Chapter 2

Linear Algebra

2.1 Introduction

We discuss vectors, matrices, transposes, covariance, correlation, diagonal and inverse matrices, orthogonality, subspaces and eigenanalysis. An alterntive source for much of this material is the excellent book by Strang [58].

2.2 Transposes and Inner Products

A collection of variables may be treated as a single entity by writing them as a *vector*. For example, the three variables x_1, x_2 and x_3 may be written as the vector

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \tag{2.1}$$

Bold face type is often used to denote vectors (scalars - single variables - are written with normal type). Vectors can be written as *column vectors* where the variables go down the page or as *row vectors* where the variables go across the page (it needs to be made clear when using vectors whether $\boldsymbol{x}$ means a row vector or a column vector - most often it will mean a column vector and in our text it will *always* mean a column vector, unless we say otherwise). To turn a column vector into a row vector we use the *transpose* operator

$$\boldsymbol{x}^T = [x_1, x_2, x_3] \tag{2.2}$$

The transpose operator also turns row vectors into column vectors. We now define the *inner product* of two vectors

$$\begin{aligned} \boldsymbol{x}^T \boldsymbol{y} &= [x_1, x_2, x_3] \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \qquad (2.3) \\ &= x_1 y_1 + x_2 y_2 + x_3 y_3 \end{aligned}$$

$$= \sum_{i=1}^{3} x_i y_i$$

which is seen to be a scalar. The *outer product* of two vectors produces a matrix

$$\begin{aligned} \boldsymbol{x}\boldsymbol{y}^T &= \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} [y_1, y_2, y_3] \\ &= \begin{bmatrix} x_1y_1 & x_1y_2 & x_1y_3 \\ x_2y_1 & x_2y_2 & x_2y_3 \\ x_3y_1 & x_3y_2 & x_3y_3 \end{bmatrix} \end{aligned} \tag{2.4}$$

An $N \times M$ matrix has N rows and M columns. The ijth entry of a matrix is the entry on the jth column of the ith row. Given a matrix $\boldsymbol{A}$ (matrices are also often written in bold type) the ijth entry is written as $\boldsymbol{A}_{ij}$. When applying the transpose operator to a matrix the ith row becomes the ith column. That is, if

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{2.5}$$

then

$$\boldsymbol{A}^T = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{bmatrix} \tag{2.6}$$

A matrix is *symmetric* if $\boldsymbol{A}_{ij} = \boldsymbol{A}_{ji}$. Another way to say this is that, for symmetric matrices, $\boldsymbol{A} = \boldsymbol{A}^T$.

Two matrices can be multiplied if the number of columns in the first matrix equals the number of rows in the second. Multiplying $\boldsymbol{A}$, an $N \times M$ matrix, by $\boldsymbol{B}$, an $M \times K$ matrix, results in $\boldsymbol{C}$, an $N \times K$ matrix. The ijth entry in $\boldsymbol{C}$ is the inner product between the ith row in $\boldsymbol{A}$ and the jth column in $\boldsymbol{B}$. As an example

$$\begin{bmatrix} 2 & 3 & 4 \\ 5 & 6 & 7 \end{bmatrix} \begin{bmatrix} 1 & 3 & 7 & 2 \\ 4 & 3 & 4 & 1 \\ 5 & 6 & 4 & 2 \end{bmatrix} = \begin{bmatrix} 34 & 39 & 42 & 15 \\ 64 & 75 & 87 & 30 \end{bmatrix} \tag{2.7}$$

Given two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ we note that

$$(\boldsymbol{A}\boldsymbol{B})^T = \boldsymbol{B}^T\boldsymbol{A}^T \tag{2.8}$$

2.2.1 Properties of matrix multiplication

Matrix multiplication is associative

$$(\boldsymbol{A}\boldsymbol{B})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{B}\boldsymbol{C}) \tag{2.9}$$

distributive

$$\boldsymbol{A}(\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{A}\boldsymbol{B} + \boldsymbol{A}\boldsymbol{C} \tag{2.10}$$

but not commutative

$$\boldsymbol{A}\boldsymbol{B} \neq \boldsymbol{B}\boldsymbol{A} \tag{2.11}$$

2.3 Types of matrices

2.3.1 Covariance matrices

In the previous chapter the covariance, σ_{xy}, between two variables x and y was defined. Given p variables there are $p \times p$ covariances to take account of. If we write the covariances between variables x_i and x_j as σ_{ij} then all the covariances can be summarised in a *covariance matrix* which we write below for $p = 3$

$$\boldsymbol{C} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix} \tag{2.12}$$

The ith diagonal element is the covariance between the ith variable and itself which is simply the variance of that variable; we therefore write σ_i^2 instead of σ_{ii}. Also, note that because $\sigma_{ij} = \sigma_{ji}$ covariance matrices are symmetric.

We now look at computing a covariance matrix from a given data set. Suppose we have p variables and that a single observation $\boldsymbol{x}_i$ (a row vector) consists of measuring these variables and suppose there are N such observations. We now make a matrix $\boldsymbol{X}$ by putting each $\boldsymbol{x}_i$ into the ith row. The matrix $\boldsymbol{X}$ is therefore an $N \times p$ matrix whose rows are made up of different observation vectors. If all the variables have zero mean then the covariance matrix can then be evaluated as

$$\boldsymbol{C} = \frac{1}{N-1}\boldsymbol{X}^T\boldsymbol{X} \tag{2.13}$$

This is a multiplication of a $p \times N$ matrix, $\boldsymbol{X}^T$, by a $N \times p$ matrix, $\boldsymbol{X}$, which results in a $p \times p$ matrix. To illustrate the use of covariance matrices for time series, figure 2.1 shows 3 time series which have the following covariance relation

$$\boldsymbol{C}_1 = \begin{bmatrix} 1 & 0.1 & 1.6 \\ 0.1 & 1 & 0.2 \\ 1.6 & 0.2 & 2.0 \end{bmatrix} \tag{2.14}$$

and mean vector

$$\boldsymbol{m}_1 = [13, 17, 23]^T \tag{2.15}$$

2.3.2 Diagonal matrices

A *diagonal matrix* is a square matrix ($M = N$) where all the entries are zero except along the diagonal. For example

$$\boldsymbol{D} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 6 \end{bmatrix} \tag{2.16}$$

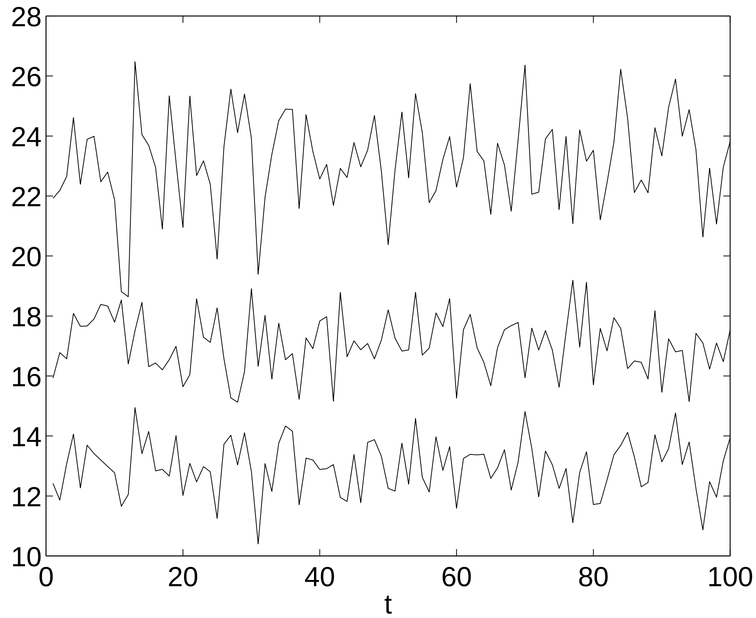

Figure 2.1: *Three time series having the covariance matrix* $\boldsymbol{C}_1$ *and mean vector* $\boldsymbol{m}_1$ *shown in the text. The top and bottom series have high covariance but none of the other pairings do.*

There is also a more compact notation for the same matrix

$$\boldsymbol{D} = diag([4, 1, 6]) \tag{2.17}$$

If a covariance matrix is diagonal it means that the covariances between variables are zero, that is, the variables are all uncorrelated. Non-diagonal covariance matrices are known as *full* covariance matrices. If $\boldsymbol{V}$ is a vector of variances $\boldsymbol{V} = [\sigma_1^2, \sigma_2^2, \sigma_3^2]^T$ then the corresponding diagonal covariance matrix is $\boldsymbol{V}_d = diag(\boldsymbol{V})$.

2.3.3 The correlation matrix

The correlation matrix, $\boldsymbol{R}$, can be derived from the covariance matrix by the equation

$$\boldsymbol{R} = \boldsymbol{BCB} \tag{2.18}$$

where $\boldsymbol{B}$ is a diagonal matrix of inverse standard deviations

$$\boldsymbol{B} = diag([1/\sigma_1, 1/\sigma_2, 1/\sigma_3]) \tag{2.19}$$

2.3.4 The identity matrix

The identity matrix is a diagonal matrix with ones along the diagonal. Multiplication of any matrix, $\boldsymbol{X}$ by the identity matrix results in $\boldsymbol{X}$. That is

$$\boldsymbol{IX} = \boldsymbol{X} \tag{2.20}$$

The identity matrix is the matrix equivalent of multiplying by 1 for scalars.

2.4 The Matrix Inverse

Given a matrix $\boldsymbol{X}$ its inverse $\boldsymbol{X}^{-1}$ is defined by the properties

$$\begin{aligned} \boldsymbol{X}^{-1}\boldsymbol{X} &= \boldsymbol{I} \\ \boldsymbol{X}\boldsymbol{X}^{-1} &= \boldsymbol{I} \end{aligned} \tag{2.21}$$

where $\boldsymbol{I}$ is the identity matrix. The inverse of a diagonal matrix with entries d_{ii} is another diagonal matrix with entries $1/d_{ii}$. This satisfies the definition of an inverse, eg.

$$\begin{bmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 6 \end{bmatrix} \begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.22}$$

More generally, the calculation of inverses involves a lot more computation. Before looking at the general case we first consider the problem of solving simultaneous equations. These constitute relations between a set of *input or independent* variables $\boldsymbol{x}_i$ and a set of *output or dependent* variables y_i. Each input-output pair constitutes an observation. In the following example we consider just $N = 3$ observations and $p = 3$ dimensions per observation

$$\begin{aligned} 2w_1 \quad & +w_2 + w_3 && = 5 \\ 4w_1 \quad & -6w_2 && = -2 \\ -2w_1 \quad & +7w_2 + 2w_3 && = 9 \end{aligned}$$

which can be written in matrix form

$$\begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 5 \\ -2 \\ 9 \end{bmatrix} \tag{2.23}$$

or in matrix form

$$\boldsymbol{X}\boldsymbol{w} = \boldsymbol{y} \tag{2.24}$$

This system of equations can be solved in a systematic way by subtracting multiples of the first equation from the second and third equations and then subtracting multiples of the second equation from the third. For example, subtracting twice the first equation from the second and -1 times the first from the third gives

$$\begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 8 & 3 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 5 \\ -12 \\ 4 \end{bmatrix} \tag{2.25}$$

Then, subtracting -1 times the second from the third gives

$$\begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 5 \\ -12 \\ 2 \end{bmatrix} \tag{2.26}$$

This process is known as *forward elimination*. We can then substitute the value for w_3 from the third equation into the second etc. This process is *back-substitution*. The two processes are together known as *Gaussian elimination*. Following this through for our example we get $\boldsymbol{w} = [1, 1, 2]^T$.

When we come to invert a matrix (as opposed to solve a system of equations as in the previous example) we start with the equation

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I} \tag{2.27}$$

and just write down all the entries in the $\boldsymbol{A}$ and $\boldsymbol{I}$ matrices in one big matrix

$$\begin{bmatrix} 2 & 1 & 1 & 1 & 0 & 0 \\ 4 & -6 & 0 & 0 & 1 & 0 \\ -2 & 7 & 2 & 0 & 0 & 1 \end{bmatrix} \tag{2.28}$$

We then perform forward elimination [1] until the part of the matrix corresponding to $\boldsymbol{A}$ equals the identity matrix; the matrix on the right is then $\boldsymbol{A}^{-1}$ (this is because in equation 2.27 if $\boldsymbol{A}$ becomes $\boldsymbol{I}$ then the left hand side is $\boldsymbol{A}^{-1}$ and the right side must equal the left side). We get

$$\begin{bmatrix} 1 & 0 & 0 & \frac{12}{16} & \frac{-5}{16} & \frac{-6}{16} \\ 0 & 1 & 0 & \frac{4}{8} & \frac{-3}{8} & \frac{-2}{8} \\ 0 & 0 & 1 & -1 & 1 & 1 \end{bmatrix} \tag{2.29}$$

This process is known as the *Gauss-Jordan* method. For more details see Strang's excellent book on Linear Algebra [58] where this example was taken from.

Inverses can be used to solve equations of the form $\boldsymbol{X}\boldsymbol{w} = \boldsymbol{y}$. This is achieved by multiplying both sides by $\boldsymbol{X}^{-1}$ giving

$$\boldsymbol{w} = \boldsymbol{X}^{-1}\boldsymbol{y} \tag{2.30}$$

Hence,

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} \frac{12}{16} & \frac{-5}{16} & \frac{-6}{16} \\ \frac{4}{8} & \frac{-3}{8} & \frac{-2}{8} \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ -2 \\ 9 \end{bmatrix} \tag{2.31}$$

which also gives $\boldsymbol{w} = [1, 1, 2]^T$.

The inverse of a product of matrices is given by

$$(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1} \tag{2.32}$$

Only square matrices are invertible because, for $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, if $\boldsymbol{y}$ and $\boldsymbol{x}$ are of different dimension then we will not necessarily have a one-to-one mapping between them.

[1] We do not perform back-substitution but instead continue with forward elimination until we get a diagonal matrix.

2.5 Orthogonality

The length of a d-element vector $\boldsymbol{x}$ is written as $||\boldsymbol{x}||$ where

$$\begin{aligned} ||\boldsymbol{x}||^2 &= \sum_{i=1}^{d} x_i^2 \\ &= \boldsymbol{x}^T \boldsymbol{x} \end{aligned} \tag{2.33}$$

Two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are *orthogonal* if

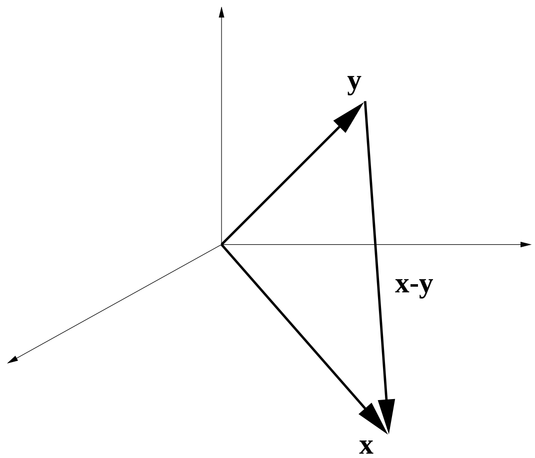

Figure 2.2: *Two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. These vectors will be orthogonal if they obey Pythagoras' relation ie. that the sum of the squares of the sides equals the square of the hypoteneuse.*

$$||\boldsymbol{x}||^2 + ||\boldsymbol{y}||^2 = ||\boldsymbol{x} - \boldsymbol{y}||^2 \tag{2.34}$$

That is, if

$$x_1^2 + ... + x_d^2 + y_1^2 + ... + y_d^2 = (x_1 - y_1)^2 + ... + (x_d - y_d)^2 \tag{2.35}$$

Expanding the terms on the right and re-arranging leaves only the cross-terms

$$\begin{aligned} x_1 y_1 + ... + x_d y_d &= 0 \\ \boldsymbol{x}^T \boldsymbol{y} &= 0 \end{aligned} \tag{2.36}$$

That is, two vectors are orthogonal if their inner product is zero.

2.5.1 Angles between vectors

Given a vector $\boldsymbol{b} = [b_1, b_2]^T$ and a vector $\boldsymbol{a} = [a_1, a_2]^T$ we can work out that

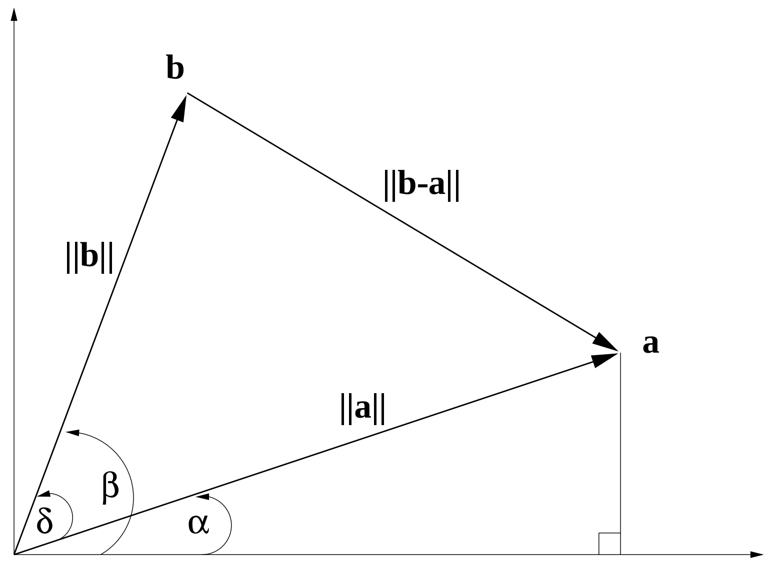

Figure 2.3: *Working out the angle between two vectors.*

$$
\begin{aligned}
\cos\alpha &= \frac{a_1}{||\boldsymbol{a}||} \qquad (2.37)\\
\sin\alpha &= \frac{a_2}{||\boldsymbol{a}||}\\
\cos\beta &= \frac{b_1}{||\boldsymbol{b}||}\\
\sin\beta &= \frac{b_2}{||\boldsymbol{b}||}
\end{aligned}
$$

$$(2.38)$$

Now, $cos\delta = cos(\beta - \alpha)$ which we can expand using the trig identity

$$\cos(\beta - \alpha) = \cos\beta\cos\alpha + \sin\beta\sin\alpha \tag{2.39}$$

Hence

$$\cos(\delta) = \frac{a_1 b_1 + a_2 b_2}{||\boldsymbol{a}||||\boldsymbol{b}||} \tag{2.40}$$

More generally, we have

$$\cos(\delta) = \frac{\boldsymbol{a}^T\boldsymbol{b}}{||\boldsymbol{a}||||\boldsymbol{b}||} \tag{2.41}$$

Because, $\cos\pi/2 = 0$, this again shows that vectors are orthogonal for $\boldsymbol{a}^T\boldsymbol{b} = 0$. Also, because $|\cos\delta| \leq 1$ where $|x|$ denotes the absolute value of x we have

$$|\boldsymbol{a}^T\boldsymbol{b}| \leq ||\boldsymbol{a}||||\boldsymbol{b}|| \tag{2.42}$$

which is known as the *Schwarz Inequality.*

2.5.2 Projections

The projection of a vector $\boldsymbol{b}$ onto a vector $\boldsymbol{a}$ results in a projection vector $\boldsymbol{p}$ which is the point on the line $\boldsymbol{a}$ which is closest to the point $\boldsymbol{b}$. Because $\boldsymbol{p}$ is a point on $\boldsymbol{a}$ it

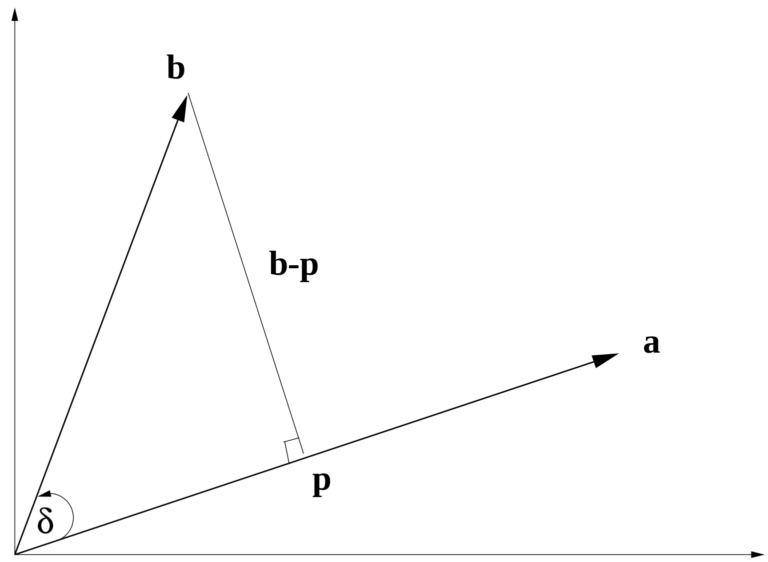

Figure 2.4: *The projection of* $\boldsymbol{b}$ *onto* $\boldsymbol{a}$ *is the point on* $\boldsymbol{a}$ *which is closest to* $\boldsymbol{b}$*.*

must be some scalar multiple of it. That is

$$\boldsymbol{p} = w\boldsymbol{a} \tag{2.43}$$

where w is some coefficient. Because $\boldsymbol{p}$ is the point on $\boldsymbol{a}$ *closest* to $\boldsymbol{b}$ this means that the vector $\boldsymbol{b} - \boldsymbol{p}$ is orthogonal to $\boldsymbol{a}$. Therefore

$$\begin{aligned} \boldsymbol{a}^T(\boldsymbol{b} - \boldsymbol{p}) &= 0 \\ \boldsymbol{a}^T(\boldsymbol{b} - w\boldsymbol{a}) &= 0 \end{aligned} \tag{2.44}$$

Re-arranging gives

$$w = \frac{\boldsymbol{a}^T\boldsymbol{b}}{\boldsymbol{a}^T\boldsymbol{a}} \tag{2.45}$$

and

$$\boldsymbol{p} = \frac{\boldsymbol{a}^T\boldsymbol{b}}{\boldsymbol{a}^T\boldsymbol{a}}\boldsymbol{a} \tag{2.46}$$

We refer to $\boldsymbol{p}$ as the *projection vector* and to w as the *projection*.

2.5.3 Orthogonal Matrices

The set of vectors $\boldsymbol{q}_1..\boldsymbol{q}_k$ are *orthogonal* if

$$\boldsymbol{q}_j^T\boldsymbol{q}_k = \begin{array}{ll} 0 & j \neq k \\ d_{jk} & j = k \end{array} \tag{2.47}$$

If these vectors are placed in columns of the matrix $\boldsymbol{Q}$ then

$$\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{Q}^T = \boldsymbol{D} \tag{2.48}$$

2.5.4 Orthonormal Matrices

The set of vectors $\boldsymbol{q}_1..\boldsymbol{q}_k$ are *orthonormal* if

$$\boldsymbol{q}_j^T \boldsymbol{q}_k = \begin{array}{ll} 0 & j \neq k \\ 1 & j = k \end{array} \tag{2.49}$$

If these vectors are placed in columns of the matrix $\boldsymbol{Q}$ then

$$\boldsymbol{Q}^T \boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{Q}^T = \boldsymbol{I} \tag{2.50}$$

Hence, the transpose equals the inverse

$$\boldsymbol{Q}^T = \boldsymbol{Q}^{-1} \tag{2.51}$$

The vectors $\boldsymbol{q}_1..\boldsymbol{q}_k$ are said to provide an *orthonormal basis*. This means that *any* vector can be written as a linear combination of the basis vectors. A trivial example is the two-dimensional cartesian coordinate system where $\boldsymbol{q}_1 = [1,0]^T$ (the x-axis) and $\boldsymbol{q}_2 = [0,1]^T$ (the y-axis). More generally, to represent the vector $\boldsymbol{x}$ we can write

$$\boldsymbol{x} = \tilde{x}_1 \boldsymbol{q}_1 + \tilde{x}_2 \boldsymbol{q}_2 + ... + \tilde{x}_d \boldsymbol{q}_d \tag{2.52}$$

To find the appropriate coefficients $\tilde{x}_k$(the co-ordinates in the new basis), multiply both sides by $\boldsymbol{q}_k^T$. Due to the orthonormality property all terms on the right disappear except one leaving

$$\tilde{x}_k = \boldsymbol{q}_k^T \boldsymbol{x} \tag{2.53}$$

The new coordinates are the projections of the data onto the basis functions (re. equation 2.45, there is no denominator since $\boldsymbol{q}_k^T \boldsymbol{q}_k = 1$). In matrix form, equation 2.52 can be written as $\boldsymbol{x} = \boldsymbol{Q}\tilde{\boldsymbol{x}}$ which therefore has the solution $\tilde{\boldsymbol{x}} = \boldsymbol{Q}^{-1}\boldsymbol{x}$. But given that $\boldsymbol{Q}^{-1} = \boldsymbol{Q}^T$ we have

$$\tilde{\boldsymbol{x}} = \boldsymbol{Q}^T \boldsymbol{x} \tag{2.54}$$

Transformation to an orthonormal basis preserves lengths. This is because [2]

$$\begin{aligned} ||\tilde{\boldsymbol{x}}|| &= ||\boldsymbol{Q}^T \boldsymbol{x}|| \\ &= (\boldsymbol{Q}^T \boldsymbol{x})^T \boldsymbol{Q}^T \boldsymbol{x} \\ &= \boldsymbol{x}^T \boldsymbol{Q}\boldsymbol{Q}^T \boldsymbol{x} \\ &= \boldsymbol{x}^T \boldsymbol{x} \\ &= ||\boldsymbol{x}|| \end{aligned} \tag{2.55}$$

Similarly, inner products and therefore angles between vectors are preserved. That is

$$\begin{aligned} \tilde{\boldsymbol{x}}^T \tilde{\boldsymbol{y}} &= (\boldsymbol{Q}^T \boldsymbol{x})^T \boldsymbol{Q}^T \boldsymbol{y} \\ &= \boldsymbol{x}^T \boldsymbol{Q}\boldsymbol{Q}^T \boldsymbol{y} \\ &= \boldsymbol{x}^T \boldsymbol{y} \end{aligned} \tag{2.56}$$

Therefore, transformation by an orthonormal matrix constitutes a *rotation* of the co-ordinate system.

[2]Throughout this chapter we will make extensive use of the matrix identities $(\boldsymbol{AB})^T = \boldsymbol{B}^T \boldsymbol{A}^T$ and $(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$. We will also use $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$.

2.6 Subspaces

A space is, for example, a set of real numbers. A subspace S is a set of points $\{x\}$ such that (i) if we take two vectors from S and add them we remain in S and (ii) if we take a vector from S and multiply by a scalar we also remain in S (S is said to be closed under addition and multiplication). An example is a 2-D plane in a 3-D space. A subspace can be defined by a basis.

2.7 Determinants

The determinant of a two-by-two matrix

$$\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{2.57}$$

is given by

$$\det(\boldsymbol{A}) = ad - bc \tag{2.58}$$

The determinant of a three-by-three matrix

$$\boldsymbol{A} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \tag{2.59}$$

is given by

$$\det(\boldsymbol{A}) = a \det\left(\begin{bmatrix} e & f \\ h & i \end{bmatrix}\right) - b \det\left(\begin{bmatrix} d & f \\ g & i \end{bmatrix}\right) + c \det\left(\begin{bmatrix} d & e \\ g & h \end{bmatrix}\right) \tag{2.60}$$

Determinants are important because of their properties. In particular, if two rows of a matrix are equal then the determinant is zero eg. if

$$\boldsymbol{A} = \begin{bmatrix} a & b \\ a & b \end{bmatrix} \tag{2.61}$$

then

$$\det(\boldsymbol{A}) = ab - ba = 0 \tag{2.62}$$

In this case the transformation from $\boldsymbol{x} = [x_1, x_2]^T$ to $\boldsymbol{y} = [y_1, y_2]^T$ given by

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y} \tag{2.63}$$

reduces two pieces of information (x_1 and x_2) to one piece of information

$$y = y_1 = y_2 = ax_1 + bx_2 \tag{2.64}$$

In this case it is not possible to reconstruct $\boldsymbol{x}$ from $\boldsymbol{y}$; the transformation is not invertible - the matrix $\boldsymbol{A}$ does not have an inverse and the value of the determinant provides a test for this: If $\det(\boldsymbol{A}) = 0$ the matrix $\boldsymbol{A}$ is not invertible; it is *singular*.

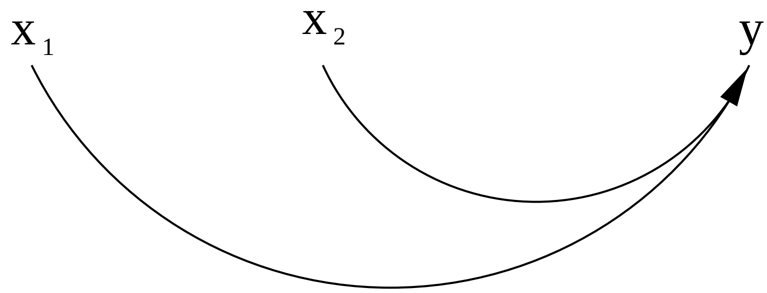

Figure 2.5: *A singular (non-invertible) transformation.*

Conversely, if $\det(\boldsymbol{A}) \neq 0$ then $\boldsymbol{A}$ *is* invertible. Other properties of the determinant are

$$\begin{aligned} \det(\boldsymbol{A}^T) &= \det(\boldsymbol{A}) \\ \det(\boldsymbol{AB}) &= \det(\boldsymbol{A}) \det(\boldsymbol{B}) \\ \det(\boldsymbol{A}^{-1}) &= 1/\det(\boldsymbol{A}) \\ \det(\boldsymbol{A}) &= \prod_k a_{kk} \end{aligned} \tag{2.65}$$

Another important property of determinants is that they measure the 'volume' of a matrix. For a 3-by-3 matrix the three rows of the matrix form the edges of a cube. The determinant is the volume of this cube. For a d-by-d matrix the rows form the edges of a 'parallepiped'. Again, the determinant is the volume.

2.8 Eigenanalysis

The square matrix $\boldsymbol{A}$ has eigenvalues λ and eigenvectors $\boldsymbol{q}$ if

$$\boldsymbol{Aq} = \lambda \boldsymbol{q} \tag{2.66}$$

Therefore

$$(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{q} = 0 \tag{2.67}$$

To satisfy this equation either $\boldsymbol{q} = 0$, which is uninteresting, or the matrix $\boldsymbol{A} - \lambda \boldsymbol{I}$ must reduce $\boldsymbol{q}$ to the null vector (a single point). For this to happen $\boldsymbol{A} - \lambda \boldsymbol{I}$ must be singular. Hence

$$\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0 \tag{2.68}$$

Eigenanalysis therefore proceeds by (i) solving the above equation to find the eigenvalues λ_i and then (ii) substituting them into equation 2.66 to find the eigenvectors. For example, if

$$\boldsymbol{A} = \begin{bmatrix} 4 & -5 \\ 2 & -3 \end{bmatrix} \tag{2.69}$$

then

$$\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = (4 - \lambda)(-3 - \lambda) - (-5)(2) = 0 \tag{2.70}$$

which can be rearranged as

$$\begin{aligned} \lambda^2 - \lambda - 2 &= 0 \\ (\lambda + 1)(\lambda - 2) &= 0 \end{aligned} \tag{2.71}$$

Hence the eigenvalues are $\lambda = -1$ and $\lambda = 2$. Substituting back into equation 2.66 gives an eigenvector $\boldsymbol{q}_1$ which is any multiple of $[1, 1]^T$. Similarly, eigenvector $\boldsymbol{q}_2$ is any multiple of $[5, 2]^T$.

We now note that the determinant of a matrix is also equal to the product of its eigenvalues

$$\det(\boldsymbol{A}) = \prod_k \lambda_k \tag{2.72}$$

We also define the *Trace* of a matrix as the sum of its diagonal elements

$$Tr(\boldsymbol{A}) = \sum_k a_{kk} \tag{2.73}$$

and note that it is also equal to the sum of the eigenvalues

$$Tr(\boldsymbol{A}) = \sum_k \lambda_k \tag{2.74}$$

Eigenanalysis applies only to *square* matrices.

2.9 Gram-Schmidt

A general class of procedures for finding eigenvectors are the *deflation methods* of which QR-decomposition and Gram-Schmidt orthogonalization are examples.

In Gram-Schmidt, we are given a set of vectors, say $\boldsymbol{a}$,$\boldsymbol{b}$ and $\boldsymbol{c}$ and we wish to find a set of corresponding orthonormal vectors which we'll call $\boldsymbol{q}_1$,$\boldsymbol{q}_2$ and $\boldsymbol{q}_3$. To start with we let

$$\boldsymbol{q}_1 = \frac{\boldsymbol{a}}{||\boldsymbol{a}||} \tag{2.75}$$

We then compute $\boldsymbol{b}'$ which is the original vector b minus the projection vector (see equation 2.46) of $\boldsymbol{b}$ onto q_1

$$\boldsymbol{b}' = \boldsymbol{b} - \boldsymbol{q}_1^T \boldsymbol{b} \boldsymbol{q}_1 \tag{2.76}$$

The second orthogonal vector is then a unit length version of b'

$$\boldsymbol{q}_2 = \frac{\boldsymbol{b}'}{||\boldsymbol{b}'||} \tag{2.77}$$

Finally, the third orthonormal vector is given by

$$\boldsymbol{q}_3 = \frac{\boldsymbol{c}'}{||\boldsymbol{c}'||} \tag{2.78}$$

where

$$\boldsymbol{c}' = \boldsymbol{c} - \boldsymbol{q}_1^T \boldsymbol{c} \boldsymbol{q}_1 - \boldsymbol{q}_2^T \boldsymbol{c} \boldsymbol{q}_2 \tag{2.79}$$

In QR-decomposition the Q terms are given by $\boldsymbol{q}_i$ and the R terms by $\boldsymbol{q}_i^T \boldsymbol{c}$.

2.9.1 Diagonalization

If we put the eigenvectors into the columns of a matrix

$$Q = \begin{bmatrix} | & | & . & | \\ | & | & . & | \\ \boldsymbol{q}_1 & \boldsymbol{q}_2 & . & \boldsymbol{q}_d \\ | & | & . & | \\ | & | & . & | \end{bmatrix} \tag{2.80}$$

then, because, $\boldsymbol{A}\boldsymbol{q}_k = \lambda_k \boldsymbol{q}_k$, we have

$$\boldsymbol{A}\boldsymbol{Q} = \begin{bmatrix} | & | & . & | \\ | & | & . & | \\ \lambda_1\boldsymbol{q}_1 & \lambda_2\boldsymbol{q}_2 & . & \lambda_d\boldsymbol{q}_d \\ | & | & . & | \\ | & | & . & | \end{bmatrix} \tag{2.81}$$

If we put the eigenvalues into the matrix $\boldsymbol{\Lambda}$ then the above matrix can also be written as $\boldsymbol{Q}\boldsymbol{\Lambda}$. Therefore,

$$\boldsymbol{A}\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{\Lambda} \tag{2.82}$$

Pre-multiplying both sides by $\boldsymbol{Q}^{-1}$ gives

$$\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q} = \boldsymbol{\Lambda} \tag{2.83}$$

This shows that any square matrix can be converted into a diagonal form (provided it has distinct eigenvalues; see eg. [58] p. 255). Sometimes there won't be d distinct eigenvalues and sometimes they'll be complex.

2.9.2 Spectral Theorem

For any real *symmetric* matrix all the eigenvalues will be real and there will be d distinct eigenvalues and eigenvectors. The eigenvectors will be orthogonal (if the matrix is not symmetric the eigenvectors won't be orthogonal). They can be normalised and placed into the matrix $\boldsymbol{Q}$. Since $\boldsymbol{Q}$ is now orthonormal we have $\boldsymbol{Q}^{-1} = \boldsymbol{Q}^T$. Hence

$$\boldsymbol{Q}^T\boldsymbol{A}\boldsymbol{Q} = \boldsymbol{\Lambda} \tag{2.84}$$

Pre-multiplying by $\boldsymbol{Q}$ and post-multiplying by $\boldsymbol{Q}^T$ gives

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T \tag{2.85}$$

which is known as the *spectral theorem*. It says that any real, symmetric matrix can be represented as above where the columns of $\boldsymbol{Q}$ contain the eigenvectors and $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues, λ_i. Equivalently,

$$\boldsymbol{A} = \begin{bmatrix} | & | & . & | \\ | & | & . & | \\ \boldsymbol{q}_1 & \boldsymbol{q}_2 & . & \boldsymbol{q}_d \\ | & | & . & | \\ | & | & . & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & & \\ & & & \lambda_d \end{bmatrix} \begin{bmatrix} - & - & \boldsymbol{q}_1 & - & - \\ - & - & \boldsymbol{q}_2 & - & - \\ . & . & . & & . \\ - & - & \boldsymbol{q}_d & - & - \end{bmatrix} \tag{2.86}$$

This can also be written as a summation

$$\boldsymbol{A} = \sum_{k=1}^{d} \lambda_k \boldsymbol{q}_k \boldsymbol{q}_k^T \tag{2.87}$$

2.10 Complex Matrices

If

$$\boldsymbol{A} = \begin{bmatrix} 3+2i & 4 & 6+3i \\ -2+i & 3+2i & 7+4i \end{bmatrix} \tag{2.88}$$

then the complex transpose or Hermitian transpose is given by

$$\boldsymbol{A}^H = \begin{bmatrix} 3-2i & -2-i \\ 4 & 3-2i \\ 6-3i & 7-4i \end{bmatrix} \tag{2.89}$$

ie. each entry changes into its complex conjugate (see appendix) and we then transpose the result. Just as $\boldsymbol{A}^{-T}$ denotes the transpose of the inverse so $\boldsymbol{A}^{-H}$ denotes the Hermitian transpose of the inverse.

If $\boldsymbol{A}^H \boldsymbol{A}$ is a diagonal matrix then $\boldsymbol{A}$ is said to be a *unitary matrix*. It is the complex equivalent of an orthogonal matrix.

2.11 Quadratic Forms

The quadratic function

$$f(\boldsymbol{x}) = a_{11}x_1^2 + a_{12}x_1x_2 + a_{21}x_2x_1 + ... + a_{dd}x_d^2 \tag{2.90}$$

can be written in matrix form as

$$f(\boldsymbol{x}) = [x_1, x_2, ..., x_d] \begin{bmatrix} a_{11} & a_{12} & a_{1d} \\ a_{21} & a_{22} & a_{2d} \\ & & \\ & & \\ a_{d1} & a_{d2} & a_{dd} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ . \\ x_d \end{bmatrix} \tag{2.91}$$

which is written compactly as

$$f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} \tag{2.92}$$

If $f(\boldsymbol{x}) > 0$ for any non-zero $\boldsymbol{x}$ then $\boldsymbol{A}$ is said to be positive-definite. Similarly, if $f(\boldsymbol{x}) \geq 0$ then $\boldsymbol{A}$ is positive-semi-definite.

If we substitute $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T$ and $\boldsymbol{x} = \boldsymbol{Q}\boldsymbol{y}$ where $\boldsymbol{y}$ are the projections onto the eigenvectors, then we can write

$$\begin{aligned} f(\boldsymbol{x}) &= \boldsymbol{y}^T \boldsymbol{\Lambda} \boldsymbol{y} \\ &= \sum_i y_i^2 \lambda_i \end{aligned} \tag{2.93}$$

Hence, for positive-definiteness we must therefore have $\lambda_i > 0$ for all i (ie. positive eigenvalues).

2.11.1 Ellipses

For 2-by-2 matrices if $\boldsymbol{A} = \boldsymbol{I}$ then we have

$$f = x_1^2 + x_2^2 \tag{2.94}$$

which is the equation of a circle with radius $\sqrt{f}$. If $\boldsymbol{A} = k\boldsymbol{I}$ we have

$$\frac{f}{k} = x_1^2 + x_2^2 \tag{2.95}$$

The radius is now $\sqrt{f/k}$. If $\boldsymbol{A} = diag([k_1, k_2])$ we have

$$f = k_1 x_1^2 + k_2 x_2^2 \tag{2.96}$$

which is the equation of an ellipse. For $k_1 > k_2$ the major axis has length $\sqrt{f/k_2}$ and the minor axis has length $\sqrt{f/k_1}$.

For a non-diagonal $\boldsymbol{A}$ we can diagonalise it using $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T$. This gives

$$f = \lambda_1 \tilde{x}_1^2 + \lambda_2 \tilde{x}_2^2 \tag{2.97}$$

where the ellipse now lives in a new co-ordinate system given by the rotation $\tilde{\boldsymbol{x}} = \boldsymbol{x}^T\boldsymbol{Q}$. The major and minor axes have lengths $\sqrt{f/\lambda_2}$ and $\sqrt{f/\lambda_1}$.