Linear Algebra

William G. Faris

May 4, 2005

Contents

Chapter 1

Matrices

1.1 Matrix operations

In the following there will be two cases: numbers are real numbers, or numbers are complex numbers. Much of what we say will apply to either case. The real case will be in mind most of the time, but the complex case will sometimes prove useful. Sometimes numbers are called scalars.

An m be n matrix A is an array of numbers with m rows and n columns. The entry in the ith row and jth column is denoted A_{ij}. There are three kinds of operations.

(1) Vector space operations. If A and B are both m by n matrices, then the *sum* matrix $A + B$ satisfies $(A + B)_{ij} = A_{ij} + B_{ij}$. Furthermore, if λ is a number, then the *scalar multiple* matrix λA satisfies $(\lambda A)_{ij} = \lambda A_{ij}$.

There is always a matrix $-A = (-1)A$. Subtraction is a special case of addition, since $A - B = A + (-B)$. There is an m by n *zero matrix* 0 with all entries equal to zero.

(2) Matrix multiplication. If A is an m by n matrix, and B is an n by p matrix, then the *product* matrix AB is defined by $(AB)_{ik} = \sum_{j=1}^{n} A_{ij}B_{jk}$.

For each n there is an n by n matrix I called the n bt n *identity matrix*. It is defined by $I_{ij} = \delta_{ij}$, where the object on the right is the Kronecker symbol.

(3) Transpose or conjugate transpose. In the real case the *transpose* A' is defined by $(A')_{ij} = A_{ji}$. In the complex case the conjugate transpose A^* is defined by $(A^*)_{ij} = \bar{A}_{ji}$.

A m by 1 matrix may be identified with an m component *vector* (or column vector). A 1 by n matrix may be identified may be identified with an n component *covector* or *linear form* (or row vector). If $m = n$ the product of a covector with a vector (in that order) is a 1 by 1 matrix, which may be identified with a scalar. The product of a vector with a covector (in that order) is an m by n matrix.

If u and v are real vectors, then when they have the same number of components their *inner product* is $u'v$. In general their *outer product* is uv'. The inner product is a scalar, while the outer product is a square matrix. (In the complex case one would use the conjugate transpose instead of the transpose.)

1.2 Problems

1. A real square matrix R is said to be orthogonal if $R'R = I$. Prove that the product of two orthogonal matrices is orthogonal. That is, show that if R is orthogonal and S is orthogonal, then RS is orthogonal.
2. Let R be a matrix with $R^2 = I$ and $R = R'$. Prove that R is orthogonal.
3. Let R be a 2 by 2 matrix with $R^2 = I$ and $R = R'$. Show that from these conditions it follows that
$$R = \begin{bmatrix} a & b \\ b & -a \end{bmatrix}$$
or R is diagonal. Give an exact description of all such matrices, including whatever additional conditions the entries must satisfy.
4. Let P be a real square matrix with $P^2 = P$ and $P = P'$. Let Q be another such matrix. Show that if P and Q commute, then PQ is yet another such matrix.
5. Let P be a 2 by 2 matrix with $P^2 = P$ and $P = P'$. Show that from these conditions it follows that
$$P = \begin{bmatrix} a & \pm\sqrt{ad} \\ \pm\sqrt{ad} & d \end{bmatrix}$$
with $0 \leq a \leq 1$, $0 \leq d \leq 1$ and $a + d = 1$, or P is diagonal. Give an exact description of all such matrices, including whatever additional conditions the entries must satisfy.
6. Show that in the non-diagonal case of the previous problem the matrix P may be written
$$P = \begin{bmatrix} c^2 & cs \\ cs & s^2 \end{bmatrix}.$$
Write this in the form of an outer product.
7. Consider a real square matrix N with $N^2 = 0$. Is it possible that each entry of N is non-zero? Give a proof that your answer is correct.
8. Consider a real square matrix N with $N^2 = 0$. Suppose that N is symmetric, that is, $N' = N$. Does it follow that $N = 0$? Prove or disprove.
9. If $M^2 = 0$ and $N^2 = 0$, does it follow that $(MN)^2 = 0$? Prove or disprove.
10. If $M^2 = 0$ and $N^2 = 0$, does it follow that $(M+N)^2 = 0$? Give a complete argument.

Chapter 2

Applications

2.1 Matrix transformations

An m by n matrix A can define a *linear transformation* from $\mathbf{R}^n$ to $\mathbf{R}^m$ by defining the value of A on the column vector $\mathbf{x}$ to be the transformed vector $\mathbf{x}' = A\mathbf{x}$.

This is particularly interesting when A is a square n by n matrix, so the transformation is from $\mathbf{R}^n$ to itself. Then the transformation can be repeated.

The following examples are 2 by 2 examples. The original input vector and the transformed output vector are both in $\mathbf{R}^2$. Thus the transformed vector is

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \tag{2.1}$$

In the first example the vector is a geometric vector in the plane. In the second example the vector is a probability vector.

2.2 Orthogonal matrices

A matrix is *orthogonal* if $R'R = I$. Define the rotation matrix corresponding to angle θ by

$$R_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}. \tag{2.2}$$

This matrix is an orthogonal matrix. Notice that every matrix of the form

$$R = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \tag{2.3}$$

with $a^2 + b^2 = 1$ may be written as a rotation matrix. The rotation matrix acts as a linear transformation of vectors. It rotates them.

Now define the projection matrix P_χ as the outer product

$$P_\chi = \begin{bmatrix} \cos(\chi) \\ \sin(\chi) \end{bmatrix} \begin{bmatrix} \cos(\chi) & \sin(\chi) \end{bmatrix} = \begin{bmatrix} \cos^2(\chi) & \sin(\chi)\cos(\chi) \\ \sin(\chi)\cos(\chi) & \sin^2(\chi) \end{bmatrix}. \tag{2.4}$$

This projects vectors from the plane onto a line through the origin. Notice that $P_\chi^2 = P_\chi$ and $P_\chi = P_\chi'$. It is easy to see that this is equal to

$$P_\chi = \frac{1}{2}\begin{bmatrix} 1+\cos(2\chi) & \sin(2\chi) \\ \sin(2\chi) & 1-\cos(2\chi) \end{bmatrix}. \tag{2.5}$$

Now define the reflection matrix $H_\chi = P_\chi - (I - P_\chi) = 2P_\chi - I$. This reflects vectors across the same line. Then

$$H_\chi = \begin{bmatrix} \cos(2\chi) & \sin(2\chi) \\ \sin(2\chi) & -\cos(2\chi) \end{bmatrix}. \tag{2.6}$$

This matrix is an orthogonal matrix. Notice that every matrix of the form

$$H = \begin{bmatrix} a & b \\ b & -a \end{bmatrix} \tag{2.7}$$

with $a^2 + b^2 = 1$ may be written as a reflection matrix.

2.3 Approach to equilibrium

Consider a land where each day is either a day of rain or a day of sun. Suppose that the conditional probability of tomorrow's weather given the past weather depends only on today's weather. In fact, suppose that the probability of sun tomorrow given rain today is p, and the probability of rain tomorrow given sun today is q. For the moment all we know is that $0 \leq p \leq 1$ and $0 \leq q \leq 1$.

Suppose that on a certain day the probability of rain is a and the probability of sun is b. Here $a + b = 1$. Then the next day the probability of rain is a' and the probability of sun is b'. The relation is that $a' = (1-p)a + qb$ and $b' = pa + (1-q)b$. Notice that also $a' + b' = 1$.

This equation may be written in matrix form as

$$\begin{bmatrix} a' \\ b' \end{bmatrix} = \begin{bmatrix} 1-p & q \\ p & 1-q \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}. \tag{2.8}$$

The first question is whether there is an equilibrium probability where $a' = a$ and $b' = b$. The answer is obtained by solving the system. The answer is that the equilibrium is given by $a^* = q/(p+q)$ and $b^* = p/(p+q)$. At least this makes sense when either $p > 0$ or $q > 0$. (If they are both zero, then every choice of a and b with $a + b = 1$ gives an equilibrium probability. This is because in this situation the weather never changes.)

A more interesting question is the approach to equilibrium. Say that you start with initial probabilities $a = a^* + c$ and $b = b^* - c$. Then the probability after n days is given by the matrix power

$$\begin{bmatrix} a_n \\ b_n \end{bmatrix} = \begin{bmatrix} 1-p & q \\ p & 1-q \end{bmatrix}^n \begin{bmatrix} a \\ b \end{bmatrix}. \tag{2.9}$$

This may be written

$$\begin{bmatrix} a_n \\ b_n \end{bmatrix} = \begin{bmatrix} 1-p & q \\ p & 1-q \end{bmatrix}^n \begin{bmatrix} a^* \\ b^* \end{bmatrix} + \begin{bmatrix} 1-p & q \\ p & 1-q \end{bmatrix}^n \begin{bmatrix} c \\ -c \end{bmatrix}. \tag{2.10}$$

Each time one applies the matrix to the equilibrium probability one just gets the same equilibrium probability. (The equilibrium is an eigenvector with eigenvalue 1.) Each time one applies the matrix to the vector with sum zero one simply multiplies the vector by a scalar factor of $\lambda = 1 - p - q$. (The sum zero vector is an eigenvector with eigenvalue λ.) So

$$\begin{bmatrix} a_n \\ b_n \end{bmatrix} = \begin{bmatrix} a^* \\ b^* \end{bmatrix} + (1-p-q)^n \begin{bmatrix} c \\ -c \end{bmatrix}. \tag{2.11}$$

The multiplicative factor satisfies $-1 \leq \lambda = 1 - p - q \leq 1$. If it satisfies $-1 < \lambda = 1 - p - q < 1$, then the powers approach zero, and in the long run the weather will settle down to its equilibrium.

The result may also be written simply as a property of the matrix power. In fact, the first column of the matrix power may be recovered by taking $a = 1$ and $b = 0$, and the second column comes from $a = 0$ and $b = 1$. The conclusion is that

$$\begin{bmatrix} 1-p & q \\ p & 1-q \end{bmatrix}^n = \begin{bmatrix} a^* & a^* \\ b^* & b^* \end{bmatrix} + (1-p-q)^n \begin{bmatrix} b^* & -a^* \\ -b^* & a^* \end{bmatrix}. \tag{2.12}$$

Note: In most serious treatments of Markov chains the probability vectors are row vectors, and the transition matrices have row sum equal to one. In other words, everything is the transpose of what we have here. The only reason for having probability vectors as column vectors and transition matrices with column sum one is to have the matrices act on the vectors to the right, according to the custom in elementary linear algebra.

2.4 Problems

1. Prove the projection matrix formula involving the double angle. That is, start with the projection matrix formula involving $\cos(\chi)$ and $\sin(\chi)$ that is obtained directly from the computing the outer product. Derive from this the following projection matrix formula involving $\cos(2\chi)$ and $\sin(2\chi)$.

2. Show that the product of two rotation matrices with angles θ_2 and θ_1 is a rotation matrix with angle $\theta_1 + \theta_2$.

3. Show that the product of two reflection matrices with angles χ_2 and χ_1 is a rotation matrix with a certain angle. What is this angle?

4. Solve the system of equations involving the transition probability matrix to obtain the equilibrium values for a^* and b^*.

5. Prove the assertions about the eigenvalues of the transition probability matrix.

6. If $p = 1/3$ and $q = 1/6$, then roughly how many days does it take until an equilibrium is reached, at least for practical purposes?

7. What happens if $p = q = 1$? Describe both mathematically and in terms of a story about the weather. How would a weather forecaster deal with such a weather pattern?

8. What happens if $p + q = 1$? What are the equilibrium probabilities? What weather pattern over time does this describe? How would a weather forecaster deal with such a weather pattern?

Chapter 3

Systems of Linear Equations

3.1 Gauss elimination

Gaussian elimination is a method for solving systems of linear equations. The basic technique is to replace an equation by the the sum of the equation with another equation. When this is done correctly, at each stage one more zero is introduced, without destroying the ones that have already been introduced. In the end one gets a row echelon form system, for which the solution is easy.

Gaussian Example 1. Say that the equation is $A\mathbf{x} = \mathbf{b}$, specifically

$$\begin{bmatrix} 2 & 3 & -1 \\ 4 & 4 & -3 \\ -2 & 3 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \\ 1 \end{bmatrix}. \tag{3.1}$$

All we care about is the coefficients in the augmented matrix

$$Ab = \begin{bmatrix} 2 & 3 & -1 & 5 \\ 4 & 4 & -3 & 3 \\ -2 & 3 & -1 & 1 \end{bmatrix}. \tag{3.2}$$

The rows correspond to the equations, and the first three columns correspond to the three variables.

Add -2 times the first row to the second row. Add 1 times the first row to the third row. The result is two zeros in the first column. Add 3 times the second row to the third row. The result is one zero in the second column. So a row echelon form is

$$Jd = \begin{bmatrix} 2 & 3 & -1 & 5 \\ 0 & -2 & -1 & -7 \\ 0 & 0 & -5 & -15 \end{bmatrix}. \tag{3.3}$$

The corresponding equation $J\mathbf{x} = \mathbf{d}$ is

$$\begin{bmatrix} 2 & 3 & -1 \\ 0 & -2 & -1 \\ 0 & 0 & -5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 5 \\ -7 \\ -15 \end{bmatrix}. \tag{3.4}$$

Such an equation may be solved by first solving for x_3, then for x_2, and finally for x_1. Thus the solution is $x_3 = 3$, $x_2 = 2$, and $x_1 = 1$.

Here is the algorithm for *Gaussian elimination*. It works to solve an arbitrary linear systems $A\mathbf{x} = \mathbf{b}$, where A is an m by n matrix. The way it works it this. At a certain stage of the computation one has a matrix with the first j columns in *row echelon form*. Each of these columns is either a pivot column or a non-pivot column. A pivot column has a non-zero entry (the pivot entry) with all entries to the left and below equal to zero. A non-pivot column has only non-zero entries in rows that correspond to pivot entries to the left.

Suppose that there are k pivot columns among the j columns. To get the $j + 1$st column in the appropriate form, see if all the entries from $k + 1$ down are zero. If so, then this is a non-pivot column, and nothing needs to be done. Otherwise, interchange two rows in the range from $k+1$ to m to bring a non-zero entry to the $k + 1, j + 1$ pivot position. Then put zeros below this by replacing the rows below by their sums with appropriate multiples of the $k + 1$ row. The result is a pivot column. So now the first $j+1$ columns are in row echelon form.

Gaussian Example 2. Say that the equation is $A\mathbf{x} = \mathbf{b}$, specifically

$$\begin{bmatrix} 2 & 1 & 2 \\ 4 & 2 & 4 \\ -2 & -1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 10 \\ 20 \\ -5 \end{bmatrix}. \tag{3.5}$$

All we care about is the coefficients in the augmented matrix

$$Ab = \begin{bmatrix} 2 & 1 & 2 & 10 \\ 4 & 2 & 4 & 20 \\ -2 & -1 & 3 & -5 \end{bmatrix}. \tag{3.6}$$

The rows correspond to the equations, and the first three columns correspond to the three variables.

Add -2 times the first row to the second row. Add 1 times the first row to the third row. The result is two zeros in the first column. The first column is a pivot column. The second column is a non-pivot column. Interchange the last two rows. The third column is a non-pivot column. The fourth column is non-pivot. So a row echelon form is

$$Jd = \begin{bmatrix} 2 & 1 & 2 & 10 \\ 0 & 0 & 5 & 5 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{3.7}$$

The corresponding equation $J\mathbf{x} = \mathbf{d}$ is

$$\begin{bmatrix} 2 & 1 & 2 \\ 0 & 0 & 5 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 10 \\ 5 \\ 0 \end{bmatrix}. \tag{3.8}$$

To solve such an equation, look at the variables corresponding to non-pivot columns. These may be assigned arbitrary values. Then solve for the pivot column variables. Thus the solution is $x_3 = 1$ and $x_1 = 4 - \frac{1}{2}x_2$.

3.2 Linear combinations

Let $\mathbf{a}_1, \ldots, \mathbf{a}_n$ be a list (finite sequence) of m-component vectors. If $c_1, \ldots, c_n$ are scalars, then the vector $c_1\mathbf{a}_1 + \cdots + c_n\mathbf{a}_n$ is called a *linear combination* of the list of vectors. (If the list has zero vectors in it, then there is precisely one linear combination, the zero vector.)

Let $\mathbf{a}_1, \ldots, \mathbf{a}_n$ be a list of vectors. The set of all possible linear combinations of these vectors is the *span* of the list of vectors.

Theorem 3.1 (Linear dependence theorem) *A list of vectors is linearly dependent if and only if one of the vectors is a linear combination of a list involving the other vectors.*

Theorem 3.2 (Vector-Matrix Correspondence) *Let A be an m by n matrix with column vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$. Let $\mathbf{c}$ be the column vector with entries $c_1, \ldots, c_n$. Then*

$$A\mathbf{c} = \sum_{j=1}^{n} c_j \mathbf{a}_j. \tag{3.9}$$

Thus a matrix times a column vector is the same as a linear combination of the columns of the matrix.

Theorem 3.3 (Span-Range Correspondence) *Let A be a matrix with columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$. Then a vector $\mathbf{b}$ is in the span of the columns, that is,*

$$\sum_{j=1}^{n} x_j \mathbf{a}_j = \mathbf{b} \tag{3.10}$$

if and only if there is a solution $\mathbf{x}$ of the matrix equation

$$A\mathbf{x} = \mathbf{b}. \tag{3.11}$$

Let $\mathbf{a}_1, \ldots, \mathbf{a}_n$ be a list of vectors. It is said to be *linearly independent* if whenever $c_1\mathbf{a}_1 + \cdots c_n\mathbf{a}_n = \mathbf{0}$, then it follows that $c_1 = 0, \ldots, c_n = 0$.

Theorem 3.4 (Linear Independence—Trivial Null-space Correspondence) *Let A be a matrix with columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$. The fundamental identity implies the columns are linearly independent if and only if the only solution $\mathbf{x}$ of the homogeneous matrix equation*

$$A\mathbf{x} = \mathbf{0} \tag{3.12}$$

is the zero solution $\mathbf{x} = \mathbf{0}$.

Theorem 3.5 (Uniqueness of coefficients) *Let A be a matrix with columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$. Suppose that the columns are linearly independent. If for some vector $\mathbf{b}$ we have $x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{b}\mathbf{a}_n = \mathbf{b}$, then the coefficients $x_1, \ldots, x_n$ are uniquely determined.*

Theorem 3.6 (Uniqueness of solution) *If the columns of A are linearly independent, then the matrix equation*

$$A\mathbf{x} = \mathbf{b} \tag{3.13}$$

can have at most one solution $\mathbf{x}$.

3.3 The Hermite matrix

Let A be an m by n matrix with columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$. A column $\mathbf{a}_j$ is called a *pivot column* if it is not a linear combination of previously defined pivot columns to the left of it. The number of pivot columns is the *rank* of A.

The pivot columns of A are linearly independent. To prove this, consider a linear combination of the pivot columns that gives the zero vector. Suppose that there were a coefficient that was non equal to zero. Take the last such coefficient. The corresponding pivot vector would then be a linear combination of the previous pivot vectors. This is a contradiction. Therefore every coefficient must be zero.

Each non-pivot column of A is a linear combination of the pivot columns of A to the left. The coefficients in such a linear combination are uniquely determined, because of the linear independence of the pivot columns.

Example: Take

$$A = \begin{bmatrix} 1 & 2 & 1 & 2 & 0 & 2 & 4 \\ 3 & 6 & 0 & 3 & 3 & 3 & 6 \\ 0 & 0 & 2 & 2 & -2 & 1 & 2 \\ 2 & 4 & 3 & 5 & -1 & 4 & 9 \end{bmatrix}. \tag{3.14}$$

The pivot columns are 1, 3, 6. The rank is 3.

Let H be an m by n matrix. Then A is a *Hermite* (or *reduced row echelon form*) matrix if it has the following form. The pivot columns are the standard basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_r$, ordered from left to right. Each remaining column is a linear combination of the pivot columns that occur to the left of it.

Each m by n matrix A defines a unique Hermite matrix H by the following rule. The pivot columns of A define the pivot columns of H. The non-pivot columns of A are expressed as linear combinations of pivot columns to the left with certain coefficients, and the same coefficients are used for H.

The Hermite matrix H of A is a description of how the columns of A depend on the pivot columns of A to the left.

Example: The associated Hermite matrix is

$$H = \begin{bmatrix} 1 & 2 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{3.15}$$

3.4 Invertible matrices

A square matrix A is *invertible* if there is a matrix B with $AZ = ZA = I$. If A is invertible, then its *inverse* $Z = A^{-1}$ is unique, and

$$AA^{-1} = A^{-1}A = I. \tag{3.16}$$

The identity matrix I is invertible, and it is its own inverse. If A is invertible, then so is A^{-1}. Finally, if A, B are both invertible, then their matrix product AB is invertible, and

$$(AB)^{-1} = B^{-1}A^{-1}. \tag{3.17}$$

Theorem 3.7 (Left Multiplication) *Let A be an m by n matrix. Let $\mathbf{b}$ be an n component vector. Left multiplication by an invertible m by n matrix does not change the solution set of $A\mathbf{x} = \mathbf{b}$. That is, $\mathbf{x}$ is a solution of $A\mathbf{x} = \mathbf{b}$ if and only if $\mathbf{x}$ is a solution of $EA\mathbf{x} = E\mathbf{b}$.*

As a special case, left multiplication of A by an invertible matrix does not change the linear dependence relations among the columns of A. This is because the linear dependence relations among the columns of A are just the solutions of the homogeneous system $A\mathbf{x} = \mathbf{0}$.

Let A be an m by n matrix. There are three elementary row operations that one can perform on A. They are:

1. Interchange two rows;
2. Replace a row by a non-zero scalar multiple of itself;
3. Replace a row by the sum of the row with a scalar multiple of another row.

Let I be the m by m identity matrix. There are three kinds of *elementary matrices* obtained by performing the three kinds of row operations on this identity matrix. They are:

1. *Reflect* across a diagonal;
2. *Multiply* by a non-zero scalar in a coordinate direction.
3. *Shear* in a coordinate direction according to a multiple of the value of another coordinate.

Each of these elementary matrices is invertible.

Let A be an m by n matrix, and let E be an elementary matrix. Then EA is the matrix obtained from A by performing the elementary row operation. In other words, an elementary row operation is the same as left multiplication by an elementary matrix.

3.5 Computing the Hermite matrix

Theorem 3.8 (Hermite form) *Let A be an m by n matrix. Then there is an invertible matrix E such that $EA = H$ is the Hermite matrix of A.*

Proof: *Gauss-Jordan elimination* is Gauss elimination followed by further operations to bring the matrix to Hermite form. The proof uses Gauss-Jordan elimination.

The invertible matrix is built up as a product of elementary matrices. The Hermite matrix is built up from left to right, column by column. Suppose that by a product of elementary matrices A has been transformed so that the first $j \geq 0$ columns form a Hermite matrix. Suppose that there are $k \leq j$ pivot vectors in the first j columns. Look at column $j+1$. If the entries from $k+1, j+1$ to $m, j+1$ are all zero, then the first $j+1$ columns are already in Hermite form. The column is a non-pivot column. Otherwise, interchange two rows in the range from $k+1$ to m to get a non-zero element in the $k+1, j+1$ position. Multiply the $k+1$st row to make this element equal to 1. Then use multiples of the $k+1$st row to produce zeros in the entries from $k+1, j+1$ to $m, j+1$. Also use multiplies of the $k+1$st row to produce zeros in the entries from $1, j+1$ to $k-1, j+1$. This produces a standard basis vector in the $j+1$ column. In this case the column is a pivot column. The process may be continued until the last column on the right is reached. This ends the proof.

3.6 Solving linear systems

Theorem 3.9 (General solution of homogeneous equation) *Let A be an m by n matrix, and consider the homogeneous equation*

$$A\mathbf{x} = \mathbf{0}. \tag{3.18}$$

Let H be the Hermite form of A. Then the solutions are the same as the solutions of $H\mathbf{x} = \mathbf{0}$. Suppose that A and H each have rank r. Let $\mathbf{y}$ be an $n-r$ component vector consisting of variables corresponding to the non-pivot columns of H. The solutions are of the form $\mathbf{x} = N\mathbf{y}$, where the nullspace matrix N is an n by $n-r$ matrix of rank $n-r$. Thus every solution of the homogeneous equation may be expressed as a linear combination of the columns of N.

Theorem 3.10 (Particular solution of inhomogeneous equation) *Let A be an m by n matrix, and consider the equation*

$$A\mathbf{x} = \mathbf{b}. \tag{3.19}$$

Augment A with an extra column $\mathbf{b}$ or the right. Let $H, \mathbf{c}$ be the Hermite form of $A, \mathbf{b}$. There is a solution if and only if $\mathbf{c}$ is a non-pivot column. Then a particular solution is obtained by solving $H\mathbf{x} = \mathbf{c}$ with the variables corresponding to non-pivot columns of H set equal to zero. This expresses $\mathbf{b}$ as a linear combination of the pivot columns of A.

Theorem 3.11 (General solution of inhomogeneous equation) *Let A be an m by n matrix, and consider the equation*

$$A\mathbf{x} = \mathbf{b}. \tag{3.20}$$

Let $\mathbf{p}$ be a particular solution of the homogeneous equation. Then the general solution

$$\mathbf{x} = \mathbf{p} + N\mathbf{y} \tag{3.21}$$

is the sum of the particular solution with the general solution of the homogeneous equation.

Example: Say that after Gauss-Jordan elimination the equation reduces to associated Hermite matrix is

$$\begin{bmatrix} 1 & 0 & a & b & 0 & c & 0 \\ 0 & 1 & d & e & 0 & f & 0 \\ 0 & 0 & 0 & 0 & 1 & g & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} p \\ q \\ r \\ s \end{bmatrix}. \tag{3.22}$$

The pivot columns are 1, 2, 5. The general solution is expressed in terms of variables corresponding to the non-pivot columns. Thus it is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} p \\ q \\ 0 \\ 0 \\ r \\ 0 \\ s \end{bmatrix} + \begin{bmatrix} -a & -b & -c \\ -d & -e & -f \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -g \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_3 \\ x_4 \\ x_6 \end{bmatrix}. \tag{3.23}$$

Every solution may be expressed as a sum of the particular solution column vector with a linear combination of the three columns vectors of the nullspace matrix.

3.7 Canonical forms

There are two basic computations. The first starts with a matrix A and computes its Hermite matrix H. This matrix is also called the reduced row echelon form or the row canonical form. This matrix displays the dependencies of later columns on earlier columns.

The second starts with a matrix A and computes its null-space matrix N. This is the matrix obtained by solving $H\mathbf{x} = \mathbf{0}$ and expressing the solution $\mathbf{x} = N\mathbf{y}$, where $\mathbf{y}$ is the vector of non-pivot variables, in order of increasing index. The null-space matrix N gives a parametric description of all solutions of the homogeneous equation $A\mathbf{x} = \mathbf{0}$.

As an example, suppose that

$$A = \begin{bmatrix} 1 & 1 & -2 & 4 & 5 \\ 2 & 2 & -3 & 1 & 3 \\ 3 & 3 & -4 & -2 & 1 \end{bmatrix}. \tag{3.24}$$

Then the Hermite form (reduced row echelon form) is

$$H = \begin{bmatrix} 1 & 1 & 0 & -10 & -9 \\ 0 & 0 & 1 & -7 & -7 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{3.25}$$

This tells us, for instance, that in the original matrix A the fourth column is -10 times the first column plus -7 times the third column. Also, we learn that the pivot columns 1 and 3 in the original matrix A are linearly independent.

The Hermite form above has three non-pivot columns. Therefore the null-space matrix N has three columns. It is

$$N = \begin{bmatrix} -1 & 10 & 9 \\ 1 & 0 & 0 \\ 0 & 7 & 7 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{3.26}$$

If one reflects this matrix across the diagonal that runs from lower right to upper left (not the usual transpose), one gets the matrix

$$N_h = \begin{bmatrix} 1 & 0 & 7 & 0 & 9 \\ 0 & 1 & 7 & 0 & 10 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}. \tag{3.27}$$

Curiously enough, this matrix is a Hermite matrix. (What is its null-space matrix?)

Say that we wanted to solve the inhomogeneous system for which A is the augmented matrix. This is equivalent to solving the homogeneous system with one more variable, where the last column is the right hand side. Then one only looks at solutions where the last variable is given the value -1. In other words, in the N matrix above one uses all the columns but the last (without the bottom row) to get the general solution of the homogeneous equation, and one uses the negative of the last column (without the bottom entry) to get the particular solution of the inhomogeneous equation.

3.8 Problems

1. Consider the four columns of the matrix in Gaussian example 1. Are the first two columns linearly independent? Are the first three columns linearly independent? Are the first four columns linearly independent. Give complete proofs.

2. Consider the four columns of the matrix in Gaussian example 2. Are the first two columns linearly independent? Are the first three columns linearly independent? Are the first four columns linearly independent. Give complete proofs.

3. Consider the four columns of the matrix in Gaussian example 1. Is the fourth column in the span of the first three columns? Give a complete proof.

4. Consider the matrix in Gaussian example 1. Is every 3 component vector in the span of the first three columns? Give a complete proof.

5. Consider the four columns of the matrix in Gaussian example 2. Is the fourth column in the span of the first three columns? Give a complete proof.

6. Consider the matrix in Gaussian example 2. Is every 3 component vector in the span of the first three columns? Give a complete proof.

7. Consider the 3 by 4 augmented matrix in Gaussian example 1. Find its Hermite matrix. Also find its 4 by 1 null-space matrix.

8. Show how the previous problem gives the solution of the original inhomogeneous equation in three unknowns (as a 3 by 1 column vector).

9. Consider the 3 by 4 augmented matrix in Gaussian example 2. Find its Hermite matrix. Also find its 4 by 2 null-space matrix.

10. Show how the previous problem gives the solution of the inhomogeneous equation in three unknowns as the sum of a particular solution vector (3 by 1) and the null-space solution (given by a 3 by 1 matrix).

Chapter 4

Invertible matrices

4.1 Left and right inverses

Consider an m by n matrix A. The *null-space* of A is the set of all n-component vector solutions $\mathbf{x}$ of the homogeneous equation $A\mathbf{x} = \mathbf{0}$. The *range* is the set of all m-component vectors of the form $A\mathbf{x}$, for some vector $\mathbf{x}$.

Theorem 4.1 (Left inverse theorem) *If A has a left inverse B with $BA = I$, then the null-space of A is trivial, that is, it consists only of the n-component vector with all zeros.*

Theorem 4.2 (Right inverse theorem) *If A has a right inverse B with $AB = I$, then every m-component vector is in the range of A.*

An m by n matrix is A *invertible* if there is another n by n matrix B with both $AB = I$ and $BA = I$. This matrix is denoted A^{-1}. So if A^{-1} exists, then

$$AA^{-1} = A^{-1}A = I. \tag{4.1}$$

Theorem 4.3 (Two sided inverse theorem) *Suppose that A is a square matrix that has a left inverse B and also has a right inverse C. Then $B = C$, and so A has an inverse.*

The proof is purely algebraic. Suppose that $BA = I$ and $AC = I$. Then

$$B = BI = BAC = IC = C. \tag{4.2}$$

For 2 by 2 matrices the inverse is easy. Suppose

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}. \tag{4.3}$$

Then

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}. \tag{4.4}$$

This formula should be memorized.

4.2 Finding inverses

Theorem 4.4 (Null space criterion) *Suppose that A is a square matrix with trivial null-space. Then A is invertible.*

The proof follows by writing $A = ER$, where E is a product of elementary matrices, and R is a Hermite matrix. Suppose that R is not the identity matrix. Since R is a square matrix, it then must have at least one non-pivot column. However then the equation $R\mathbf{x} = \mathbf{0}$ has a non-trivial solution $\mathbf{x}$. Therefore $A\mathbf{x} = \mathbf{0}$. This contradicts the assumption that A has trivial null-space. We conclude that $R = I$. Hence $A = E$ is a product of elementary matrices, and hence A is invertible.

It is now clear how to compute the inverse of a matrix A. Start with A and augment it by I. Apply the elementary row operations to A and the same elementary row operations to I. Then we get R augmented by E, where $R = EA$ is in Hermite form, and E is a product of elementary matrices. Then A is invertible precisely when $R = I$. In that case $I = EA$, and so $A^{-1} = E$.

Theorem 4.5 (Range criterion) *Suppose that A is a n by n matrix with all n-component vectors in its range. Then A is invertible.*

The proof follows by writing $A = ER$, where E is a product of elementary matrices, and R is a Hermite matrix. Suppose that R is not the identity matrix. Since R is a square matrix, it then must have at least one non-pivot column, so the bottom row must be zero. Consider the vector $\mathbf{c}$ that is zero except for a 1 in the bottom entry. Let $\mathbf{b} = E^{-1}\mathbf{c}$. Suppose that $A\mathbf{x} = \mathbf{b}$. Then $R\mathbf{x} = \mathbf{c}$. However this implies that $0 = 1$, which leads to a contradiction. Therefore $R = I$. Hence $A = E$ is a product of elementary matrices, and hence A is invertible.

4.3 Problems

1. Find the inverse of the 3 by 3 matrix in Gaussian example 1.

2. Let A be a square matrix. Its quadratic form is the function that sends a column vector $\mathbf{x}$ to the number $\mathbf{x}'A\mathbf{x}$. Prove that A and A' have the same quadratic form, that is, $\mathbf{x}'A\mathbf{x} = \mathbf{x}'A'\mathbf{x}$. Also, show that if $A' = -A$, then its quadratic form is zero.

3. Show that if A is a square matrix with $A' = -A$, then $I - A$ has an inverse. Hint: Show that $(I - A)\mathbf{x} = \mathbf{0}$ implies $\mathbf{x} = \mathbf{0}$. To accomplish this, use the previous problem.

4. Suppose that A is a square matrix with $A' = -A$. Prove that $R = (I + A)(I - A)^{-1}$ is an orthogonal matrix.

5. Say that $P(t)$ is a square matrix that depends on time, and suppose that $P(0) = I$. Suppose that $A = -A'$ is a square matrix, and that

$$\frac{dP(t)}{dt} = AP(t). \tag{4.5}$$

Prove that for each t the matrix $P(t)$ is orthogonal. Hint: Differentiate $P(t)'P(t)$.

6. Say that A is the matrix

$$A = \begin{bmatrix} 0 & -\omega \\ \omega & 0 \end{bmatrix}. \tag{4.6}$$

Find $P(t)$. Hint: Write this explicitly as a system of two ordinary differential equations.

7. For computer calculations it is common to approximate the differential equation by fixing a small $h \neq 0$ and solving

$$\frac{1}{h}(P(t+h) - P(t)) = \frac{1}{2}(AP(t+h) + AP(t)) \tag{4.7}$$

for $P(t+h)$ in terms of $P(t)$. The average on the right is supposed to help with the stability of the calculation. Show that $P(t+h)$ is an orthogonal matrix Q times $P(h)$.

8. For the two by two matrix A above, compute this orthogonal matrix Q explicitly.

9. We know that if AB has an inverse, then so does BA. Show that if $I - AB$ has an inverse, then so does $I - BA$.

Chapter 5

Vector spaces

5.1 Axioms

In the following the field of scalars will be either the real number field or the complex number field. In most of our thinking it will be the real number field.

The elements of a field form an abelian group under addition, with 0 as the additive identity and negation giving the additive inverse. The non-zero elements of a field form an abelian group under multiplication, with 1 as the multiplicative identity and the reciprocal as the multiplicative inverse. The multiplication is related to the addition by the distributive law.

Explicitly, the additive axioms state that

$$\begin{aligned} a+b &= b+a && (5.1)\\ (a+b)+c &= a+(b+c) && (5.2)\\ a+0 &= a && (5.3)\\ a+(-a) &= 0. && (5.4) \end{aligned}$$

The multiplicative axioms state that for every $a \neq 0, b \neq 0, c \neq 0$

$$\begin{aligned} ab &= ba && (5.5)\\ (ab)c &= a(bc) && (5.6)\\ a1 &= a && (5.7)\\ aa^{-1} &= 1. && (5.8) \end{aligned}$$

The distributive law states that

$$a(b+c) = ab + ac. \tag{5.9}$$

A *vector space* is an abelian group together with scalar multiplication satisfying certain axioms.

The abelian group axioms describe the addition of vectors. For every ordered pair of vectors $\mathbf{u}, \mathbf{v}$ in the vector space V there must be a sum vector $\mathbf{u}+\mathbf{v}$. This operation is commutative and associative. There is a zero vector $\mathbf{0}$, and every vector $\mathbf{u}$ has an additive inverse vector $-\mathbf{u}$. These axioms are summarized as

$$\begin{aligned} \mathbf{u}+\mathbf{v} &= \mathbf{v}+\mathbf{u} &(5.10)\\ (\mathbf{u}+\mathbf{v})+\mathbf{w} &= \mathbf{u}+(\mathbf{v}+\mathbf{w}) &(5.11)\\ \mathbf{u}+\mathbf{0} &= \mathbf{u} &(5.12)\\ \mathbf{u}+(-\mathbf{u}) &= \mathbf{0}. &(5.13) \end{aligned}$$

There is also an operation that sends every pair $a, \mathbf{u}$ into a vector $a\mathbf{u}$. This scalar multiplication satisfies the axioms

$$\begin{aligned} a(\mathbf{u}+\mathbf{v}) &= a\mathbf{u}+a\mathbf{v} &(5.14)\\ (a+b)\mathbf{u} &= a\mathbf{u}+b\mathbf{u} &(5.15)\\ (ab)\mathbf{u} &= a(b\mathbf{u}) &(5.16)\\ 1\mathbf{u} &= \mathbf{u}. &(5.17) \end{aligned}$$

The first axiom is group addition to group addition, the second axiom is field addition to group addition, the third axiom is field multiplication to composition of transformations, and the fourth axiom is field multiplicative identity to identity transformation.

5.2 Subspaces

A *subspace* (or vector subspace) of a vector space V is a subset W that is itself a vector space when one restricts the vector space operations to vectors in W.

In order for a subset W to be a subspace, one needs three conditions:

1. $\mathbf{0} \in W$
2. $\mathbf{u}, \mathbf{v} \in W$ imply $\mathbf{u}+\mathbf{v} \in W$
3. a scalar, $\mathbf{u} \in W$ imply $a\mathbf{u} \in W$

Examples: Consider the vector space $V = \mathbf{R}^3$. There are four kinds of subspaces. One consists only of the zero vector. Another is a line through the origin. Another is a plane through the origin. The final one is V itself.

Consider a list (finite sequence) of vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$. A *linear combination* of these vectors is a vector of the form

$$\mathbf{u} = c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k. \tag{5.18}$$

The set of all linear combinations of a list of vectors is called the *span* of the list of vectors. The span of a list of vectors is always a subspace.

The standard example of a vector space with real scalars is the vector space $\mathbf{R}^n$ of n-component column vectors. Once we have this example, we can find others by looking at subspaces.

Another example of a vector space is the space $C([a, b]$ of all continuous real functions p on the interval $[a, b]$. An example of a subspace would be the set of all such functions satisfying the additional condition $\int_a^b p(y)\, dy = 0$.

5.3 Bases and dimension

A list of vectors is *linearly independent* if the only linear combination of the vectors that gives the zero vector is the trivial linear combination in which each of the scalar coefficients is equal to zero.

A list of vectors has *span* W if every vector in W is a linear combination of the vectors in the list.

A *basis* for W is a list of vectors that are linearly independent and span W. (In some treatments this is called an *ordered basis.*)

Theorem 5.1 (Dimension comparison theorem) *Let* $\mathbf{u}_1, \ldots, \mathbf{u}_k$ *be linearly independent in* W*. Let* $\mathbf{v}_1, \ldots, \mathbf{v}_m$ *span* W*. Then* $k \leq m$*.*

The proof starts by writing each $\mathbf{u}_j$ as a linear combination

$$\mathbf{u}_j = \sum_{i=1}^{m} b_{ij}\mathbf{v}_i. \tag{5.19}$$

This can be done because the $\mathbf{v}_i$ span W. Suppose that a linear combination of the columns of the m by k matrix B is zero, that is, that

$$\sum_{j=1}^{n} c_j b_{ij} = 0 \tag{5.20}$$

for each i. Then

$$\sum_{j=1}^{n} c_j \mathbf{u}_j = \sum_{i=1}^{m} \sum_{j=1}^{n} c_j b_{ij} \mathbf{v}_i = \mathbf{0}. \tag{5.21}$$

Since the $\mathbf{u}_j$ are linearly dependent, it follows that the coefficients c_j are each zero. This argument proves that the columns of the m by k matrix B are linearly independent. Thus the Hermite form of the m by k matrix B has all pivot columns, which forces $k \leq m$.

Theorem 5.2 (Dimension characterization theorem) *Let* W *be a vector space with a finite basis. Then every basis for* W *has the same number of vectors.*

This theorem says that if $\mathbf{u}_1, \ldots, \mathbf{u}_k$ is a basis for W, and $\mathbf{v}_1, \ldots, \mathbf{v}_m$ is a basis for W, then $k = m$. The proof is easy. Since $\mathbf{u}_1, \ldots, \mathbf{u}_k$ is linearly independent, and $\mathbf{v}_1, \ldots, \mathbf{v}_m$ span W, it follows that $k \leq n$. Since $\mathbf{v}_1, \ldots, \mathbf{v}_m$ is linearly independent, and $\mathbf{u}_1, \ldots, \mathbf{u}_k$ span W, it follows that $m \leq k$. These two inequalities imply that $k = m$.

The number of vectors in a basis for W is called the *dimension* of W.

Example. Consider a list of column vectors in $\mathbf{R}^n$. Let W be their span. The problem is to find a basis for this subspace W. The solution is simple. Let A be the matrix with the vectors as columns. The pivot columns of A are the basis. (To find the pivot columns, one must find the Hermite form of the matrix. However the pivot columns that form the basis are the pivot columns of the original matrix A.)

As a specific example, consider the matrix

$$A = \begin{bmatrix} -1 & -2 & 2 & 1 \\ 1 & 2 & 2 & 3 \\ 4 & 8 & 2 & 6 \end{bmatrix}. \tag{5.22}$$

The four columns of this matrix span a subspace of $\mathbf{R}^3$. A basis for this subspace is given by the columns of

$$B = \begin{bmatrix} -1 & 2 \\ 1 & 2 \\ 4 & 2 \end{bmatrix}. \tag{5.23}$$

This shows that the subspace has dimension two. It is a plane through the origin. Of course all four columns of A are in this plane, but both the second and fourth column of A may be expressed as a linear combination of the first and third columns.

Example: Consider a homogeneous equation $A\mathbf{x} = \mathbf{0}$. The set of solutions of such an equation is a subspace, the null-space of A. The problem is to find a basis for this subspace. The solution is simple. Let N be the null-space matrix constructed from A. Then the columns of N are a basis for the null-space of A.

As a specific example, consider the problem of finding a basis for the solutions of $x_1 + x_2 + x_3 = 0$. The matrix $A = [1\ 1\ 1]$. The subspace is spanned by the columns of

$$N = \begin{bmatrix} -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{5.24}$$

The null-space has dimension two. It is a plane through the origin. Both the column vectors are in this plane, since the sum of the entries is zero.

5.4 The standard basis

In some contexts it is useful to speak of the *standard basis* for $\mathbf{R}^n$. This is the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ consisting of the columns of the identity matrix I. The vector $\mathbf{e}_j$ is called the jth *standard unit basis vector*.

5.5 Problems

1. Consider the matrix

$$A = \begin{bmatrix} 1 & 1 & 3 & 1 & 5 \\ 2 & 3 & 8 & 4 & 13 \\ 1 & 3 & 7 & 6 & 13 \\ 3 & 5 & 13 & 9 & 25 \\ 2 & 3 & 8 & 7 & 19 \end{bmatrix}. \tag{5.25}$$

Find a basis for its null-space (the space of solutions of the homogeneous equation). Find the dimension of the null-space.

2. Consider the subspace spanned by the five columns of the matrix A in the preceding problem. Find a basis for this subspace. Find the dimension.

3. Let $\mathbf{R}^\infty$ be the set consisting of all sequences $c = (c_1, c_2, c_3, \ldots)$ of real numbers. If a, b are two such sequences, define their sum by $(a + b)_j = a_j + b_j$. Define the scalar multiple ta by $(ta)_j = ta_j$. Which vector space axioms are satisfied by $\mathbf{R}^\infty$, and which are not?

4. Let ℓ^2 be the set consisting of all sequences $c = (c_1, c_2, c_3, \ldots)$ of real numbers such that the squared length $\sum_{j=1}^\infty c_j^2 < \infty$. Prove that this set of sequences is a subspace of the vector space of the previous problem. (This vector space is called *Hilbert space.*) Hint: What you need to prove is that if a and b are such sequences, then $c = a + b$ is such a sequence. Here $c_j = a_j + b_j$. First prove that $(a_j + b_j)^2 = a_j^2 + 2a_jb_j + b_j^2 \leq 2a_j^2 + 2b_j^2$.

Chapter 6

Linear transformations

6.1 Functions

A *function* $f : S \to T$ is a rule that assigns to each element of a set S (the *domain*) a unique element of another set T (the *target*). The value of the function f on x in S is $f(x)$ in T.

Consider a function $f : S \to T$. The function f is *injective* (or one-to-one) if for every y in T there is at most one x in S with $f(x) = y$. (That is, f is injective if $f(x) = f(x')$ implies $x = x'$.) The function f is *surjective* (or onto) if for every y in T there is at least one x in S with $f(x) = y$. The function f is *bijective* if it is both injective and surjective.

The *range* or *image* of a function S is the set of all y in T for which there exists x in S with $f(x) = y$. A function is surjective precisely when its range is the same as the target.

Say that $f : S \to T$ and $g : R \to S$ are functions. Their *composition* $f \circ g$ is the function defined by $(f \circ g)(t) = f(g(t))$. If S is a set, then the *identity* function $\mathbf{1}_S$ is defined by $\mathbf{1}_S(x) = x$ for x in S. Notice that $f \circ \mathbf{1}_S = f$ and $\mathbf{1}_T \circ f = f$.

If $f : S \to T$ is a bijective function, then there is a unique *inverse function* $f^{-1} : T \to S$. Thus $f^{-1}(y) = x$ if and only if $f(x) = y$. The relation between f and f^{-1} may also be expressed by saying that $f^{-1} \circ f = \mathbf{1}_S$ and $f \circ f^{-1} = \mathbf{1}_T$. In other words, f^{-1} is a (two-sided) inverse of f under composition.

A great part of mathematics falls naturally within this framework. Say that $f : S \to T$ is a function, and y is in T. A common problem is to solve the equation $f(x) = y$ for x. This is called an implicit description of x. If f is injective, the solution is unique. If f is surjective, then the solution exists for each y in T. When the solution is not unique, then there is an entire set $S_y \subseteq S$ of solutions of $f(x) = y$. Then an explicit or parametric solution is an injective function $g : R \to S$ with range S_y. The set R is the parameter space. Finding such a function g is often what is meant by "solving" a math problem.

Here is the obvious linear algebra example. For a homogeneous system

$A\mathbf{x} = \mathbf{0}$ of m linear equations in n unknowns, the solution set is the null-space, defined implicitly. It is a space of dimension $n - r$, where r is the rank. When one "solves" the system, this amounts to giving the solution in parametric form. This is determined by a n by $n - r$ matrix N that defines an injective function from $\mathbf{R}^{n-r}$ (the non-pivot variables) to the solution space.

6.2 Linear transformations

Let V and W be vector spaces. A *linear transformation* (or *linear mapping* $f : V \to W$ is a function that always satisfies

$$f(\mathbf{u} + \mathbf{v}) = f(\mathbf{u}) + f(\mathbf{v}) \tag{6.1}$$

$$f(a\mathbf{u}) = af(\mathbf{u}). \tag{6.2}$$

The standard example of a linear transformation is when $V = \mathbf{R}^n$ and $W = \mathbf{R}^m$ and $f(\mathbf{x}) = A\mathbf{x}$, where A is an m by n matrix.

Consider the vector spaces $C([a, b])$ consisting of all continuous real functions on the interval $[a, b]$. Consider the subspace $C^1([a, b])$ consisting of all real functions on the interval $[a, b]$ that have a derivative that is continuous on $[a, b]$. An example of a linear transformation would be $D : C^1([a, b]) \to C([a, b]$, that is, differentiation. Thus

$$(Dp)(x) = p'(x). \tag{6.3}$$

Another example would be the integration transformation I_a from $C([a, b])$ to $C^1([a, b])$ defined by

$$(I_a q)(x) = \int_a^x q(y)\, dy. \tag{6.4}$$

Notice that the range of this transformation is the subspace of functions p that have the zero value $p(a) = 0$ at a. The relation between these two transformations is

$$DI_a q = q, \tag{6.5}$$

that is, the derivative of the integral is the original function. On the other hand, we have a more complicated relation in the other direction. Thus

$$I_a Dp = p - p(a), \tag{6.6}$$

that is, the integral of the derivative is the original function with a suitably adjusted constant of integration.

Example: Consider the following linear transformation. Its domain consists of the subspace of $C^1([a, b])$ consisting of all p with $p(a) = 0$. Its target is $C([a, b])$. The transformation sends p to $Dp + cp$. Show that it is bijective. To do this, we must solve the differential equation

$$\frac{d}{dx}p(x) + cp(x) = s(x) \tag{6.7}$$

for an arbitrary continuous function s and find a unique solution. Multiply by e^{cx}. This gives

$$\frac{d}{dx}(e^{cx}p(x)) = e^{cx}. \tag{6.8}$$

Integrate and multiply by the decay factor e^{-cx}. This gives

$$p(x) = \int_a^x e^{-c(x-y)}s(y)\,dy. \tag{6.9}$$

Because of the boundary condition $p(a) = 0$ the constant of integration is uniquely determined to be zero.

Given a linear transformation $f : V \to W$, there are two associated subspaces. The *null-space* or *kernel* of f consists of all solutions $\mathbf{u}$ of the equation $f(\mathbf{u}) = \mathbf{0}$. It is a subspace of the domain V. The *range* or *image* of f is the set of all vectors $\mathbf{w}$ such that there is a solution of the equation $f(\mathbf{x}) = \mathbf{w}$. It is a subspace of the target space W.

It is clear that a linear transformation $f : V \to W$ is surjective if and only if its range is W. When is it injective? Clearly, if it is injective, then its null-space consists only of the zero vector.

Theorem 6.1 (Null space theorem) *Assume $f : V \to W$ is a linear transformation. Suppose its null-space consists only of the zero vector. Then f is injective.*

This theorem is more or less obvious. Suppose that the null-space of f consists only of the zero vector. Suppose that $f(\mathbf{u}) = f(\mathbf{v})$. Then by linearity $f(\mathbf{u}-\mathbf{v}) = f(\mathbf{u}) - f(\mathbf{v}) = \mathbf{0}$. Therefore $\mathbf{u}-\mathbf{v}$ is in the null-space. Hence $\mathbf{u}-\mathbf{v} = \mathbf{0}$ and so $\mathbf{u} = \mathbf{v}$. This proves that $f(\mathbf{u}) = f(\mathbf{v})$ implies $\mathbf{u} = \mathbf{v}$. This is enough to prove that f is injective.

Say that $f : V \to W$ is linear and both injective and surjective. Then f is called a *linear isomorphism*, or, when the context is clear, an *isomorphism*.

Theorem 6.2 (Rank-nullity theorem) *Let $f : V \to W$ be a linear transformation from an n finite-dimensional vector space to an m dimensional vector space. The the dimension r of the range (the rank) plus the dimension of the null-space (the nullity) equals the dimension n of the domain space.*

Here is an illustration of the theorem in the case of a matrix transformation. Let A be an m by n matrix. Let N be its null-space matrix. This has columns corresponding to the non-pivot columns of A. Take also r columns corresponding to the pivot-columns of A, in such a way that the jth column is the jth unit basis vector. (The images of these columns form a basis for the range of A.) Let L be n by n matrix with these columns. Then the columns of L form a basis for $\mathbf{R}^n$ that includes a basis for the null-space of A.

As an example, take

$$A = \begin{bmatrix} 1 & 1 & -2 & 4 & 5 \\ 2 & 2 & -3 & 1 & 3 \\ 3 & 3 & -4 & -2 & 1 \end{bmatrix}. \tag{6.10}$$

The pivot columns are the first and third. The basis matrix L is

$$L = \begin{bmatrix} 1 & -1 & 0 & 10 & 9 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 7 & 7 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{6.11}$$

The first and third columns are unit basis vectors, corresponding to the fact that the rank is 2. The second, fourth, and fifth columns are basis vectors for the null-space, corresponding to the fact that the nullity is 3. And indeed, the sum of the rank and the nullity is 5.

6.3 Affine transformations

In advanced mathematics it is customary to distinguish between linear transformations and affine transformations. (In elementary mathematics these are all called linear.)

An example of a linear transformation is the transformation that sends the column vector $\mathbf{x}$ to the column vector $\mathbf{y} = A\mathbf{x}$.

A scalar constant c can define a linear transformation by sending $\mathbf{x}$ to $c\mathbf{x}$. This is the same as the linear transformation cI. For this kind of linear transformation the output is proportional to the input. A linear transformation given by a matrix corresponds to a more general concept of proportionality.

An example of an affine transformation is the transformation that sends the column vector $\mathbf{x}$ to the column vector $\mathbf{y} = A\mathbf{x} + \mathbf{b}$. Here $\mathbf{b}$ is a fixed column vector. According to the definitions of linear algebra, this is not a linear transformation (unless $\mathbf{b}$ is the zero vector).

There is a trick that reduces a affine transformations to linear transformations acting on a special kind of vector. Thus one can write

$$\begin{bmatrix} \mathbf{y} \\ 1 \end{bmatrix} = \begin{bmatrix} A & \mathbf{b} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix}. \tag{6.12}$$

So that is why we mainly concentrate on linear transformations.

6.4 Problems

1. List all 64 functions from $S = \{1, 2, 3\}$ to $T = \{a, b, c, d\}$.

2. Say that S and T are finite, with S having k elements and T having n elements. How many functions $f : S \to T$ are there?

3. Say that S is finite with k elements and T has two points. How many functions $f : S \to T$ are there? How many subsets of S are there? What is the explanation for the coincidence that you observe?

4. List all 24 injective functions from $S = \{1, 2, 3\}$ to $T = \{a, b, c, d\}$. List all 4 subsets of T with precisely 3 elements.

5. Say that S and T are finite, with k and n elements. How many injective functions $f : S \to T$ are there? How many subsets of T with k elements are there? What is the relation between the two results?

6. Say that $f : S \to T$ is a function. Suppose that it has a left inverse $g : T \to S$ with $g \circ f = 1_S$. Show that f is injective.

7. Say that $g : T \to S$ is a function. Suppose that it has a right inverse $f : S \to T$ with $g \circ f = 1_S$. Show that g is surjective.

8. Say that $f : S \to T$ is an injective function and that $S \neq \emptyset$. Show that it has a left inverse $g : T \to S$ with $g \circ f = 1_S$.

9. Say that $g : T \to S$ is a surjective function. Show that it has a right inverse $f : S \to T$ with $g \circ f = 1_S$.

10. Consider the matrix

$$A = \begin{bmatrix} 1 & 1 & 3 & 1 & 5 \\ 2 & 3 & 8 & 4 & 13 \\ 1 & 3 & 7 & 6 & 13 \\ 3 & 5 & 13 & 9 & 25 \\ 2 & 3 & 8 & 7 & 19 \end{bmatrix}. \tag{6.13}$$

    Each column of the null-space matrix belongs to $\mathbf{R}^5$, and these columns form a basis for a subspace of $\mathbf{R}^5$. Find a basis for $\mathbf{R}^5$ that consists of null-space matrix columns together with unit basis vectors.

11. Consider the following linear transformation. Its domain consists of $C^1([a, b])$, and its target is $C([a, b])$. The transformation sends p to $Dp + cp$. Show that it is surjective. Find its null-space.

12. Consider the following linear transformation. Its domain consists of the functions p in $C^1([a, b])$ with $p(a) = 0$ and $p(b) = 0$. Its target is $C([a, b])$. The transformation sends p to $Dp + cp$. Show that it is injective. Find its range.

Chapter 7

Linear transformations and matrices

7.1 From vector to coordinates

Let V be a vector space. Let $\mathbf{u}_1, \ldots, \mathbf{u}_n$ be a list of vectors in V. The corresponding linear transformation $L : \mathbf{R}^n \to V$ is defined by

$$Lc = \sum_{j=1}^{n} c_j \mathbf{u}_j. \tag{7.1}$$

This transformation associates to each column vector a corresponding vector. It is tempting to write this as $L = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$ as if the vectors were the column vectors of a matrix. This transformation could be called the *coordinate to vector transformation*.

Theorem 7.1 (Linear independence-injective correspondence) *Consider a list of n vectors in V. Then they are linearly independent if and only if the corresponding linear transformation $L : \mathbf{R}^n \to V$ is injective.*

Theorem 7.2 (Span-surjective correspondence) *Consider a list of n vectors in V. Then they span V if and only if the corresponding linear transformation $L : \mathbf{R}^n \to V$ is surjective.*

Theorem 7.3 (Basis-isomorphism correspondence) *Consider a list of n vectors in V. Then they are a basis for V if and only if the corresponding linear transformation $L : \mathbf{R}^n \to V$ is an isomorphism (in particular bijective).*

In the case when L represents a basis, there is an inverse transformation $L^{-1} : V \to \mathbf{R}^n$. It takes a vector $\mathbf{v} = \sum_{j=1}^{n} c_j \mathbf{u}_j$ and sends it to the coordinate column vector with components c_j. This could be called the *vector to coordinate transformation*. Some texts introduce a special notation for this: the column vector $L^{-1}\mathbf{v}$ is called $[\mathbf{v}]_L$.

7.2 From linear transformation to matrix

A matrix A defines a linear transformation from $\mathbf{R}^n$ to $\mathbf{R}^n$ that sends each column vector $\mathbf{x}$ to the matrix product $A\mathbf{x}$. Conversely, a linear transformation from $\mathbf{R}^n$ to $\mathbf{R}^n$ defines a unique matrix that gives the linear transformation in this way. This is called the *standard matrix* of the linear transformation. We shall see later that this is the matrix of the linear transformation with respect to the standard basis (the basis consisting of the columns of the identity matrix).

Say that $f : V \to V$ is a linear transformation from a vector space to the same vector space. Let $\mathbf{u}_1, \ldots, \mathbf{u}_n$ be a basis for V. Let L be the corresponding linear transformation. Then the matrix of f with respect to this basis is the standard matrix of

$$A = L^{-1} f L : \mathbf{R}^n \to \mathbf{R}^n. \tag{7.2}$$

Thus

$$\sum_{i=1}^{n} (Ac)_i \mathbf{u}_i = f(\sum_{j=1}^{n} c_j \mathbf{u}_j). \tag{7.3}$$

Theorem 7.4 (Linear transformation-matrix correspondence) *Let $f : V \to V$ be a linear transformation. Suppose that $\mathbf{u}_1, \ldots, \mathbf{u}_n$ is a basis for V. Let A be the matrix of L with respect to this basis. Then A may be computed directly from the action of f on basis vectors by expanding*

$$f(\mathbf{u}_j) = \sum_{i=1}^{n} A_{ij} \mathbf{u}_i. \tag{7.4}$$

In other words, the jth column of A is the coordinate vector of $f(\mathbf{u}_j)$ with respect to the basis.

The proof is a computation. By definition and the linearity of f we have

$$\sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} c_j \mathbf{u}_i = \sum_{j=1}^{n} c_j f(\mathbf{u}_j). \tag{7.5}$$

The only way this can happen is that for each j

$$\sum_{i=1}^{n} A_{ij} \mathbf{u}_i = f(\mathbf{u}_j). \tag{7.6}$$

Example. Consider the 2-dimensional vector space of functions with basis $\cos(\theta), \sin(\theta)$. Let $f = d/d\theta$. The matrix of f with respect to this basis is obtained by finding the matrix of the linear transformation that sends the column vector $[c_1\ c_2]'$ to $c_1 \cos(\theta) + c_2 \sin(\theta)$ to $c_2 \cos(\theta) - c_1 \sin(\theta)$ to the column vector $[c_2\ -c1]'$. This matrix is a rotation by $-\pi/2$.

The most remarkable fact about this correspondence is that composition of linear transformations gives rise to multiplication of matrices. Say that $g : V \to V$ is another linear transformation. Then $B = L^{-1} g L$ is the matrix of g with respect to the basis for V. The composition $g \circ f : V \to V$ then has a matrix $L^{-1} g \circ f L = L^{-1} g L L^{-1} f L = BA$.

Some texts use a notation like $[f]_L$ for the matrix of f with respect to the basis L. Then this fact about composition could be written $[f \circ g]_L = [f]_L [g]_L$, where the operation on the right is matrix multiplication. Of course this is already obvious from $L^{-1}(f \circ g)L = L^{-1} f \circ L L^{-1} \circ g L = (L^{-1} f L)(L^{-1} g L)$.

Say that $V = \mathbf{R}^n$ consists of column vectors, and $f : V \to V$ is multiplication by an n by n matrix F. Then these ideas still apply. If the basis is the *standard basis* given by the columns of the identity matrix, then the matrix of F with respect to this basis is F itself. However if we take some other basis L, then F is represented by the matrix $A = L^{-1} F L$.

7.3 Similar matrices

Say that $L : \mathbf{R}^n \to V$ and $\tilde{L} : \mathbf{R}^n \to V$ are two bases for V. Thus $L^{-1} : V \to \mathbf{R}^n$ and $\tilde{L}^{-1} : V \to \mathbf{R}^n$ are both coordinate mappings. The *coordinate transition matrix* from the $\tilde{L}$ coordinates to the L coordinates is the standard matrix of

$$Q = L^{-1} \tilde{L} : \mathbf{R}^n \to \mathbf{R}^n. \tag{7.7}$$

That is, the jth column of Q consists of the L coordinates of the jth vector in the basis $\tilde{L}$.

Theorem 7.5 (change of basis: similarity) *Let $f : V \to V$ be a linear transformation. Say that $L : \mathbf{R}^n \to V$ is the linear transformation associated with one basis, and $\tilde{L} : \mathbf{R}^n \to V$ is the linear transformation associated with another basis. Let $A = L^{-1} f L$ be the matrix of f with respect to L, and let $\tilde{A} = \tilde{L}^{-1} f \tilde{L}$ be the matrix of f with respect to $\tilde{L}$. Let Q be the coordinate transition matrix from $\tilde{L}$ to L. Then*

$$\tilde{A} = Q^{-1} A Q. \tag{7.8}$$

The proof is just

$$\tilde{A} = \tilde{L}^{-1} f \tilde{L} = \tilde{L}^{-1} L L^{-1} f L L^{-1} \tilde{L} = Q^{-1} A Q. \tag{7.9}$$

When two matrices $A, \tilde{A}$ are related by $\tilde{A} = Q^{-1} A Q$, then they are said to be *similar*. The theorem says that if two matrices represent the same linear transformation from a vector space to itself, then the two matrices are similar. This is an exciting concept which will eventually lead to the important concept of eigenvalue.

In the notation used by some texts, $[\mathbf{v}]_L$ is the coordinate representation of the vector $\mathbf{v}$ with respect to the basis L. Also $[f]_L$ is the matrix representation of f with respect to the basis L. If Q is the coordinate transition matrix from basis $\tilde{L}$ to basis L, then for every vector $\mathbf{v}$ we have $Q[\mathbf{v}]_{\tilde{L}} = [\mathbf{v}]_L$. The similarity relation is expressed by $[f]_{\tilde{Q}} = Q^{-1}[f]_L Q$. Of course both these relations are obvious: the first from $(L^{-1}\tilde{L})\tilde{L}^{-1}\mathbf{v} = L^{-1}\mathbf{v}$, the second from $\tilde{L}^{-1} f \tilde{L} = (\tilde{L}^{-1} L)(L^{-1} f L)(L^{-1} \tilde{L})$.

7.4 Appendix: The case of two vector spaces

Say that $f : U \to V$ is a linear transformation from a vector space U to the vector space V. Let $K : \mathbf{R}^n \to U$ represent a basis for U, and let $L : \mathbf{R}^m \to V$ represent a basis for V. Then the matrix of f with respect to these bases is the matrix of

$$A = L^{-1} f K : \mathbf{R}^n \to \mathbf{R}^m. \tag{7.10}$$

Say that $g : V \to W$ is another linear transformation. Let $M : \mathbf{R}^p \to W$ represent a basis for W. Then $B = M^{-1} g L$ is the matrix of g with respect to the bases for V and W. The composition $g \circ f : U \to W$ then has a matrix $M^{-1} g \circ f K = M^{-1} g L L^{-1} f K = BA$. The composition of linear transformations corresponds to the product of matrices.

Theorem 7.6 (change of basis: equivalence) *Let $f : U \to V$ be a linear transformation. Say that $K : \mathbf{R}^n \to U$ is the linear transformation associated with one basis, and $\tilde{K} : \mathbf{R}^n \to U$ is the linear transformation associated with another basis. Say that $L : \mathbf{R}^m \to V$ is the linear transformation associated with one basis, and $\tilde{L} : \mathbf{R}^m \to V$ is the linear transformation associated with another basis. Let $A = L^{-1} f K$ be the matrix of f with respect to K, L, and let $\tilde{A} = \tilde{L}^{-1} f \tilde{K}$ be the matrix of f with respect to $\tilde{K}, \tilde{L}$. Let the coordinate transition matrix from $\tilde{K}$ to K be the matrix P, and let the coordinate transition matrix from $\tilde{L}$ to L be the matrix Q. Then*

$$\tilde{A} = Q^{-1} A P. \tag{7.11}$$

The proof is just

$$\tilde{A} = \tilde{L}^{-1} f \tilde{K} = \tilde{L}^{-1} L L^{-1} f K K^{-1} \tilde{K} = Q^{-1} A P. \tag{7.12}$$

When two matrices $A, \tilde{A}$ are related by $\tilde{A} = Q^{-1} A P$, then they are said to be *equivalent*. The theorem says that if two matrices represent the same linear transformation from a vector space to another vector space, then the two matrices are equivalent. This is a boring concept; all that matters is the rank of the matrix.

7.5 The standard matrix

It is easy to get confused about linear transformations defined by matrices. If F is an n by n matrix, then it defines a linear transformation on the vector space $\mathbf{R}^n$ by $f(\mathbf{x}) = F\mathbf{x}$ (matrix multiplication). If L is an n by n matrix whose columns form a basis for $\mathbf{R}^n$, then the matrix of this linear transformation with respect to L is $A = L^{-1} F L$.

How do we get the original matrix? Take $L = I$ to be the matrix whose columns form the standard basis for $\mathbf{R}^n$. Then the matrix of the linear transformation with respect to this standard basis is the *standard matrix* F itself.

The same ideas apply to an m by n matrix F. Think of F as defining a linear transformation from $\mathbf{R}^n$ to $\mathbf{R}^m$. Let the columns of K form a basis for $\mathbf{R}^n$, and let the columns of L form a basis for $\mathbf{R}^m$. The matrix of the linear transformation with respect to these bases is $A = L^{-1}FK$.

If we take in particular the standard bases $K = I_n$ and $L = I_m$, then we get the matrix of the linear transformation to be the standard matrix F that we started with.

7.6 Problems

1. Consider the 5 dimensional space spanned by the functions 1, $\sin(\theta)$, $\cos(\theta)$, $\sin(2\theta)$, $\cos(2\theta)$. Find the matrix of $d/d\theta$ with respect to this basis. Hint: Each column is found by expressing the derivative of one of these functions as a linear combination of all five, and then extracting the coefficients.

2. Consider the 5 dimensional space spanned by the functions 1, $\sin(\theta)$, $\cos(\theta)$, $\sin^2(\theta)$, $\sin(\theta)\cos(\theta)$. Find the matrix of $d/d\theta$ with respect to this basis.

3. This problem refers to the previous two problems. Find the change of coordinates matrix from the coordinates given by the second basis to the coordinates given by the first basis. Check that this gives the correct relation between the two matrices of the linear transformation $d/d\theta$. In other words, show by explicit calculation that they are similar. Hint: Each column is found by taking one of the five functions in the second basis and expressing it as a linear combination of all five in the first basis, then extracting the coefficients.

4. An n by n matrix F can always be thought of as the standard matrix associated with a linear transformation. (That is, it is the matrix of the linear transformation with respect to the standard basis of $\mathbf{R}^n$.) However the linear transformation may have a particularly nice matrix with respect to some other basis than the standard basis. If the basis consists of the columns of L, then the matrix with respect to this basis is $A = L^{-1}FL$. Use this to find a nice matrix representation of the linear transformation with standard matrix
$$F = \begin{bmatrix} 2 & 1 & -2 \\ 2 & 3 & -4 \\ 1 & 1 & -1 \end{bmatrix}. \tag{7.13}$$
Hint: Take the basis to be the columns of
$$L = \begin{bmatrix} 1 & 2 & 1 \\ -1 & 0 & 2 \\ 0 & 1 & 1 \end{bmatrix}. \tag{7.14}$$

Chapter 8

Determinants

8.1 Permutations

A function $f : S \to S$ is called a (discrete time) *dynamical system*. An *orbit* of an element x of S consists of the sequence $x, f(x), f(f(x)), f(f(f(x))), \ldots$.

If S is finite, each orbit eventually enters a *cycle*, that is, it assumes a certain value and that value is repeated later, periodically.

As example, we can take $S = \{1, 2, 3, \ldots, n\}$. There are several ways of describing dynamical systems. One is to simply list the values of the function in order. Thus the function $3, 5, 5, 1, 3, 1$ is the function f with $f(1) = 3$, $f(2) = 5$, $f(3) = 5$, $f(4) = 1$, $f(5) = 3$, $f(6) = 1$.

Another way to describe a dynamical system is to describe its cycles and the way the function feeds into the cycles. The example above has the cycle $3, f(3) = 5$. Since $f(5) = 3$ this is a cycle of period 2. The element 2 feeds into the cycle at 5. The elements 6 and 4 both feed into 1, while 1 feeds into the cycle at 3.

If S is finite and $f : S \to S$ is a bijection, then f is called a *permutation*. Then every orbit is a cycle.

In studying permutations it is common to take $S = \{1, 2, 3, \ldots, n\}$. There are several ways of describing permutations. One is to simply list the values of the function in order. Thus the permutation $5, 3, 6, 2, 1, 4$ is the function f with $f(1) = 5$, $f(2) = 3$, $f(3) = 6$, $f(4) = 2$, $f(5) = 1$, $f(6) = 4$.

Another way is to describe a permutation is to describe its cycles. The permutation in the example above has the cycle $1, f(1) = 5, f(5) = 1$ and the cycle $f(2) = 3$, $f(3) = 6$, $f(6) = 4$, and $f(4) = 2$. An abbreviated notation for this is to say that the cycles are $(1, 5)$ and $(2, 3, 6, 4)$. (In this example it would be equivalent to say that the cycles are $(6, 4, 2, 3)$ and $(5, 1)$.)

Example: Take $S = \{1, 2, 3\}$. The six permutations may be listed as sequences $1, 2, 3$ and $2, 3, 1$ and $3, 1, 2$ and $1, 3, 2$ and $3, 2, 1$ and $2, 1, 3$.

Example: Take $S = \{1, 2, 3\}$. The six permutations may be listed as cycles. They are $(1)(2)(3)$ and (123) and (132) and $(1)(23)$ and $(2)(13)$ and $(12)(3)$. It is often convenient to make the writing shorter by leaving out the one-cycles. With this convention one would write the six permutations as ( ) and (123) and (132) and (23) and (13) and (12).

A two-cycle is called a *transposition*. It just interchanges two thing and leaves everything else alone. An $k+1$-cycle is may be obtained by successively applying k transpositions. For instance, the cycle (12345) is obtained by first applying (12), then (23), then (34) and finally (45). This is written (12345) = (45)(34)(23)(12).

A permutation is said to be *even* if it may be written by applying an even number of transpositions. Otherwise it is *odd*. Thus an $k+1$ cycle is even if k is even. A permutation with any number of even cycles and with an even number of odd cycles is even. The only way a permutation can be odd is to have an odd number of odd cycles.

Example: Take $S = \{1, 2, 3, 4\}$. The twelve even permutations are listed as cycles. They are ( ) and (123) and (213) and (124) and (214) and (134) and (143) and (234) and (243) and (12)(34) and (13)(24) and (14)(23). The twelve odd permutations are (12) and (13) and (14) and (23) and (24) and (34) and (1234) and (1243) and (1324) and (1342) and (1423) and (1432).

If the set S has n elements, then there are $n!$ permutations in all. It is remarkable fact that half of the permutations are even and half of the permutations are odd.

8.2 The determinant of a matrix

If A is an n by n matrix, then the *determinant* of A is a sum over all $n!$ permutations of certain signed products. More precisely,

$$\det A = \sum_{\sigma} (-1)^{\sigma} \prod_{j=1}^{n} a_{\sigma(j),j}. \tag{8.1}$$

For a 2 by 2 matrix there are only two permutations. So

$$\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{21}a_{12}. \tag{8.2}$$

For a 3 by 3 matrix there are six permutations. So

$$\det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = a_{11}a_{22}a_{33}+a_{21}a_{32}a_{13}+a_{31}a_{12}a_{23}-a_{11}a_{32}a_{23}-a_{31}a_{22}a_{13}-a_{21}a_{12}a_{33}. \tag{8.3}$$

The three positive terms correspond to the identity ( ) and the two 3-cycles (123) and (132). The three negative terms correspond to the three 2-cycles $(23), (13), (12)$.

Three fundamental properties of the determinant are that it is multi-linear in the columns, alternating in the columns, and the determinant of I is 1.

Theorem 8.1 (product property of determinants) *Let A, B be n by n matrices. Then*

$$\det(AB) = \det(A)\det(B). \tag{8.4}$$

This theorem may be proved by a calculation. Let $C = AB$. Since $c_{ik} = \sum_j a_{ij}b_{jk}$, it follows that the kth column of C is

$$\mathbf{c}_k = \sum_{j=1}^{n} b_{jk}\mathbf{a}_j. \tag{8.5}$$

By multi-linearity

$$\det C = \det[\mathbf{c}_1, \ldots, \mathbf{c}_n] = \sum_{j_1=1}^{n} \ldots \sum_{j_n=1}^{n} bj_1\, 1 \cdots b_{j_n\, n} \det[\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_n}]. \tag{8.6}$$

However $\det[\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_n}] = 0$ whenever two columns are the same. So the only contributions to the sum are from bijections. Thus the equation may be written

$$\det C = \sum_{\sigma} b_{\sigma(1)1} \cdots b_{\sigma(n)n} \det[\mathbf{a}_{\sigma(1)}, \ldots, \mathbf{a}_{\sigma(n)}]. \tag{8.7}$$

By the alternating property, we can bring the columns of the matrix to their standard order at the price of introducing a sign. Thus

$$\det C = \sum_{\sigma} (-1)^{\sigma} b_{\sigma(1)1} \cdots b_{\sigma(n)n} \det[\mathbf{a}_1, \ldots, \mathbf{a}_n]. \tag{8.8}$$

This immediately gives the conclusion.

From this it is easy to see that the determinant of an inverse matrix is the reciprocal of the determinant. In particular, if the matrix has an inverse, then the determinant is non-zero.

This theorem gives a practical way of calculating determinants. Write $EA = H$, where E is a product of elementary matrices. Then $A = FH$, where F is a product of elementary matrices (the inverses in opposite order). The determinants of the elementary matrices are easy to compute. The determinant of a reflection across a diagonal is -1. The determinant of a multiplication of a coordinate by $s \neq 0$ is s. The determinant of a shear is 1. (And the determinants of their inverses are -1, $1/s$ and 1.) Finally, the determinant of the Hermite matrix H is 1 if $H = I$ and 0 otherwise.

In particular, this shows that if the determinant of A is not zero, then the matrix A has an inverse.

8.3 The determinant of a linear transformation

Let $f : V \to V$ be a linear transformation of a finite-dimensional vector space. Let $L : \mathbf{R}^n \to V$ determine a basis for V. Define $A = L^{-1}AL$ to be the matrix of f. Then $\det f$ is defined to be $\det A$.

One could worry that this depends on the choice of basis. However, if $\tilde{A} = \tilde{L}^{-1} f \tilde{L}$ is the matrix of f with respect to some other basis, then $\tilde{A} = Q^{-1}AQ$. So $\tilde{A}$ and A have the same determinant.

8.4 Problems

1. Find the cycle structure of the permutation that sends $1, 2, 3, 4, 5, 6, 7, 8, 9$ to $2, 7, 3, 5, 8, 9, 1, 6, 4$. Is it even or odd?

2. Consider the 3-dimensional vector space spanned by 1, $\cos(\theta)$, and $\sin(\theta)$. Let f be the linear transformation $d/d\theta + c$, where c is a constant. Find the determinant of f.

3. A group is an algebraic system with an identity that is also closed under multiplication and inverse. The group of all real n by n matrices that have inverses is called $GL(n, \mathbf{R})$ (the General Linear group). The subgroup of all real n by n matrices that have determinant equal to one is called $SL(n, \mathbf{R})$ (the Special Linear group). Prove that this is in fact a subgroup, that is, that the product of two matrices with determinant one is a matrix with determinant one, and the inverse of a matrix with determinant one is a matrix with determinant one.

4. The group of all real n by n orthogonal matrices is called $O(n)$ (the Orthogonal group). Show that every matrix in this group has determinant ± 1. (Each such matrix is a product of rotations and reflections.)

5. The group of all real n by n orthogonal matrices with determinant one is called $SO(n)$ (the Special Orthogonal group). Prove that $SO(n)$ is a subgroup of $O(n)$. (Each such matrix is a product of rotations and reflections, with an even number of reflections.)

Chapter 9

Eigenvalues

9.1 Eigenvalues and eigenvectors of matrices

Let $f : V \to V$ be a linear transformation. If there is a vector $\mathbf{v}$ in V that is not the zero vector, and if

$$f\mathbf{v} = \lambda\mathbf{v}, \tag{9.1}$$

then $\mathbf{v}$ is said to be an *eigenvector* of f with eigenvalue λ.

Theorem 9.1 (Characterization of eigenvalues) *A scalar λ is an eigenvalue of f if and only if the transformation $\lambda I - f$ is not invertible.*

Theorem 9.2 (Linear independence of eigenvectors) *Let f be a linear transformation from V to V. Let $\lambda_1, \ldots, \lambda_r$ be eigenvalues of f with corresponding eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_r$. If the eigenvalues $\lambda_1, \ldots, \lambda_r$ are all distinct, then the eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are linearly independent.*

The proof uses the fact that if $p(x)$ is a polynomial in x, then for each eigenvector

$$p(f)\mathbf{v} = p(\lambda)\mathbf{v}. \tag{9.2}$$

Suppose that the eigenvalues λ_i are all distinct. Let

$$c_1\mathbf{v}_1 + \cdots + c_j\mathbf{v}_j + \cdots + c_r\mathbf{v}_r = \mathbf{0}. \tag{9.3}$$

The goal is to show that all the coefficients $c_1, \ldots, c_j, \ldots, c_r$ are zero. This proves linear independence.

Fix j. The following argument will show that $c_j = 0$. Since j is arbitrary, this is all that is needed to prove linear independence.

Define a polynomial $p(x)$ by multiplying factors $(x - \lambda_i)$ for all i not equal to j. This is expressed in symbols by

$$p(x) = \prod_{i \neq j}(x - \lambda_i). \tag{9.4}$$

Then for each i not equal to j we have $p(\lambda_i) = 0$. Furthermore, since the eigenvalues are all distinct, we have $p(\lambda_j) \neq 0$.

Now apply the linear transformation $p(f)$. Since $p(f)$ is linear, we have

$$c_1 p(f)\mathbf{v}_1 + \cdots + c_j p(f)\mathbf{v}_j + \cdots + c_r p(f)\mathbf{v}_r = p(f)(c_1\mathbf{v}_1 + \cdots + c_j\mathbf{v}_j + \cdots + c_r\mathbf{v}_r) = p(f)\mathbf{0} = \mathbf{0}. \tag{9.5}$$

However $p(f)\mathbf{v}_i = p(\lambda_i)\mathbf{v}_i$. So this says that

$$c_1 p(\lambda_1)\mathbf{v}_1 + \cdots + c_j p(\lambda_j)\mathbf{v}_j + \cdots + c_r p(\lambda_r)\mathbf{v}_r = \mathbf{0}. \tag{9.6}$$

From the choice of the polynomial, this says that

$$c_j p(\lambda_j)\mathbf{v}_j = \mathbf{0}. \tag{9.7}$$

Since $\mathbf{v}_j \neq \mathbf{0}$, it follows that

$$c_j p(\lambda_j) = 0. \tag{9.8}$$

Since $p(\lambda_j) \neq 0$, it follows that $c_j = 0$. This completes the proof.

A linear transformation $f : V \to V$ is said to be *diagonalizable* if there is a basis of V such consisting of eigenvectors of f.

Theorem 9.3 (Diagonalization theorem) *Suppose that $f : V \to V$ is diagonalizable. Let $L : \mathbf{R}^n \to V$ be the linear transformation associated with the basis of eigenvalues. Then*

$$fL = LD, \tag{9.9}$$

where D is diagonal.

The proof is immediate. We have $f\mathbf{v}_j = \lambda_j\mathbf{v}_j$. Hence

$$f(\sum_{j=1}^{n} c_j\mathbf{v}_j) = \sum_{j=1}^{n} \lambda_j c_j \mathbf{v}_j. \tag{9.10}$$

This says that

$$fLc = LDc, \tag{9.11}$$

where $(D_c)_j = \lambda_j c_j$. In other words, $fL = LD$.

It is important to note that the diagonalization may be written in various forms, all of which say that f and D are similar. Thus it says that f may be transformed to a diagonal matrix:

$$D = L^{-1} f L. \tag{9.12}$$

Equivalently, it gives a representation of f as

$$f = LDL^{-1}. \tag{9.13}$$

This makes it clear that powers (iterates) of f may be computed by computing powers of the eigenvalues:

$$f^n = LD^nL^{-1}. \tag{9.14}$$

Let $f : V \to V$ be a linear transformation of a finite-dimensional vector space. The *characteristic polynomial* of f is the polynomial

$$p(\lambda) = \det(\lambda I - f). \tag{9.15}$$

The eigenvalues are the roots of the characteristic polynomial.

Theorem 9.4 (Distinct root criterion) *Let $f : V \to V$ be a linear transformation of a finite-dimensional vector space. If the characteristic polynomial of f has n distinct roots, then f is diagonalizable.*

Here is an example of how all this works. Say that F is a linear transformation given by an n by n matrix. The eigenvalues λ of F are the solutions of the polynomial equation

$$\det(\lambda I - F) = 0. \tag{9.16}$$

Say that this polynomial has n distinct roots $\lambda_1, \ldots, \lambda_n$. Then there are n independent column eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$. These form a basis with matrix $L = [\mathbf{u}_1, \ldots, \mathbf{u}_n$. Then

$$L^{-1}FL = D, \tag{9.17}$$

, where D is diagonal with $\lambda_1, \ldots, \lambda_n$ on the diagonal. In other words, F may be represented as

$$F = LDL^{-1}. \tag{9.18}$$

Take, for instance,

$$F = \begin{bmatrix} 4 & 2 \\ 3 & -1 \end{bmatrix}. \tag{9.19}$$

The characteristic polynomial is

$$\det(\lambda I - F) = \det \begin{bmatrix} \lambda - 4 & -2 \\ -3 & \lambda + 1 \end{bmatrix} = (\lambda - 4)(\lambda + 1) - 6 = \lambda^2 - 3\lambda - 10. \tag{9.20}$$

Since $\lambda^2 - 3\lambda - 10 = (\lambda - 5)(\lambda + 2)$, the roots are $\lambda_1 = 5$ and $\lambda_2 = -2$. The first eigenvector is obtained by finding a non-zero column vector in the null space of

$$5I - F = \begin{bmatrix} 1 & -2 \\ -3 & 6 \end{bmatrix}. \tag{9.21}$$

The second eigenvector is obtained by finding a non-zero column vector in the null space of

$$-2I - F = \begin{bmatrix} -6 & -2 \\ -3 & -1 \end{bmatrix}. \tag{9.22}$$

These two column vectors combine to form a basis consisting of the columns of the matrix

$$L = \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix}. \tag{9.23}$$

9.2 The trace

The *trace* of a square matrix is the sum of the diagonal entries. It is generally true that $\mathrm{tr}(AB) = \mathrm{tr}(BA)$. From this it follows that $\mathrm{tr}(Q^{-1}AQ) = \mathrm{tr}(AQQ^{-1}) = \mathrm{tr}(A)$. In other words, similar matrices have the same trace.

This gives a nice check on eigenvalue calculations. For example, in the last section there was a matrix F with trace $4 - 1 = 3$. It was similar to a diagonal matrix with diagonal entries $5, -2$. This matrix also has trace $5 - 2 = 3$. Maybe the calculation was correct!

9.3 Problems

1. Let
$$F = \begin{bmatrix} 5 & 6 \\ 3 & -2 \end{bmatrix}. \tag{9.24}$$
Find the eigenvalues of F.

2. In the preceding problem, find a basis consisting of eigenvectors of F.

3. In the preceding problem, let L be a matrix whose columns are the basis vectors. Compute $D = L^{-1}FL$.

4. In the preceding problem, find the matrix
$$G = \lim_{n\to\infty} \frac{1}{7^n} F^n. \tag{9.25}$$
Hint: Use $F = LDL^{-1}$.

Chapter 10

Inner product spaces

10.1 Inner products

An *inner product* is the same as a *scalar product* or *dot product*. It is a function whose inputs are ordered pairs of vectors in a vector space V and whose outputs are numbers. There are two notations in common use, the bracket notation and the dot notation:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v} \tag{10.1}$$

The axioms are

1. Symmetry: $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$.
2. Vector addition: $\mathbf{u} + \mathbf{v} \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$.
3. Scalar multiplication: $c\mathbf{u} \cdot \mathbf{v} = c\mathbf{u} \cdot \mathbf{v}$.
4. Positivity: $\mathbf{u} \cdot \mathbf{u} \geq 0$, and $\mathbf{u} \cdot \mathbf{u} = 0$ implies $\mathbf{u} = \mathbf{0}$.

The standard example is when the vector space is $\mathbf{R}^n$ and the inner product of the column vector $\mathbf{x}$ with the column vector $\mathbf{y}$ is $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}'\mathbf{y}$, where $\mathbf{x}'$ is the row vector corresponding to $\mathbf{x}$.

The *length* (or *norm*) of vector $\mathbf{u}$ is

$$\|\mathbf{u}\| = \sqrt{\mathbf{u} \cdot \mathbf{u}}. \tag{10.2}$$

A basic computation is

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + 2\mathbf{u} \cdot \mathbf{v} + \|\mathbf{v}\|^2. \tag{10.3}$$

A vector is a *unit vector* if its length is one. For unit vectors $\mathbf{u}, \mathbf{v}$ we have $0 \leq \|\mathbf{u} \mp \mathbf{v}\|^2 = 2 \mp 2\mathbf{u} \cdot \mathbf{v}$. Hence for unit vectors $\mathbf{u}, \mathbf{v}$ it follows that $\pm\mathbf{u} \cdot \mathbf{v} \leq 1$. This is a special case of the Cauchy-Schwarz inequality.

Theorem 10.1 (Cauchy-Schwarz inequality)

$$\pm\mathbf{u} \cdot \mathbf{v} \leq \|\mathbf{u}\|\|\mathbf{v}\|. \tag{10.4}$$

The proof of the Cauchy-Schwarz inequality is to notice that it is automatically true if either vector is the zero vector. Otherwise, $\mathbf{u}/\|\mathbf{u}\|$ and $\mathbf{v}/\|\mathbf{v}\|$ are unit vectors, and the previous special case applied to these unit vectors gives the result.

The inner product has a geometrical significance. In fact, we can write

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos(\theta), \tag{10.5}$$

where θ measures the angle between the two vectors. The Cauchy-Schwarz inequality guarantees that this makes sense, that is, that the cosine is between -1 and 1.

Two vectors $\mathbf{u}, \mathbf{v}$ are said to be *orthogonal* (or *perpendicular*) if $\mathbf{u} \cdot \mathbf{v} = 0$, and in this case we write $\mathbf{u} \perp \mathbf{v}$.

Theorem 10.2 (Theorem of Pythagoras) *If* $\mathbf{u} \perp \mathbf{v}$*, then*

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}^2 + \|\mathbf{v}\|^2. \tag{10.6}$$

10.2 Projections

Given vectors $\mathbf{u}_1, \ldots, \mathbf{u}_p$, their *Gram matrix* is the matrix of inner products

$$G_{jk} = \mathbf{u}_j \cdot \mathbf{u}_k. \tag{10.7}$$

The key property of the Gram matrix is that

$$\|c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p\|^2 = \sum_{i=1}^{p} \sum_{j=1}^{p} c_i G_{ij} c_j. \tag{10.8}$$

Theorem 10.3 (Gram matrix condition) *A list of vectors is linearly independent if and only if its Gram matrix is invertible.*

First we prove that invertibility of the Gram matrix implies linear independence. Suppose that $\sum_j c_j \mathbf{u}_j = 0$. Take the inner product with $\mathbf{u}_i$. This gives

$$\sum_{j=1}^{p} G_{ij} c_j = 0 \tag{10.9}$$

for each i. Since the matrix is invertible, it follows that the coefficients c_j are all zero. This proves linear independence.

Then we prove that linear independence implies that the Gram matrix is invertible. Consider a coordinate vector of c_j that is in the null space of the matrix G, that is, such that

$$\sum_{j=1}^{p} G_{ij} c_j = 0 \tag{10.10}$$

for each i. It follows that

$$\sum_{i=1}^{p} c_i \sum_{j=1}^{p} G_{ij} c_j = 0. \tag{10.11}$$

From the key property, it follows that $\sum_{j=1}^{p} c_j \mathbf{u}_j = \mathbf{0}$. By linear dependence the coefficients c_j are all zero. This proves that the null space of G is trivial. Therefore G is invertible.

If V is a vector space with an inner product, and W is a subspace of V, then the orthogonal projection of a vector $\mathbf{v}$ onto W is a vector $\mathbf{w}$ with the properties:

1. $\mathbf{w}$ is in W.
2. $\mathbf{v} - \mathbf{w}$ is orthogonal to W.

Theorem 10.4 (Orthogonal projection theorem) *Let $\mathbf{u}_1, \ldots, \mathbf{u}_p$ be linearly independent. Then the orthogonal projection onto the span of these vectors is the vector $\mathbf{w}$ given by*

$$\mathbf{w} = E\mathbf{v} = \sum_{j=1}^{p} \mathbf{u}_i \left(\sum_{j=1}^{p} G_{ij}^{-1} \mathbf{u}_j \cdot \mathbf{v} \right). \tag{10.12}$$

To see this, write

$$\mathbf{w} = \sum_{i=1}^{p} c_i \mathbf{u}_i. \tag{10.13}$$

In order to have $\mathbf{v} - \mathbf{w}$ orthogonal to each $\mathbf{u}_j$ it is enough to have

$$\mathbf{u}_j \cdot \mathbf{w} = \mathbf{u}_j \cdot \mathbf{v}. \tag{10.14}$$

This says that

$$\sum_{i=1}^{p} G_{ij} c_i = \mathbf{u}_j \cdot \mathbf{v}. \tag{10.15}$$

So to get c_i one just has to solve this equation involving the Gram matrix. The solution is given by the inverse to the Gram matrix

$$c_i = \sum_{j=1}^{p} G_{ij}^{-1} \mathbf{u}_j \cdot \mathbf{v}. \tag{10.16}$$

10.3 Projection matrices

In this section the same ideas are presented in matrix language. For each subspace W of $\mathbf{R}^m$ there is a *projection* matrix E. This is the matrix that defines the linear transformation of orthogonal projection onto W. It is characterized by the following properties:

1. For each $\mathbf{y}$ in $\mathbf{R}^m$ the projected vector $\hat{\mathbf{y}} = E\mathbf{y}$ is in W.
2. For each $\mathbf{y}$ in $\mathbf{R}^m$ the vector $\mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - E\mathbf{y}$ is perpendicular to W.

Thus $\mathbf{y} = \hat{\mathbf{y}} + (\mathbf{y} - \hat{\mathbf{y}})$ is the decomposition of $\mathbf{y}$ into the sum of a vector in W and a vector in $W^\perp$.

Theorem 10.5 (Orthogonal projection) *Let A be a m by p matrix with linearly independent columns. Let W be the column space of A, a p dimensional subspace of $\mathbf{R}^m$. Then the Gram matrix $A'A$ is an invertible p by p matrix, and the orthogonal projection onto W is given by the m by m matrix*

$$E = A(A'A)^{-1}A'. \tag{10.17}$$

Furthermore $E = E'$ and $E^2 = E$.

Proof:

The matrix $A'A$ is a p by p square matrix. If $\mathbf{x}$ is in the null space of this matrix, then $A'A\mathbf{x} = \mathbf{0}$. In particular, $\mathbf{x}'A'A\mathbf{x} = 0$. However this says that $(A\mathbf{x})'(A\mathbf{x}) = 0$. It follows that $A\mathbf{x} = \mathbf{0}$. Since the null space of A is the zero subspace, it follows that $\mathbf{x} = \mathbf{0}$. This shows that the null space of $A'A$ is the zero subspace. Since $A'A$ is square, it follows that $A'A$ is invertible.

The next two steps show that the formula for E gives the projection onto the column space of A.

1. Clearly $E\mathbf{y} = A\mathbf{x}$, where $\mathbf{x} = (A'A)^{-1}A'\mathbf{y}$. This shows that $E\mathbf{y}$ is in the column space of A.

2. Let $\mathbf{z} = A\mathbf{x}$ be in the column space of A. Then the inner product of $\mathbf{y} - E\mathbf{y}$ with $\mathbf{z}$ is $(\mathbf{y} - E\mathbf{y})'\mathbf{z} = \mathbf{y}^T A\mathbf{x} - \mathbf{y}'A(A'A)^{-1}A'A\mathbf{x} = \mathbf{y}'A\mathbf{x} - \mathbf{y}'A\mathbf{x} = \mathbf{0}$.

The proof that $E = E'$ and the proof that $E^2 = E$ are both simple computations.

10.4 Least squares

Let A be an m by p matrix (the design matrix), and let $\mathbf{y}$ be a vector in $\mathbf{R}^m$ (the observation vector). A least squares vector is a vector $\mathbf{x}$ in $\mathbf{R}^p$ (a parameter vector) such that the sum of squares $\|A\mathbf{x} - \mathbf{y}\|^2$ is minimal.

Theorem 10.6 (Least squares solution) *A vector $\mathbf{x}$ is a least squares vector if and only if it is a solution of the normal equations*

$$A'A\mathbf{x} = A'\mathbf{y}. \tag{10.18}$$

Proof: Let $\hat{\mathbf{y}}$ (the predicted vector) be the projection of $\mathbf{y}$ onto the column space of A. Then a least squares vector is a vector such that $A\mathbf{x} = \hat{\mathbf{y}}$.

The condition that $A\mathbf{x}$ is the projection onto the column space of A is that $A\mathbf{x}$ is in the column space of A and that $\mathbf{y} - A\mathbf{x}$ is orthogonal to the column space of A. The first condition is obviously satisfied. The second condition says that $(A\mathbf{z})'(\mathbf{y} - A\mathbf{x}) = 0$ for all $\mathbf{z}$ in $\mathbf{R}^p$. This is the same as requiring that $\mathbf{z}'A'(\mathbf{y} - A\mathbf{x}) = 0$ for all $\mathbf{z}$ in $\mathbf{R}^p$.

Here is a summary of these ideas. There is a m by n matrix A, the *parametrizing matrix*, with linearly independent columns that span a subspace of $\mathbf{R}^n$. The matrix $A'A$ is the *Gram matrix* and is invertible. The m by m matrix $E = A(A'A)^{-1}A'$ is the *orthogonal projection* onto the subspace. The m by n matrix $(A'A)^{-1}A'$ is sometimes called the *pseudo-inverse* of A.

Say that $\mathbf{y}$ is a *data vector* in $\mathbf{R}^m$. Then the *parameter vector* $\mathbf{x}$ is the *least squares solution* of the problem of minimizing the size of $A\mathbf{x}-\mathbf{y}$. The formula for $\mathbf{x}$ is given by the pseudo-inverse applied to the data vector, so $\mathbf{x} = (A'A)^{-1}A'\mathbf{y}$. The *fitted vector* $\hat{\mathbf{y}}$ is $E\mathbf{y} = A\mathbf{x}$. The *residual vector* is $\mathbf{y} - \hat{\mathbf{y}}$.

10.5 Euclidean geometry

In elementary geometry a Euclidean space is a space consisting of points, but there is no preferred origin. However, the notion of Euclidean space is closely related to that of vector space, provided that we realize that while that points p and q of Euclidean space are not vectors, their difference $p - q$ is a vector $\mathbf{v}$. Even though it does not make sense to add points in Euclidean space, it is possible to add a vector $\mathbf{v}$ to a point q and get another point p.

It is possible to have points p and q and other points r and s with $p-q = r-s$ being the same vector. Thus if $p - q = \mathbf{v}$, and $\mathbf{u} + q = s$, $\mathbf{u} + p = r$, then $\mathbf{v} + s = \mathbf{v} + \mathbf{u} + q = \mathbf{u} + \mathbf{v} + q = \mathbf{u} + p = r$, so also $r - s = \mathbf{v}$.

An *affine space* is a non-empty set P together with a finite-dimensional vector space V (the vector space of translations). There is also a function that associates to every p in P and $\mathbf{v}$ in V another point $\mathbf{v}+p$ in P. Thus the sum of a vector and a point is another point. It must satisfy the following properties:

1. Action of zero: $\mathbf{0} + p = p$.
2. Action of vector sum: $(\mathbf{v} + \mathbf{w}) + p = \mathbf{v} + (\mathbf{w} + p)$
3. For every two points q and p in P, there is exactly one vector $\mathbf{v}$ such that $\mathbf{v} + q = p$.

If P is an affine space, the unique vector from q to p is a vector in V, and this vector is denoted by $p - q$.

Thus the difference of two points is a vector. This operation satisfies the following properties:

1. $p - q = \mathbf{0}$ is equivalent to $p = q$.
2. $(p - q) + (q - r) = p - r$.
2. To each q in P and $\mathbf{v}$ in V there is exactly one p in P with $p - q = \mathbf{v}$.

Since the difference $p - q$ of two points p and q is a vector, for every real number t the scalar multiple $t(p - q)$ is a vector. Thus $t(p - q) + q$ is a point. Thus it is possible to define

$$tp + (1 - t)q = t(p - q) + q.$$

The line passing through p and q consists of all points $tp + (1 - t)q$ for t real. The segment between p and q consists of all points $tp + (1 - t)q$ for $0 \leq t \leq 1$.

Example: Let p, q, and r be three distinct points determining a triangle. The point $\frac{1}{2}p + \frac{1}{2}q$ is the midpoint of the segment from p to q. The point

$$\frac{2}{3}(\frac{1}{2}p + \frac{1}{2}q) + \frac{1}{3}r = \frac{1}{3}p + \frac{1}{3}q + \frac{1}{3}r$$

is on the segment from this midpoint to r. It is not hard to see that the three such lines, from the midpoints to the opposite corners, all meet in this point.

A *Euclidean space* is an affine space P for which the vector space of translations is an inner product space.

In an Euclidean space the distance between two points p and q is $|p - q|$, the length of the vector $p - q$.

Example: A right triangle is determined by three distinct points p, q, r such that the inner product of the vectors $p - q$ and $q - r$ is zero. Then $p - r = (p - q) + (q - r)$ as a vector sum, so when we compute the inner product of $p - r$ with itself the cross term drops out, and we get

$$|p - q|^2 + |q - r|^2 = |p - r|^2.$$

This is the theorem of Pythagoras.

10.6 Problems

1. Say that $\mathbf{u}_1, \ldots, \mathbf{u}_p$ are orthogonal, in the sense that for each $i \neq j$ we have $\mathbf{u}_i \cdot \mathbf{u}_j = 0$. When are they linear independent? Relate this to the invertibility of the Gram matrix.

2. Say that $\mathbf{u}_1, \ldots, \mathbf{u}_p$ are orthogonal and linearly independent. What is the projection of a vector $\mathbf{v}$ onto their span in this case? Give the explicit formula. Relate this to the formula in terms of the Gram matrix.

3. Let A be the 11 by 3 matrix whose ith row is $1, i, i^2$, for $i = 0, 1, 2, \ldots, 9, 10$. Find the Gram matrix, the pseudo-inverse matrix, and the projection matrix.

4. In the preceding problem, let the data vector $\mathbf{y} = [120\ 82\ 173\ 148\ 92\ 155\ 152\ 103\ 43\ 22\ 35]'$. Find the parameter vector $\mathbf{x}$. Find the fitted vector $\hat{\mathbf{y}}$. Verify the theorem of Pythagoras for the data vector, the fitted vector, and the residual vector $\mathbf{y} - \hat{\mathbf{y}}$.

Chapter 11

Self-adjoint transformations

11.1 The adjoint

Let $f : V \to W$ be a linear transformation from a real inner product to another. Then the *adjoint* f^* is the transformation from W to V that satisfies

$$f^*(\mathbf{w}) \cdot \mathbf{v} = \mathbf{w} \cdot f(\mathbf{v}) \tag{11.1}$$

for all $\mathbf{v}$ in V and $\mathbf{w}$ in W.

For real matrices the adjoint with respect to the standard inner product is the transpose.

11.2 Orthogonal transformations

Let $f : V \to W$ be a linear transformation from a real inner product space to another. Then f is said to be *inner product preserving* if $f(\mathbf{u}) \cdot f(\mathbf{v}) = \mathbf{u} \cdot \mathbf{v}$ for each $\mathbf{u}, \mathbf{v}$ in V. (Another name for such a transformation is *isometry.*) An inner product preserving transformation is automatically norm preserving, since $\|f(\mathbf{u})\|^2 = \|\mathbf{u}\|^2$. It follows that it is injective.

Theorem 11.1 (Inner product preservation) *A linear transformation $f : V \to W$ from a real inner product space to another is inner product preserving if and only if $f^* f = I$.*

A list of vectors $\mathbf{u}_j$ in an inner product space is an orthonormal family if

$$\mathbf{u}_j \cdot \mathbf{u}_k = \delta_{jk}. \tag{11.2}$$

Theorem 11.2 (Orthonormal families) *A linear transformation $f : \mathbf{R}^n \to W$ from $\mathbf{R}^n$ with the standard inner product to the real inner product space W is inner product preserving if and only if it is of the form*

$$f(c) = \sum_{j=1}^{n} c_j \mathbf{u}_j, \tag{11.3}$$

where the $\mathbf{u}_j$ *form an orthonormal family of vectors.*

Let $f : V \to W$ be a linear transformation from a real inner product space to another. Then f is said to be an *inner product isomorphism* if it is inner product preserving and is a bijection. (Another name for such a transformation is *orthogonal transformation.*)

Theorem 11.3 (Inner product preservation) *A linear transformation* $f : V \to W$ *from a real inner product space to another is an inner product isomorphism if and only if* $f^*f = I$ *and* $ff^* = I$.

Theorem 11.4 (Orthonormal bases) *A linear transformation* $f : \mathbf{R}^n \to W$ *from* $\mathbf{R}^n$ *with the standard inner product to the real inner product space* W *is an inner product isomorphism if and only if it is of the form*

$$f(c) = \sum_{j=1}^{n} c_j \mathbf{u}_j, \tag{11.4}$$

where the $\mathbf{u}_j$ *form an orthonormal basis.*

11.3 Self-adjoint transformations

Theorem 11.5 (Spectral theorem) *Let* $f : V \to V$ *be a self-adjoint linear transformation from an* n *dimensional real inner product space to itself. Then there exists an orthonormal basis* $\mathbf{u}_1, \ldots, \mathbf{u}_n$ *of* V *such that*

$$f\mathbf{u}_j = \lambda_j \mathbf{u}_j, \tag{11.5}$$

where the eigenvalues λ_j *are all real. Thus if we define the linear transformation* $U : \mathbf{R}^n \to V$ *by*

$$Uc = \sum_{j=1}^{n} c_j \mathbf{u}_j, \tag{11.6}$$

then U *is an inner product isomorphism, and*

$$fU = U\Lambda, \tag{11.7}$$

where Λ *is a real diagonal matrix.*

In the matrix case this says that if F is a real symmetric matrix, then there is a real orthogonal matrix U and a real diagonal matrix Λ such that $FU = U\Lambda$. In other words, every real symmetric matrix is similar to a real diagonal matrix.

The spectral theorem may also be thought of as a theorem about quadratic forms. If A is a real symmetric matrix, the its quadratic form is $\mathbf{x}'A\mathbf{x}$. The theorem says that there is a transformation $\mathbf{x} = U\mathbf{y}$ via an orthogonal matrix U such that

$$\mathbf{x}'A\mathbf{x} = \mathbf{y}'\Lambda\mathbf{y} = \lambda_1 y_1^2 + \cdots \lambda_n y_n^2 \tag{11.8}$$

is a sum of multiples of squares This works because $U^{-1} = U'$ and so $\mathbf{x}'A\mathbf{x} = \mathbf{y}'U'AU\mathbf{y} = \mathbf{y}'\Lambda\mathbf{y}$.

As an example of the material of this chapter, take the problem of finding the orthogonal diagonalization of the symmetric matrix

$$F = \begin{bmatrix} 2 & 2 & 4 \\ 2 & 5 & 8 \\ 4 & 8 & 17 \end{bmatrix}. \tag{11.9}$$

It is not too hard to compute its eigenvalues from the characteristic polynomial. But it is even easier to use a computer program to guess that they are 1, 1, 22. This means that we need to find the null spaces of

$$I - F = \begin{bmatrix} -1 & -2 & -4 \\ -2 & -4 & -8 \\ -4 & -8 & -16 \end{bmatrix} \tag{11.10}$$

and of

$$22I - F = \begin{bmatrix} 20 & -2 & -4 \\ -2 & 17 & -8 \\ -4 & -8 & 5 \end{bmatrix}. \tag{11.11}$$

A matrix whose columns span the null space of $I - F$ is

$$N_1 = \begin{bmatrix} -2 & -4 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{11.12}$$

In fact, there is really only one equation to be satisfied, namely $x_1 + 2x_2 + 4x_3 = 0$, and these two columns each satisfy this equation. However they are not orthogonal. One way to find two orthogonal vectors is to solve $x_1 + 2x_2 + 4x_3 = 0$ with the condition that the solution is orthogonal to $[-2, 1, 0]'$. This condition gives the second equation $-2x_1 + x_2 = 0$. This system of two equations in three unknowns has solution $[-4\ -8\ 5]'$ (or any multiple).

There is another way to get the same result. Project the second column $\mathbf{v}$ of N_1 onto the subspace spanned by first column $\mathbf{u}$. We get the projected vector $\mathbf{w} = \mathbf{u}(1/\mathbf{u}\cdot\mathbf{u})\mathbf{u}\cdot\mathbf{v}$. Then the difference $\mathbf{v} - \mathbf{w}$ (or any multiple of it) is a vector in the null space that is orthogonal to the first column of N_1.

With either argument we see that

$$\hat{N}_1 = \begin{bmatrix} -2 & -4 \\ 1 & -8 \\ 0 & 5 \end{bmatrix}. \tag{11.13}$$

has columns that give an orthogonal basis for the null space of $I - F$..

The null space of $22I - F$ is only one dimensional. It is spanned by

$$N_{22} = \frac{1}{4}\begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}. \tag{11.14}$$

Since $1 \neq 22$ are distinct eigenvalues, the column in N_{22} is automatically orthogonal to the columns in N_1.

To get unit vectors, we normalize each vector to have length one. This gives

$$U = \begin{bmatrix} -2/\sqrt{5} & -4/\sqrt{105} & 1/\sqrt{21} \\ 1/\sqrt{5} & -8/\sqrt{105} & 2/\sqrt{21} \\ 0 & 5/\sqrt{105} & 4/\sqrt{21} \end{bmatrix}. \tag{11.15}$$

Since U is orthogonal, we have $U^{-1} = U'$. Finally, we check that $U'FU = D$.

11.4 Problems

1. Let F be the symmetric matrix

$$F = \begin{bmatrix} 1 & -1 & 0 & 2 \\ -1 & 2 & 1 & 0 \\ 0 & 1 & 1 & 2 \\ 2 & 0 & 2 & -1 \end{bmatrix}. \tag{11.16}$$

   Find the eigenvalues. You can use the computer and then check that your answer for λ is correct by calculating the null space of $\lambda I - F$. Or you can struggle with the characteristic polynomial, which should also work, if you can guess a few roots.

2. In the previous problem, find a basis of eigenvectors (with rational entries).

3. In the previous problem, find an orthonormal basis of eigenvectors (square roots needed).

4. In the previous problem, find U orthogonal and D diagonal with $U^{-1}FU = D$.

Chapter 12

Multiplication of vectors

12.1 Dot product and cross product

In physics and engineering it is customary to multiply vectors in more than one way. The most common version in special to three dimensional space. This chapter gives a brief summary of this theory.

Let V be a 3 dimensional real vector space with an inner product. We will also assume that V has an orientation, so that we have a notion of when a basis is right-handed.

Traditionally there are two notions of multiplication. The *dot product* of $\mathbf{u}$ and $\mathbf{v}$ is

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\| \cos(\theta). \tag{12.1}$$

If $\mathbf{u}$ and $\mathbf{v}$ are neither the zero vector, then there is a uniquely defined value of θ with $0 \leq \theta \leq \pi$. Then $-1 \leq \cos(\theta) \leq 1$. When one of them is the zero vector, then the dot product is zero. The dot product measures the tendency of the two vectors to go in the same direction. For instance, if $\mathbf{u}$ is a unit vector, then the size of $\mathbf{u} \cdot \mathbf{v}$ depends on the size of the projection of $\mathbf{v}$ onto the space spanned by $\mathbf{u}$.

The *cross product* of $\mathbf{u}$ and $\mathbf{v}$ is a vector of length

$$\|\mathbf{u} \times \mathbf{v}\| = \|\mathbf{u}\|\|\mathbf{v}\| \sin(\theta). \tag{12.2}$$

Again $0 \leq \theta \leq \pi$, so $0 \leq \sin(\theta) \leq 1$. The vector $\mathbf{u} \times \mathbf{v}$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$. Furthermore, the triple $\mathbf{u}, \mathbf{v}, \mathbf{u} \times \mathbf{v}$ has a right-hand orientation. The cross product measures the tendency of the two vectors to go in different directions. If $\mathbf{u}, \mathbf{v}$ are regarded as defining a parallelogram, then the magnitude of the cross product is the area of this parallelogram.

The cross-product is anti-symmetric:

$$\mathbf{u} \times \mathbf{v} = -\mathbf{v} \times \mathbf{u}. \tag{12.3}$$

It is linear in each variable, for instance

$$(\mathbf{u} + \mathbf{v}) \times \mathbf{w} = \mathbf{u} \times \mathbf{w} + \mathbf{v} \times \mathbf{w} \tag{12.4}$$

and

$$(a\mathbf{u}) \times \mathbf{v} = a(\mathbf{u} \times \mathbf{v}). \tag{12.5}$$

It also distributes over vector addition:

$$\mathbf{u} \times (\mathbf{v} + \mathbf{w}) = \mathbf{u} \times \mathbf{v} + \mathbf{u} \times \mathbf{w}. \tag{12.6}$$

The unpleasant thing about the cross product is that it is not associative. On the one hand, the *triple vector product*

$$\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{u} \cdot \mathbf{v})\mathbf{w} \tag{12.7}$$

lies in the space spanned by $\mathbf{v}$ and $\mathbf{w}$. On the other hand,

$$(\mathbf{u} \times \mathbf{v}) \times \mathbf{w} = -\mathbf{w} \times (\mathbf{u} \times \mathbf{v}) = -(\mathbf{w} \cdot \mathbf{v})\mathbf{u} + (\mathbf{w} \cdot \mathbf{u})\mathbf{v}. \tag{12.8}$$

lies in the space spanned by $\mathbf{u}$ and $\mathbf{v}$.

Another interesting quantity is the *triple scalar product*

$$[\mathbf{u}, \mathbf{v}, \mathbf{w}] = \mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}). \tag{12.9}$$

This is like a determinant, in that every time one interchanges two vectors, the sign changes. Its magnitude represents the volume of the parallelepiped spanned by the three vectors.

Sometimes a distinction is made between scalars and vectors, on the one hand, and pseudo-scalars and pseudo-vectors, on the other hand. A pseudo-scalar or a pseudo-vector is like an ordinary scalar or vector, except that its sign depends on whether a right hand rule or a left hand rule is used.

With this distinction, the cross product of a vector and another vector is a pseudo-vector. The triple scalar of three vectors is a pseudo-scalar. However the vector triple product of three vectors is an ordinary vector. Notice that the cross product of two pseudo-vectors is another pseudo-vector.

In elementary vector algebra an orientation is fixed (usually right-handed), and there is only one kind of scalar and vector. More sophisticated treatments do not require a fixed orientation, but then the objects can have properties that are orientation-dependent. Thus there are are scalars and vectors (which do not depend on the orientation) and also pseudo-scalars and pseudo-vectors (which change sign when the orientation is reversed), and they are different objects.

Often vector algebra in three dimensions is described via a right-handed orthonormal basis $\mathbf{i}, \mathbf{j}, \mathbf{k}$. Then the cross product is defined by the equations

$$\begin{aligned}
\mathbf{i} \times \mathbf{i} &= \mathbf{0} \\
\mathbf{j} \times \mathbf{j} &= \mathbf{0} \\
\mathbf{k} \times \mathbf{k} &= \mathbf{0} \\
\mathbf{i} \times \mathbf{j} &= \mathbf{k} \\
\mathbf{j} \times \mathbf{k} &= \mathbf{i} \\
\mathbf{k} \times \mathbf{i} &= \mathbf{j} \\
\mathbf{j} \times \mathbf{i} &= -\mathbf{k} \\
\mathbf{k} \times \mathbf{j} &= -\mathbf{i} \\
\mathbf{i} \times \mathbf{k} &= -\mathbf{j}.
\end{aligned} \tag{12.10}$$

12.2 Quaternion product

Some light is cast on the cross product and dot product by combining them in a single product. The idea is to have objects that consist of a pair, a scalar and a vector. Such an object is written $a + \mathbf{u}$ as if it were a sum. Two such objects are the same if they have the same scalar part and the same vector part. This mathematical construct is called a *quaternion*.

The sum of two quaternions is defined in the obvious way:

$$(a + \mathbf{u}) + (b + \mathbf{v}) = (a + b) + (\mathbf{u} + \mathbf{v}). \tag{12.11}$$

The product is much more interesting. It is defined by

$$(a + \mathbf{u})(b + \mathbf{v}) = (ab - \mathbf{u} \cdot \mathbf{v}) + (a\mathbf{v} + b\mathbf{u} + \mathbf{u} \times \mathbf{v}). \tag{12.12}$$

In particular, the quaternion product of two vectors $\mathbf{u}$ and $\mathbf{v}$ is

$$\mathbf{u}\mathbf{v} = -\mathbf{u} \cdot \mathbf{v} + \mathbf{u} \times \mathbf{v}. \tag{12.13}$$

That is, the quaternion product combines the dot product and the cross product in one multiplication operation.

One interesting identity is

$$(\mathbf{u} \cdot \mathbf{u})(\mathbf{v} \cdot \mathbf{v}) = (\mathbf{u} \cdot \mathbf{v})^2 + \det \begin{bmatrix} \mathbf{u} \cdot \mathbf{u} & \mathbf{u} \cdot \mathbf{v} \\ \mathbf{v} \cdot \mathbf{u} & \mathbf{v} \cdot \mathbf{v} \end{bmatrix}. \tag{12.14}$$

It says that the square of the length of the quaternion product is the square of the dot product plus the square of the length of the cross product. Notice that both terms on the right are positive. The two terms in the identity measure the tendency of the two vectors to be close to lying on a single line versus their tendency to point in different directions.

Since the dot product is symmetric and the cross product is skew-symmetric, it follows that the symmetric part of the quaternion product

$$\frac{1}{2}(\mathbf{u}\mathbf{v} + \mathbf{v}\mathbf{u}) = -\mathbf{u} \cdot \mathbf{v} \tag{12.15}$$

gives the negative of the dot product, while the anti-symmetric part of the quaternion product

$$\frac{1}{2}(\mathbf{u}\mathbf{v} - \mathbf{v}\mathbf{u}) = \mathbf{u} \times \mathbf{v} \tag{12.16}$$

gives the cross product.

For every quaternion $a+\mathbf{u}$ there is a *conjugate* quaternion $a-\mathbf{u}$. The product of a quaternion with its conjugate is

$$(a - \mathbf{u})(a + \mathbf{u}) = a^2 + \mathbf{u} \cdot \mathbf{u}. \tag{12.17}$$

This is the square of the length of the quaternion. This formula shows that every non-zero quaternion has a multiplicative inverse, namely,

$$\frac{1}{a + \mathbf{u}} = \frac{a - \mathbf{u}}{a^2 + \mathbf{u} \cdot \mathbf{u}}. \tag{12.18}$$

The best thing about the quaternion multiplication is that it is associative. For example, the product of three vectors is

$$\mathbf{u}(\mathbf{v}\mathbf{w}) = \mathbf{u}(-\mathbf{v}\cdot\mathbf{w}+\mathbf{v}\times\mathbf{w}) = \mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})-(\mathbf{v}\cdot\mathbf{w})\mathbf{u}+\mathbf{u}\times(\mathbf{v}\times\mathbf{w}) = \mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})-(\mathbf{v}\cdot\mathbf{w})\mathbf{u}+(\mathbf{u}\cdot\mathbf{w})\mathbf{v}-(\mathbf{u}\cdot\mathbf{v})\mathbf{w}. \tag{12.19}$$

The other product $(\mathbf{u}\mathbf{v})\mathbf{w}$ works out to be exactly the same. So, in summary, the quaternion product of three vectors is

$$\mathbf{u}\mathbf{v}\mathbf{w} = [\mathbf{u},\mathbf{v},\mathbf{w}] + \{\mathbf{u},\mathbf{v},\mathbf{w}\}. \tag{12.20}$$

The scalar part $[\mathbf{u},\mathbf{v},\mathbf{w}]$ is the triple scalar product. Here the vector part will be written $\{\mathbf{u},\mathbf{v},\mathbf{w}\}$ and called the *quaternion triple vector product*. This vector

$$\{\mathbf{u},\mathbf{v},\mathbf{w}\} = -(\mathbf{v}\cdot\mathbf{w})\mathbf{u} + (\mathbf{u}\cdot\mathbf{w})\mathbf{v} - (\mathbf{u}\cdot\mathbf{v})\mathbf{w} \tag{12.21}$$

is a linear combination of all three input vectors. It measures the tendency of the three vectors to lie in a plane. Thus if they are all three orthogonal, then the quaternion triple vector product is the zero vector. On the other hand, if $\mathbf{u}$ is a unit vector, then the quaternion triple vector product $\{\mathbf{u},\mathbf{v},\mathbf{u}\}$ is the reflection of $\mathbf{v}$ in the plane orthogonal to $\mathbf{u}$, and so it has the same length as $\mathbf{v}$.

Again there is an interesting identity. It says that

$$(\mathbf{u}\cdot\mathbf{u})(\mathbf{v}\cdot\mathbf{v})(\mathbf{w}\cdot\mathbf{w}) = \det\begin{bmatrix} \mathbf{u}\cdot\mathbf{u} & \mathbf{u}\cdot\mathbf{v} & \mathbf{u}\cdot\mathbf{w} \\ \mathbf{v}\cdot\mathbf{u} & \mathbf{v}\cdot\mathbf{v} & \mathbf{v}\cdot\mathbf{w} \\ \mathbf{w}\cdot\mathbf{u} & \mathbf{w}\cdot\mathbf{v} & \mathbf{w}\cdot\mathbf{w} \end{bmatrix} + \{\mathbf{u},\mathbf{v},\mathbf{w}\}\cdot\{\mathbf{u},\mathbf{v},\mathbf{w}\}. \tag{12.22}$$

Each term on the right is greater than or equal to zero. The two terms in the identity measure the tendency of the three vectors to point in rather different directions versus their tendency to be close to lying in a single plane. The last term has the explicit form

$$\{\mathbf{u},\mathbf{v},\mathbf{w}\}\cdot\{\mathbf{u},\mathbf{v},\mathbf{w}\} = (\mathbf{v}\cdot\mathbf{w})^2\mathbf{u}\cdot\mathbf{u}+(\mathbf{u}\cdot\mathbf{w})^2\mathbf{v}\cdot\mathbf{v}+(\mathbf{u}\cdot\mathbf{v})^2\mathbf{w}\cdot\mathbf{w}-2(\mathbf{u}\cdot\mathbf{v})(\mathbf{v}\cdot\mathbf{w})(\mathbf{w}\cdot\mathbf{u}). \tag{12.23}$$

The quaternion product is also sometimes described via an orthonormal basis $\mathbf{i},\mathbf{j},\mathbf{k}$ together with the number 1. Then the quaternion product is defined by the equations

$$\begin{aligned} \mathbf{i}\,\mathbf{i} &= -1 \\ \mathbf{j}\,\mathbf{j} &= -1 \\ \mathbf{k}\,\mathbf{k} &= -1 \\ \mathbf{i}\,\mathbf{j} &= \mathbf{k} \\ \mathbf{j}\,\mathbf{k} &= \mathbf{i} \\ \mathbf{k}\,\mathbf{i} &= \mathbf{j} \\ \mathbf{j}\,\mathbf{i} &= -\mathbf{k} \\ \mathbf{k}\,\mathbf{j} &= -\mathbf{i} \\ \mathbf{i}\,\mathbf{k} &= -\mathbf{j}. \end{aligned} \tag{12.24}$$

12.3 Quaternions and rotations

In this section we shall see that quaternions may be used to describe reflections and rotations in three dimensional space.

Theorem 12.1 (Quaternion reflection) *Let* $\mathbf{u}$ *be a unit vector in* $\mathbf{R}^3$*. Then* u *defines a linear transformation from* $\mathbf{R}^3$ *to* $\mathbf{R}^3$ *by the quaternion product*

$$\mathbf{r} \mapsto \mathbf{uru}. \tag{12.25}$$

This is a reflection that sends $\mathbf{u}$ *to* $-\mathbf{u}$*.*

To see this, take $\mathbf{r} = \mathbf{u}$. Then $\mathbf{uru} = \mathbf{uuu} = -\mathbf{u}$. On the other hand, take $\mathbf{r} \perp \mathbf{u}$. Then $\mathbf{ru} = -\mathbf{ur}$, so $\mathbf{uru} = -\mathbf{uur} = \mathbf{r}$.

What happens when there are two reflections, one after the other? Let $\mathbf{u}$ and $\mathbf{v}$ be two unit vectors in $\mathbf{R}^3$. Say that θ is the angle from $\mathbf{u}$ to $\mathbf{v}$. Then their quaternion product is

$$\mathbf{uv} = -\cos(\theta) + \sin(\theta)\mathbf{w}, \tag{12.26}$$

where $\mathbf{w}$ is a unit vector parallel to $\mathbf{u} \times \mathbf{v}$. Reflect along $\mathbf{u}$ and then along $\mathbf{v}$. The result is the linear transformation of $\mathbf{R}^3$ defined by

$$\mathbf{r} \mapsto \mathbf{vruv} = (-\cos(\theta) - \sin(\theta)\mathbf{w})\mathbf{r}(-\cos(\theta) + \sin(\theta)\mathbf{w}). \tag{12.27}$$

We shall now see that this is a rotation.

Theorem 12.2 (Quaternion rotation) *Let* $\mathbf{w}$ *be a unit vector and define unit quaternions* $\cos(\theta) \pm \sin(\theta)\mathbf{w}$*. The linear transformation of* $\mathbf{R}^3$ *defined by*

$$\mathbf{r} \mapsto (\cos(\theta) + \sin(\theta)\mathbf{w})\mathbf{r}(\cos(\theta) - \sin(\theta)\mathbf{w}). \tag{12.28}$$

is a rotation in the plane orthogonal to $\mathbf{w}$ *by an angle* 2θ*.*

To see that it defines a rotation, first note that if we take $\mathbf{r} = \mathbf{w}$, we get $\mathbf{vuwuv} = -\mathbf{vuuwv} = \mathbf{vwv} = -\mathbf{vvw} = \mathbf{w}$. So the axis of rotation $\mathbf{w}$ is unchanged. Now let $\mathbf{p}$ and $\mathbf{q}$ be orthogonal unit vectors that are each orthogonal to the unit vector $\mathbf{w}$ such that $\mathbf{p} \times \mathbf{q} = \mathbf{w}$. We have

$$\begin{aligned}(\cos(\theta) + \sin(\theta)\mathbf{w})\mathbf{p}(\cos(\theta) - \sin(\theta)\mathbf{w}) &= (\cos(\theta) + \sin(\theta)\mathbf{w})(\cos(\theta)\mathbf{p} + \sin(\theta)\mathbf{q}) \\ &= (\cos^2(\theta) - \sin^2(\theta))\mathbf{p} + 2\sin(\theta)\cos(\theta)\mathbf{q} \\ &= \cos(2\theta)\mathbf{p} + \sin(2\theta)\mathbf{q}. \end{aligned} \tag{12.29}$$

Similarly

$$\begin{aligned}(\cos(\theta) + \sin(\theta)\mathbf{w})\mathbf{q}(\cos(\theta) - \sin(\theta)\mathbf{w}) &= (\cos(\theta) + \sin(\theta)\mathbf{w})(\cos(\theta)\mathbf{q} - \sin(\theta)\mathbf{p}) \\ &= (\cos^2(\theta) - \sin^2(\theta))\mathbf{q} - 2\sin(\theta)\cos(\theta)\mathbf{p} \\ &= \cos(2\theta)\mathbf{q} - \sin(2\theta)\mathbf{p}. \end{aligned} \tag{12.30}$$

Thus in the plane orthogonal to the axis the transformation is

$$\begin{aligned} a\mathbf{p} + b\mathbf{q} \mapsto a(\cos(2\theta)\mathbf{p} + \sin(2\theta)\mathbf{q}) + b(\cos(2\theta)\mathbf{q} - \sin(2\theta)\mathbf{p}) &= \\ (\cos(2\theta)a - \sin(2\theta)b)\mathbf{p} + (\sin(2\theta)a + \cos(2\theta)b)\mathbf{q}. & \end{aligned} \tag{12.31}$$

This is a rotation by angle 2θ.

The conclusion is that the quaternion $-\cos(\theta) + \sin(\theta)\mathbf{w}$ defines a rotation about the $\mathbf{w}$ axis by an angle 2θ. This doubling of the angle is a deep and mysterious fact. Among other things, it fundamental to the quantum theory of electron spin.

The group of unit quaternions of the form $\cos(\theta) + \sin(\theta)\mathbf{w}$ is called Spin(3). The group of rotations in called SO(3). The correspondence from Spin(3) to SO(3) is two to one. Changing the sign of the quaternion gives the same rotation. This has a beautiful geometric explanation.

Each unit quaternion in Spin(3) is described by an angle θ with $0 \leq \theta \leq \pi$ and a vector $\mathbf{z}$ of length $\sin(\theta)$. The quaternion itself is $\cos(\theta) + \mathbf{z}$. The angle θ describes the angle of rotation in SO(3). The vector $\mathbf{z}$ describes the axis of rotation. If one replaces θ by $\pi - \theta$ and $\mathbf{z}$ by $-\mathbf{z}$ this changes the sign of the quaternion in Spin(3), but it gives the same rotation in SO(3).

One can think of the group Spin(3) as a three dimensional sphere sitting in the four dimensional space of all quaternions. The north pole of this sphere is at $\theta = 0$. For $0 < \theta < \pi$ the set of constant latitude is a 2-dimensional sphere consisting of all $\mathbf{z}$ of radius $\sin(\theta)$. The south pole of the sphere is at $\theta = \pi$.

The rotation group SO(3) consists of this sphere with opposite points identified. Or one may restrict the angle θ to be in the range $0 \leq \theta \leq \pi/2$. Even then, for $\theta = \pi/2$ the axes $\mathbf{z}$ and $-\mathbf{z}$ are unit vectors describing the same rotation by angle $\pi/2$ about this axis. The group SO(3) is not a sphere; it is a geometric object known as 3-dimensional projective space.

12.4 Clifford algebra

There is a generalization of these ideas to the case of an n dimensional real vector space V with an inner product. The vectors generate an algebra of dimension 2^n called the *Clifford algebra*. This algebra has a 2^{n-1} dimensional subalgebra called the *even Clifford algebra*.

In the case $n = 3$ the even Clifford algebra has dimension 4, and it is the algebra of quaternions. These should be thought of as the scalars and the pseudo-vectors. If $\mathbf{i}, \mathbf{j}, \mathbf{k}$ form a basis for the three-dimensional space of pseudo-vectors, then $1, \mathbf{i}, \mathbf{j}, \mathbf{k}$ form a basis for the even Clifford algebra, and the usual quaternion multiplication rule holds.

In the case $n = 3$ the full Clifford algebra has dimension 8. This should be thought of as the algebra consisting of the scalars and the pseudo-vectors (the even subalgebra), together with the vectors and the pseudo-scalars.

In the case $n = 2$ the full Clifford algebra has dimension 4. One can think of the plane as sitting inside a three dimensional space. The Clifford algebra consists of the scalars, the vectors in the plane, and the one-dimensional space of pseudo-vectors orthogonal to the plane. These last are the planar equivalent of pseudo-scalars. The even subalgebra consists of the scalars and these planar pseudo-scalars. While in the case $n = 3$ the even subalgebra is the quaternion algebra, in the case $n = 2$ it is a realization of the complex number algebra.

It may be helpful to think of the case $n = 2$ in terms of an orthonormal basis. If one takes $\mathbf{j}, \mathbf{k}$ as an orthonormal basis for the vectors in the plane, then $\mathbf{jk} = \mathbf{i}$ plays the role of the pseudo-scalar. The Clifford algebra is spanned by $1, \mathbf{j}, \mathbf{k}, \mathbf{i}$. The even-subalgebra is spanned by $1, \mathbf{i}$.

In the case $n = 1$ the full Clifford has dimension 2 (scalar and vector), and the even subalgebra consists of scalars.

There is charm to the fact that the three classic numbers systems $\mathbf{R}$, $\mathbf{C}$, and $\mathbf{H}$ (real numbers, complex numbers, Hamilton's quaternions) correspond to the even Clifford algebras in dimensions 1, 2, and 3. However this analogy breaks down in higher dimensions.