\lambda_i^k\right)$ and $\boldsymbol{D}^{-1} = diag\left(\frac{1}{\lambda_i}\right)$.

6.4 Properties of symmetric matrices

For symmetric matrices some interesting properties emerge. Firstly, all eigenvalues are real. Secondly, suppose we have two eigenvalues λ_k and λ_l of the symmetric matrix $\boldsymbol{A}$ such that $\lambda_k \neq \lambda_l$. Let $\boldsymbol{u}_k$ and $\boldsymbol{u}_l$ be the corresponding eigenvectors.

Then

$$\begin{aligned}
\lambda_k \boldsymbol{u}_l' \boldsymbol{u}_k &= \boldsymbol{u}_l' \lambda_k \boldsymbol{u}_k \\
&= \boldsymbol{u}_l' \boldsymbol{A} \boldsymbol{u}_k \\
&= \boldsymbol{u}_k' \boldsymbol{A}' \boldsymbol{u}_l \\
&= \boldsymbol{u}_k' \boldsymbol{A} \boldsymbol{u}_l \\
&= \boldsymbol{u}_k' \lambda_l \boldsymbol{u}_l \\
&= \lambda_l \boldsymbol{u}_k' \boldsymbol{u}_l \\
\lambda_k \boldsymbol{u}_l' \boldsymbol{u}_k &= \lambda_l \boldsymbol{u}_l' \boldsymbol{u}_k
\end{aligned}$$

Now, since $\lambda_k \neq \lambda_l$, this implies that $\boldsymbol{u}_l' \boldsymbol{u}_k = 0$. In other words, symmetric matrices with unique eigenvalues have orthogonal eigenvectors. It can also be shown that even with repeated eigenvalues, orthogonal sets of eigenvectors can be found (with the Gram-Schmidt process) when $\boldsymbol{A}$ is symmetric.

Thus it follows that if $\boldsymbol{A}$ is symmetric, then $\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{I}$, in which case $\boldsymbol{U}' = \boldsymbol{U}^{-1}$. Using Property 5 above, we can therefore see that

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{-1} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}'.$$

This can be used to define spectral decomposition for a symmetric matrix.

6.5 Spectral decomposition

The spectral decomposition of a symmetric matrix $\boldsymbol{A}$ follows directly from $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}'$:

$$\begin{aligned}
\boldsymbol{A} &= \begin{bmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_1' \\ \vdots \\ \boldsymbol{u}_n' \end{bmatrix} \\
&= \begin{bmatrix} \lambda_1 \boldsymbol{u}_1 & \cdots & \lambda_n \boldsymbol{u}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_1' \\ \vdots \\ \boldsymbol{u}_n' \end{bmatrix} \\
&= \sum_{k=1}^{n} \lambda_k \boldsymbol{u}_k \boldsymbol{u}_k'
\end{aligned}$$

Therefore, spectral decomposition expresses a symmetric matrix $\boldsymbol{A}$ as the sum of n rank 1 matrices.

6.6 Singular value decomposition

We will now generalise this idea to non-square matrices. Let $\boldsymbol{X}: n \times p$ be any matrix of rank k where $k \leq p < n$. Through singular value decomposition (SVD) the matrix $\boldsymbol{X}$ can be expressed in the form

$$\boldsymbol{X} = \boldsymbol{U}^* \boldsymbol{D}^* \boldsymbol{V}' \tag{6}$$

where $\boldsymbol{U}^*: n \times n$ and $\boldsymbol{V}: p \times p$ are orthogonal matrices, and

$$\boldsymbol{D}^*: n \times p = \begin{bmatrix} \boldsymbol{D}: k \times k & \boldsymbol{0}: k \times (p-k) \\ \boldsymbol{0}: (n-k) \times k & \boldsymbol{0}: (n-k) \times (p-k) \end{bmatrix}$$

where $\boldsymbol{D} = diag(d_i) = diag(\sqrt{\lambda_i})$. The positive quantities $\lambda_1, \ldots, \lambda_k$ are the non-zero eigenvalues of both the symmetric matrices $\boldsymbol{X'X}: p \times p$ and $\boldsymbol{XX'}: n \times n$.

Equation 6 is often called the full version of the SVD, since it can also be written in compact form:

$${}_n\boldsymbol{X}_p = {}_n\boldsymbol{U}_p \boldsymbol{D}_p \boldsymbol{V}_p'$$

where $\boldsymbol{U}$ consists of the first p columns of $\boldsymbol{U}^*$. Note that since ${}_n\boldsymbol{U}_p$ is not square, its columns are mutually orthonormal, but not the rows, i.e. ${}_p\boldsymbol{U}_n' \boldsymbol{U}_p = \boldsymbol{I}_p$, whilst ${}_n\boldsymbol{U}_p \boldsymbol{U}_n' \neq \boldsymbol{I}_n$.

The columns of the matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ are called the left and right singular vectors of $\boldsymbol{X}$ respectively. The matrix $\boldsymbol{D}$ is a diagonal matrix containing the singular values on the diagonal. It is assumed unless stated otherwise that the singular values are ordered in decreasing order.

Without loss of generality, the singular values can always be (and are) written as positive. This can be seen from the spectral decomposition, where $\boldsymbol{X} = \sum_{k=1}^{p} d_k \boldsymbol{u}_k \boldsymbol{v}_k'$. If any singular value is negative, we can simply multiply it by -1 and multiply either the left or right corresponding singular vector by -1 as well.

Summary of the relationships between singular and eigen- vectors and values

$\boldsymbol{U}$	is the left singular vectors of $\boldsymbol{X}$
$\boldsymbol{U}$	is the eigenvectors of $\boldsymbol{XX'}$, since $(\boldsymbol{XX'})\boldsymbol{U} = \boldsymbol{UD}^2$
$\boldsymbol{V}$	is the right singular vectors of $\boldsymbol{X}$
$\boldsymbol{V}$	is the eigenvectors of $\boldsymbol{X'X}$, since $(\boldsymbol{X'X})\boldsymbol{V} = \boldsymbol{VD}^2$
$\boldsymbol{D} = \boldsymbol{L}^{\frac{1}{2}}$	contains the singular values of $\boldsymbol{X}$ on the diagonal
$\boldsymbol{D}^2 = \boldsymbol{L}$	contains the eigenvalues of $\boldsymbol{XX'}$ and $\boldsymbol{X'X}$ on the diagonal

One practical problem in applying the SVD is that eigenvectors are defined in an arbitrary directional sense – if $\boldsymbol{x}$ is an eigenvector of $\boldsymbol{A}$, then so is $-\boldsymbol{x}$. One needs to check that the directions of the eigenvectors in $\boldsymbol{U}$ and in $\boldsymbol{V}$ are consistently defined.