Part II: LINEAR ALGEBRA

Contents

1 Linear Equations

We all know how to solve the linear homogeneous equation in two real variables

$$3x + 4y = 0.$$

Each solution is a pair $(x, y) \in \mathbb{R} \times \mathbb{R}$ (for example $(-2, 3/2)$ is a solution). We also know that such pairs may be represented as 'vectors' in the Cartesian plane. The *set of all solutions* forms a straight line in the plane, passing through the origin $(0, 0)$. This means that if we add any two solutions together (using vector addition) we get another solution!

More generally, consider a *system* of m linear equations in n variables ($x_1, ..., x_n$, say), of the form

$$\sum_{i=1}^{n} a_{ji} x_i = 0 \qquad\qquad (j = 1, 2, ..., m) \tag{1}$$

where the a_{ji} are constants. Note that each *solution* of (1) is an n–tuple of numbers $(x_1, x_2, ..., x_n)$, or an 'n–component vector', and hence may be regarded as an element of $\mathbb{R}^n$. The set of solutions is therefore a subset of $\mathbb{R}^n$. As in our first example, we have a notion of 'addition' of solutions here. In this case 'addition' means the addition of real n–component vectors, i.e. objects of the form $x = (x_1, x_2, ..., x_n)$ with $x_i \in \mathbb{R}$. Namely, for $x = (x_1, x_2, ..., x_n)$, $y = (y_1, y_2, ..., y_n)$ define $x + y$ by

$$x + y = (x_1 + y_1, x_2 + y_2, ..., x_n + y_n).$$

Is it still true, as in the case $m = 1$, $n = 2$ above, that adding two solutions together gives us another solution?

To try to answer the question, suppose for example that $m = 1, n = 3$ in equation (1). Then the solution set, call it S, consists of triples (x_1, x_2, x_3), or (x, y, z) say, such that

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = 0.$$

Obviously we can't *solve* this without knowing the $\{a_{ij}\}$, but just suppose $v = (x, y, z)$ is one solution, and $v' = (x', y', z')$ another. What about $v + v'$? We require

$$a_{11}(x + x') + a_{12}(y + y') + a_{13}(z + z') = 0. \tag{2}$$

IS THIS TRUE? Well,...

$$(a_{11}x + a_{12}y + a_{13}z) + (a_{11}x' + a_{12}y' + a_{13}z') = 0 + 0 = 0 \tag{3}$$

so it IS true! In fact it is true, by a similar argument, for any m and n.
EXERCISE: Verify this by using the matrix notation for systems of simultaneous equations, and the rules of matrix algebra (from last year — see also section 2).

Since solving simultaneous linear equations is such a common problem in Mathematics it is a good idea to look at this generalised *linearity* property (cf. $m = 1, n = 2$) a bit more closely.

It has long been known how to add complex numbers together, and even how to add n–tuples of complex numbers together. This is a natural generalisation of the addition of real n–component vectors above. Recall from [Martin1998] (or elsewhere) that

Definition 1 (ABELIAN GROUP) *An* abelian group $(G, +)$ *is a non–empty set, G, with a closed associative commutative binary operation $+$, such that*
(1) there exists an element $0 \in G$, such that $0 + a = a + 0 = a$ for all $a \in G$;
(2) for each $a \in G$ there exists $(-a) \in G$ such that $a + (-a) = 0$.

Thus $(\mathbb{C}, +)$ is an abelian group, and so is $(\mathbb{C}^n, +)$ for any $n \in \mathbb{N}$ (in this case the 'zero element' is $0 = (0, 0, \cdots, 0)$).

For $x \in \mathbb{R}^n$ we define *scalar multiplication* of x by a real number α by

$$\alpha x = \alpha(x_1, x_2, ..., x_n) = (\alpha x_1, \alpha x_2, ..., \alpha x_n). \tag{4}$$

This generalises the familiar $n = 2$ case. For example

$$2 * (1, 3) = (2, 6). \tag{5}$$

Note that if x is a solution to (1) then so is αx for *any* α (exercise!). This suggests that some kind of 'scalar multiplication' is useful, together with addition, in solving linear equations.

Recall from [Martin1998] that

Definition 2 (FIELD) *A* field $(G, +, *)$ *is an abelian group $(G, +)$ together with a second closed associative commutative binary operation $*$ such that*
(1) $a * (b + c) = a * b + a * c$ for all $a, b, c \in G$;
(2) there exists $1 \in G$, $1 * a = a$ for all $a \in G$;
(3) for each $a \in G \setminus \{0\}$ there exists $a^{-1} \in G$ such that $a * a^{-1} = 1$ (that is to say, every non–zero element of G has an inverse).

Thus $(\mathbb{C}, +, \times)$ is a field, and so are $\mathbb{Q}$ and $\mathbb{R}$. However there is no suitable 'multiplication' operation to make $(\mathbb{C}^n, +)$ into a field. The scalar multiplication is not suitable here since it multiplies a scalar with a vector, rather than two vectors.

There *are* other sets we know of with two binary operations acting on them. For example the set $M_2(\mathbb{C})$ of complex 2×2 matrices — we 'know' how to add and multiply these. However $M_2(\mathbb{C})$ is not a field. Not every matrix $a \in M_2(\mathbb{C}) \setminus \{0\}$ has an inverse (here 0 denotes the zero matrix), so matrix multiplication does not fulfill the requirements for the second operation in a field.

Note that for any $m, n \in \mathbb{N}$ and field K the set $M_{mn}(K)$ of $m \times n$ matrices with entries in K forms an abelian group under matrix addition (for a review see §2). In the case $m = 1$ such a set $M_{1n}(K)$ reduces to n–component vectors. In all these cases we know how to add, but not multiply (i.e. we have an abelian group, but not a field). However, we DO know how to multiply a vector in $M_{12}(\mathbb{R})$ by a *scalar* (a real number) — e.g. $2 * (1,3) = (2,6)$ again. Indeed this works for multiplying $v \in M_{mn}(K)$ (any m, n, K) by a scalar (an element of K) — e.g. $a*(b,c) = (a*b, a*c)$, $a, b, c \in K$.

We know *scalar multiplication* is a useful idea wherever we use vectors (which is almost everywhere!). Thus it is natural to introduce the notion of VECTOR SPACE, which captures the two useful operations on vectors of *addition* and *scalar multiplication* in a formal way. Then we can recognise them wherever they occur.

1.1 Vector Spaces

Up to now we have seen *examples* of 'vectors'. The idea now is to capture the *essence* of vectors with an abstract definition:

Definition 3 (VECTOR SPACE) *A vector space V over a field K (e.g. $\mathbb{R}$ or $\mathbb{C}$) is an abelian group $(V, +)$ (with elements called 'vectors') with a further operation of* Multiplication *by an element of K (called a 'scalar') such that $\alpha \in K$ and $v \in V$ implies $\alpha v \in V$, and satisfying the following axioms:*
B_1: $\alpha \in K$ and $u, v \in V$ implies $\alpha(u + v) = \alpha u + \alpha v$;
B_2: for all $\alpha, \beta \in K$ and $v \in V$ then $(\alpha + \beta)v = \alpha v + \beta v$;
B_3: for all $\alpha, \beta \in K$ and $v \in V$ then $(\alpha\beta)v = \alpha(\beta v)$;
B_4: for all $v \in V$ if 1 is the multiplicative identity of K then $1v = v$.

Exercise 1 *Let $\underline{0}$ denote the zero vector of V. Using these axioms prove*

Theorem 1 *(i) for all $\alpha \in K$ then $\alpha\underline{0} = \underline{0}$;*
(ii) for all $v \in V$ and 0 the 'zero' of K then $0v = \underline{0}$;
(iii) $\alpha v = \underline{0}$ implies either $\alpha = 0$ or $v = \underline{0}$ (or both);
(iv) for all $\alpha \in K$ and $v \in V$ then $(-\alpha)v = -(\alpha v) = \alpha(-v)$.

Example 1 *Let V be the set of n-tuples of real numbers. Define a scalar multiplication with $K = \mathbb{R}$ by equation (4). This makes V a vector space* over $\mathbb{R}$ *(i.e. with $K = \mathbb{R}$).*

Exercise 2 *Verify the axioms in this case.*

Note that there is a geometrical interpretation of this vector space in the case $n = 2$ in terms of vectors in the Cartesian plane (as discussed in lectures). That is, the vector space of pairs of real numbers is $\mathbb{R}^2$. Similarly, for $n = 3$ the space is $\mathbb{R}^3$,.... and so on (note that for $n > 3$ the Cartesian coordinate version becomes increasingly difficult to picture!).

Example 2 *The set $\mathbb{C}^n$ may similarly form a vector space of n-tuples, this time of complex numbers. The natural field to choose is $\mathbb{C}$.*

Exercise 3 *Verify this example, and try to draw a geometric picture as above for the cases $n = 1$ and $n = 2$. What happens? What other choices of field would be possible?*

Example 3 *Let V be the set of all polynomials of degree n with real coefficients. This is a vector space under the usual addition and scalar multiplication. (Verify!).*

Example 4 *For any set X the set of functions*

$$f : X \to \mathbb{R}$$

is a vector space over $\mathbb{R}$, with addition

$$(f + g)(x) = f(x) + g(x)$$

and scalar multiplication

$$(\alpha f)(x) = \alpha f(x).$$

Definition 4 (SUBSPACE) *A subspace W of a vector space V is a subset of V which is itself a vector space with respect to the operations of V.*

If W is a subspace of V we write $W \leq V$.
Examples:
Note that the solutions of systems of linear *inhomogeneous* equations of the form

$$\sum_{i=1}^{n} a_{ji} x_i = b_j \qquad (j = 1, 2, ..., m) \tag{6}$$

may again be regarded as elements of $\mathbb{R}^n$. The solution set is again therefore a subset of $\mathbb{R}^n$. Now $\mathbb{R}^n$ may be equipped with the properties of a vector space, so is the solution set also a vector space, i.e. is it a subspace of $\mathbb{R}^n$?

Suppose again that $m = 1, n = 3$. Then the solution set, call it S, consists of triples (x_1, x_2, x_3), or (x, y, z) say, such that

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1.$$

Suppose $v = (x, y, z)$ is one solution, and $v' = (x', y', z')$ another. If S is a vector space it is closed under addition, that is $v + v' \in S$, in other words $(v + v')$ must be another solution. So for S to be a vector space

$$a_{11}(x + x') + a_{12}(y + y') + a_{13}(z + z') = b_1. \tag{7}$$

IS THIS TRUE? Well,...

$$a_{11}x + a_{12}y + a_{13}z + a_{11}x' + a_{12}y' + a_{13}z' = b_1 + b_1 = 2b_1 \tag{8}$$

so we require $2b_1 = b_1$ (comparing equation 7 with 8). In other words $S \leq \mathbb{R}^3$ only if $b_1 = 0$ (i.e. the equation is HOMOGENEOUS).

Exercise 4 *Generalize to all n.*

1.2 Linear Combinations and Linear (In)Dependence

Let V be a real (i.e. $K = \mathbb{R}$) vector space. Let $v_1, v_2, ..., v_n \in V$. (N.B. Here v_i is being defined to be vector, *not* the i^{th} component of a vector. You may prefer to write $\underline{v_i}$ to denote a vector in such cases. I will just be *careful* instead.) Then

$$\sum_{i=1}^{n} \alpha_i v_i \qquad (\alpha_i \in \mathbb{R})$$

is called a LINEAR COMBINATION of $v_1, v_2, ..., v_n$.

Definition 5 (LINEAR DEPENDENCE) *A finite set $\{v_i \mid i = 1, 2, ..., n\}$ is called linearly dependent if there exists a set $\{\alpha_i \mid i = 1, 2, ..., n\}$ of scalars ($\alpha_i \in \mathbb{R}$) NOT all zero, such that*

$$\sum_i \alpha_i v_i = \underline{0}.$$

Conversely, if no such set $\{\alpha_i\}$ exists then the set $\{v_i\}$ is LINEARLY INDEPENDENT.

Proposition 1 *The set $\{v_i\}$ is linearly independent iff*

$$\sum_i \alpha_i v_i = \underline{0} \quad \Rightarrow \quad \textit{all } \alpha_i = 0.$$

We will discuss an extended example in the lectures.

Definition 6 (SPANNING) *Let $\{v_i\}$ be a set of vectors in V and let W be the set of all linear combinations of the set $\{v_i\}$. Then W is a subspace of V, and W is called the subspace spanned by $\{v_i\}$ (or GENERATED by $\{v_i\}$).*

Geometrically: In $\mathbb{R}^3$ the space spanned by a single vector consists of all scalar multiples of it (hence a straight line through the origin). The space spanned by 2 linearly independent vectors in $\mathbb{R}^3$ is the PLANE through the origin containing them (exercise: prove that this plane is unique).

1.3 Basis

Definition 7 (BASIS) *A basis for a vector space V is a set of linearly independent vectors which spans V.*

It turns out that although a space V may have many different bases, if one of these has finite degree n, say, then they all have the same degree. The DIMENSION of a vector space is the number of vectors in a basis.

Theorem 2 *If $\{e_i\}$ is a basis for V then every vector $v \in V$ has a UNIQUE representation*

$$v = \sum_i \alpha_i e_i$$

(i.e. the scalars α_i, called coefficients or components of v w.r.t. this basis, are uniquely determined).

Proof: Let

$$v = \sum_i \alpha_i e_i = \sum_i \beta_i e_i$$

then

$$\sum_i (\alpha_i - \beta_i) e_i = \underline{0}.$$

But the $\{e_i\}$ are linearly independent, so by the definition of basis, for all i

$$\alpha_i - \beta_i = 0.$$

QED.

Conversely, the basis is not unique.

Examples: - see lectures.

In $\mathbb{R}^n$ the set of n vectors

$$e_1 = (1, 0, 0, ..., 0)$$
$$e_2 = (0, 1, 0, 0, ..., 0)$$

then similarly so that

$$e_i = (0, 0, 0, ..., 0, 1, 0, 0, ..., 0)$$

where the 1 is in the i^{th} position, and finally

$$e_n = (0, 0, 0, ..., 0, 0, 1)$$

is the STANDARD ORDERED BASIS.

Theorem 3 (Exchange Theorem) *Let $\{v_i \mid i = 1, 2, ..., n\}$ span vector space V, and $\{w_i \mid i = 1, 2, ..., m\}$ be a linearly independent set in V.*

Then $m \leq n$ and a set

$$\{w_1, w_2, ..., w_m, v_{i_1}, v_{i_2}, ..., v_{i_{n-m}}\}$$

(i.e. with last $n - m$ elements some subset of $\{v_i\}$) spans V.

Proof:

Let $S = \{w_m, v_1, v_2, ..., v_n\}$. Then S spans V and is linearly dependent. Therefore one of the v_i is a linear combination of the preceding vectors (as written in S above), that is there exists j such that

$$S' = \{w_m, v_1, v_2, ..., v_{j-1}, v_{j+1}, ..., v_n\}$$

spans V. Now let $S'' = \{w_{m-1}\} \cup S'$. Then S'' is linearly dependent and spans V. Thus one of its vectors is a linear combination of the preceding ones, when written in the order

$$S'' = \{w_{m-1}, w_m, v_1, v_2, ..., v_{j-1}, v_{j+1}, ..., v_n\}.$$

This vector cannot be w_m or w_{m-1} as these are linearly independent of each other by construction, so let it be v_k, say. Then $S'' - \{v_k\}$ spans V.

Now ITERATE this process until all the $\{w_i\}$ are included, and m of the $\{v_i\}$ are excluded. QED.

Corollary 3.1 *In a finite dim. vector space V the number of vectors in a basis is unique.*

Proof:

Let $\{v_i\}$ and $\{w_i\}$ be bases. Then $m \leq n$ and $n \leq m$ so $m = n$. QED.

Note that $Dim(\mathbb{R}^n) = n$. Also, more surprisingly, $Dim(\mathbb{C}^n) = n$ .... think about it!

Now let $Dim(V) = n$, then

Corollary 3.2 *Any set of $(n + 1)$ vectors in V is linearly dependent.*

Proof:

Let $S = \{v_i \mid i = 1, 2, ..., n\}$ span V, and let $S' = \{w_i \mid i = 1, 2, ..., n+1\}$ be linearly independent. Then $(n + 1) \leq n$ - a contradiction! QED.

Corollary 3.3 *If $W \leq V$ then $Dim(W) \leq Dim(V)$.*

Proof:

Let $Dim(W) = m$ and suppose $m > n$, then a basis for W is a set of more than n elements of V, but then they are linearly dependent by corollary 3.2 - a contradiction! QED.

Corollary 3.4 *Conversely, no set of $(n - 1)$ vectors can span V.*

Proof:

let $S = \{v_1, v_2, .., v_{n-1}\}$ span V and let $S' = \{w_1, w_2, ..., w_n\}$ be a basis for V. Then $(n-1) \geq n$ - a contradiction! QED.

Corollary 3.5 *n vectors of V form a basis iff linearly independent.*

Proof:

Let $S = \{v_1, v_2, .., v_n\}$ be a linearly independent set. Let $S' = \{w_1, w_2, ..., w_n\}$ be a basis. Then w_i's may be completely replaced by v_i's without affecting spanning, so S is a basis. QED.

Corollary 3.6 *If a set S of $m \geq n$ vectors spans V then there exists a subset of S which forms a basis for V.*

Proof:

Let $S = \{v_1, v_2, .., v_m\}$. Either S is a basis ($m = n$) or it is linearly dependent. In that case there exists a vector in S linearly dependent on the preceding ones. Discard it and iterate until linearly independent. QED.

Corollary 3.7 *Any l.i. set of vectors is part of a basis for V.*

Proof: Exercise!

Summary

1. Number of vectors in a basis is unique.

Let $Dim(V) = n$ then

2. A set of more than n vectors in V is necessarily linearly dependent.
3. Fewer than n vectors cannot span.
4. Any n linearly independent vectors form a basis.
5. Any set of $m \geq n$ vectors which spans CONTAINS a basis as a subset.
6. A l.i. set of $\leq n$ vectors can be extended to a basis by adding vectors.

We now want to study sets of vectors arising as solution sets for systems of linear equations.... So, what is the usual way of manipulating many vectors at once?

2 Matrices

A matrix is a rectangular array of scalars. For the case of m rows and n columns we call it an $m \times n$ matrix. If we think of a vector as a row of n numbers then such a matrix is a column of m vectors.

If A is a matrix then a_{ij} is the entry in the i^{th} row and j^{th} column. For example

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

is a 2×2 matrix.

2.1 Matrix Algebra

If $A = B$ then $a_{ij} = b_{ij}$ for all i, j.
If $A = 0$ then $a_{ij} = 0$ for all i, j.
ADDITION: If A, B both $m \times n$ then $C = A \pm B$ is given by $c_{ij} = a_{ij} \pm b_{ij}$.

Exercise 5 *The set $M_2(\mathbb{R})$ of 2×2 matrices with real entries forms an abelian group under addition. Check it!*
MULTIPLICATION: (1) by a scalar k: $C = kA$ is given by $c_{ij} = ka_{ij}$;
(2) of two matrices:
Matrices A, B are "conformable" for multiplication to give a product $C = AB$ (order matters!!) if the number of columns of A = no. of rows of B =p (say). Then $C = AB$ is given by

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

For example,

$$\begin{pmatrix} 2 & 4 & 6 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 5 \\ 2 & 4 \\ 6 & 0 \end{pmatrix} = \begin{pmatrix} 46 & 26 \\ 12 & 0 \end{pmatrix}.$$

Definition 8 (IDENTITY MATRIX) *A square (i.e. $n \times n$) matrix is called an identity matrix, denoted I_n, if*

$$I_n = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix}.$$

Exercise 6 *Check $AI_3 = I_3 A = A$ for all 3×3 matrices A.*

Exercise 7 *Check that no other matrix B obeys $AB = BA = A$ for all A.*

Definition 9 (INVERSE MATRIX) *An n square matrix B is called the inverse of A if $AB = I_n$.*

Proposition 2 (exercise) *If $AB = I_n$ then $BA = I_n$*

Hence we can write $B = A^{-1}$ or $A = B^{-1}$ if $AB = I_n$.

Proposition 3 *$AB = I_n$ and $AC = I_n$ implies $B = C$.*

Proof: Let $AB = I_n$ and $CA = I_n$. Then $(CA)B = C(AB)$ (exercise: prove matrix multiplication is associative!) so $I_n B = CI_n$ and so finally $B = C$. QED.

Proposition 4 (exercise)

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

Proposition 5 *$C = AB$ does not imply $C = BA$.*

Proof: By counterexample

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

but

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

Similarly, $AB = 0$ does not imply that either $A = 0$ or $B = 0$, and $AB = AC$ does not imply that $B = C$ (but $A0 = 0A = 0$ still holds).

Definition 10 (TRANSPOSE) *The transpose A^t of matrix A is obtained by interchanging rows and columns.*

e.g.

$$\begin{pmatrix} 2 & 4 & 6 \\ 0 & 0 & 2 \end{pmatrix}^t = \begin{pmatrix} 2 & 0 \\ 4 & 0 \\ 6 & 2 \end{pmatrix}.$$

Proposition 6

$$(AB)^t = B^t A^t.$$

Exercise 8 *Prove this for $n \times n$ matrices.*

Proposition 7 (exercise) *For A square, if A^{-1} exists then*

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

2.2 Matrices and Linear Equations

The system

$$\sum_{i=1}^{n} a_{ji} x_i = b_j \qquad (j = 1, 2, ..., m) \tag{9}$$

is equivalent to the matrix equation

$$A \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

or, with x denoting the transposed n-tuple, just

$$Ax = b.$$

Example 5 *The system*

$$x_1 + 2x_2 + 3x_3 = 1$$

$$x_1 + x_3 = 2$$

is equivalent to

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

The matrix $A = (a_{ij})$ is called the COEFFICIENT MATRIX of the system. The $m \times (n+1)$ matrix A' given by adding the column B to the right hand side of A is called the AUGMENTED MATRIX of the system. It contains all the information to completely define the system.

2.3 Elementary Row Operations

In fact it is usually simpler to use the language of matrices throughout when studying linear equations. Then the way we solve systems of linear equations is by using the fact that the solution set is unchanged under certain transformations of the system. For example:

changing the order in which we write the equations;
multiplying an equation through by an overall scalar constant; or
adding one equation to another,

does not affect the solution set (exercise: check it!).
These are the 'moves' we use to simplify the set of equations in the *Gauss elimination* method of solution. We use up some of the equations in eliminating some of the variables:

Example 6

$$3x + 2y + z = 4 \tag{10}$$

$$6x + 3y + 5z = 9 \tag{11}$$

Taking (11) - 2.(10), as it were, we get

$$-y + 3z = 1$$

so $y = 3z - 1$, $3x + 7z = 6$ *and thus* $z = (6 - 3x)/7$. *Finally, then, the solution set is the LINE of points in* $\mathbb{R}^3$ *given by*

$$\{(x, 3((6-3x)/7) - 1, (6-3x)/7) \text{ for all } x \in \mathbb{R}\}.$$

There must be a corresponding set of moves on the matrix side, therefore, which also do not change the solution.

Note that each linear equation is embodied in a particular ROW of A'. A matrix A is said to be ROW EQUIVALENT to a matrix B if A is mapped to B by a finite sequence of ELEMENTARY ROW OPERATIONS:

1. Interchanging i^{th} and j^{th} rows

$$R^{ij} : A \mapsto H^{ij} A$$

where H^{ij} is a square matrix with the same no. of rows as A, of the form

$$H^{ij} = \begin{pmatrix} 1 & & & & & & \\ & 1 & & & & & \\ & & 0 & & & 1 & \\ & & & 1 & & & \\ & & & & 1 & & \\ & & 1 & & & 0 & \\ & & & & & & 1 \end{pmatrix}.$$

2. Multiply i^{th} row by a scalar $k \neq 0$

$$R^{i}(k) : A \mapsto H^{i}(k) A$$

where

$$H^{i}(k) = \begin{pmatrix} 1 & & & & & & \\ & 1 & & & & & \\ & & 1 & & & & \\ & & & k & & & \\ & & & & 1 & & \\ & & & & & 1 & \\ & & & & & & 1 \end{pmatrix}.$$

3. Add j^{th} row to i^{th} row

$$P^{ij} : A \mapsto M^{ij} A$$

where

$$M^{ij} = \begin{pmatrix} 1 & & & & & & \\ & 1 & & & & & \\ & & 1 & & & 1 & \\ & & & 1 & & & \\ & & & & 1 & & \\ & & & & & 1 & \\ & & & & & & 1 \end{pmatrix}.$$

Exercise 9 *Check that these matrices act on an arbitrary matrix A in the way described.*

Now our Gaussian elimination process for solving systems of linear equations becomes "Transformation by elementary row operations to ECHELON FORM" where

Definition 11 (ECHELON FORM) *A matrix $A = (a_{ij})$ is in Echelon form if the number of zeros preceding the first non-zero entry in each row increases row by row until only zero rows remain.*

e.g.

$$A = \begin{pmatrix} 2 & 4 & 6 & 8 & 27 \\ 0 & 1 & 9 & 0 & 2 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

We call the first non-zero element of each row the DISTINGUISHED ELEMENT of that row.

Definition 12 *An echelon matrix is said to be ROW REDUCED if distinguished elements are*
(i) the only non-zero elements in their columns;
(ii) equal to 1.

e.g.

$$A = \begin{pmatrix} 1 & 0 & 6 & 0 & 27 \\ 0 & 1 & 9 & 0 & 2 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Theorem 4 *The row reduced echelon form (RREF) of a matrix is unique.*

Exercise 10 *Prove that a RREF always exists (hint: develop an algorithm to construct an echelon form using elementary row operations).*

We are trying to find the ESSENTIAL properties of systems of linear equations (resp. matrices) which DO have a bearing on their solutions. Clearly RREF is one of these. What else? ....

2.4 Row space of a matrix

Let A be an arbitrary $m \times n$ matrix (over $\mathbb{R}$). The rows of A, viewed as vectors in $\mathbb{R}^n$, span a subspace of $\mathbb{R}^n$. This is the ROW SPACE of A, $L(A)$.

Suppose we apply an elementary row operation to A and hence obtain B. Each row of B is a linear combination of rows of A so $L(B) \leq L(A)$. On the other hand if $GA = B$ is the matrix version of the elementary row operation then we can invert G (exercise) to obtain $G^{-1}B = G^{-1}GA = A$, so $L(A) \leq L(B)$ and altogether $L(A) = L(B)$. Thus row equivalent matrices have the same row space.

Definition 13 *$Dim(L(A))$ is called the RANK of A.*

Thus RANK is the number of independent rows - closely related to the solution space of any associated system of linear equations.

Example 7 *Suppose*

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 0 & -2 & 2 \\ 3 & 4 & 5 & 9 \end{pmatrix}$$

then

$$A \to \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -4 & -8 & -6 \\ 0 & -2 & -4 & -3 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -2 & -4 & -3 \\ 0 & -4 & -8 & -6 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -2 & -4 & -3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

then $rank(A) = 2$.

Hence:

Theorem 5 *The dimension of the solution space W of the homogeneous system*

$$AX = 0$$

is $n - rank(A)$, where n is the number of unknowns.

Theorem 6 *The system $AX = B$ has a solution iff $rank(A) = rank((A, B))$ (where (A, B) is the augmented matrix), whereupon, if v is a solution then*

$$\{u + v \;:\; u \in W\}$$

is the general solution, or solution set.

Proof: see later.

3 Linear Functions

We now provide one more way to look at systems of simultaneous linear equations (a way which will lead us into a number of other important applications of linear algebra).

Recall that for any two sets A, B a function $f : A \to B$ may be thought of as a map which assigns to each $a \in A$ a unique $f(a) \in B$ (e.g.

(1)

$$f : \mathbb{R} \to \mathbb{R}$$
$$f(x) = x^2$$

(2)

$$f : \mathbb{R}^3 \to \mathbb{R}^2$$
$$f(x, y, z) = (0, x + y)$$

— this function is illustrated in figure 1). Let V, U be vector spaces over $\mathbb{R}$ (say). A function $f : V \to U$ is called a linear function (or linear transformation, or linear mapping, or vector space homomorphism) if

(i) for all $v, w \in V$ we have $f(v + w) = f(v) + f(w)$;

(ii) for all $k \in \mathbb{R}$, $v \in V$ we have $f(kv) = kf(v)$.

That is, f is linear if it preserves the two basic operations $(+, .)$.
Examples: $f(x) = 3x$ is linear; (1) above is not linear; (2) is linear. (Exercise: Explicitly demonstrate the truth of these assertions.)

Exercise 11 *Verify that if f is linear then $f(0) = 0$.*
Answer: By linearity $f(0v) = 0f(v)$, but $0v = 0$. Note that this shows that $f(x) = x + 1$ is not linear by our definition, even though $y = x + 1$ is a linear equation. We will see later that this restriction is not as limiting as it may seem.

3.1 Isomorphism

If the linear function $f : V \to U$ is an isomorphism (one–to–one and onto) then V, U are *isomorphic* vector spaces, and we write $V \cong U$.

Exercise 12 *Compare $\mathbb{R}^3$ over $\mathbb{R}$ and the real vector space of polynomials of degree at most 2 with real coefficients. Construct a linear function from one to the other. Is your function an isomorphism?*

Exercise 13 *Prove that isomorphic vector spaces have the same dimension. (Hints: Ask yourself if the linear function in example (2) above is an isomorphism? Would it be possible to construct another linear function between $\mathbb{R}^3$ and $\mathbb{R}^2$ which was an isomorphism?)*

Exercise 14 *The set of all real polynomials in x form a real vector space, call it V, just as do the subset consisting of those polynomials of degree at most 2 (although V is not a finite dimensional space). Let $f \in V$, and define $D(f) = \frac{df}{dx}$. Show that D is a linear map from V to itself. Show that D is not one–to–one.*

Let $f : V \to U$ be a linear function, then

Definition 14 *The Image of f is*

$$Im\ f \;=\; \{u \in U \;:\; f(v) = u \text{ for some } v \in V\}.$$

(NB, this is the same as the image, or range, of any function. However, we will see that Image has some useful extra properties when the function is linear.) In our example, the image of f is $\{(0, y) \mid y \in \mathbb{R}\}$, that is, the y–axis in $\mathbb{R}^2$.

Definition 15 *The KERNEL of f is*

$$Ker\ f \;=\; \{v \in V \;:\; f(v) = 0\}.$$

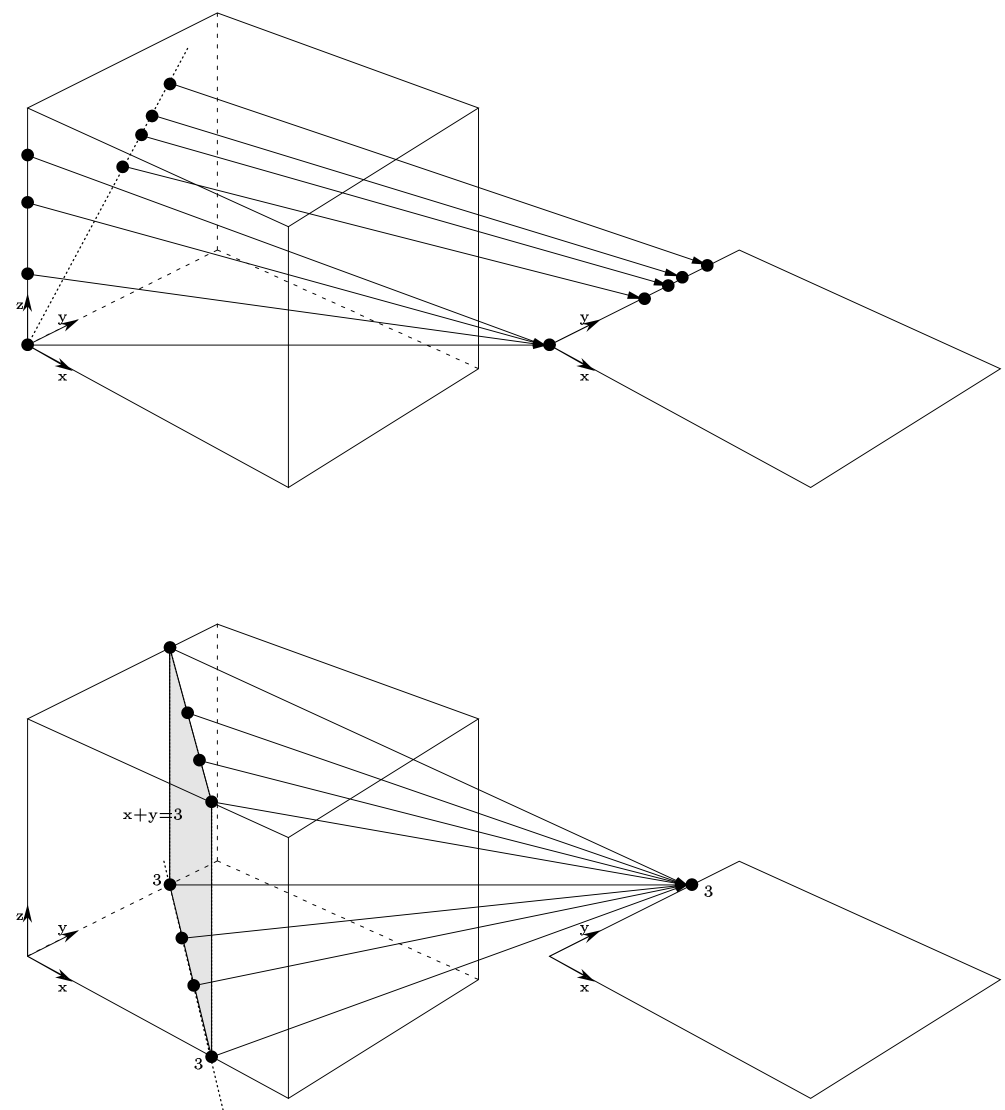

Figure 1: The function $f(x, y, z) = (0, x + y)$. The upper diagram illustrates the map on the lines $x = y = 0$ and $z - y = x = 0$ in $\mathbb{R}^3$. The lower diagram illustrates the map on the plane $x + y = 3$ in $\mathbb{R}^3$.

In our example the kernel of f is the $x + y = 0$ plane.

Theorem 7 *$Im\ f\ \leq U$ and $Ker\ f\ \leq V$.*

Proof: Exercise!

Example 8 *Suppose $f : \mathbb{R}^3 \to \mathbb{R}^3$ is given by*

$$f : (x, y, z) \mapsto (x, y, 0)$$

(called a "projection mapping"), then

$$Im\ f = \{(a, b, 0)\ :\ a, b \in \mathbb{R}\}$$

$$Ker\ f\ = \{(0, 0, c)\ :\ c \in \mathbb{R}\}.$$

Theorem 8 *For V a finite dimensional vector space and $f : V \to U$ a linear function then*

$$dim\ V\ =\ dim(Ker\ f) + dim(Im\ f).$$

Proof: Let $dim\ V = n$. Then $dim(Ker\ f) \leq n$, call it r. Let $\{v_1, v_2, ..., v_r\}$ be a basis for $Ker\ f$. Extend to a basis for V $(\{v_1, v_2, ..., v_r, v_{r+1}, ..., v_n\})$ then we just need:

Proposition 8 *A basis for $Im\ f$ is $\{f(v_{r+1}), f(v_{r+2}), ..., f(v_n)\}$.*

Proof of proposition: We must check span and linear independence.
Span: For each $w \in\ Im\ f$ there exists a $v \in V$ such that $f(v) = w$. Put

$$v = \sum_{i=1}^{n} \alpha_i v_i$$

say, but then by linearity

$$w = f(v) = \sum_{i=r+1}^{n} \alpha_i f(v_i).$$

Linear independence (by contradiction): Suppose

$$\sum_{i=r+1}^{n} \alpha_i f(v_i) = 0$$

with not all coeficients zero. By linearity this implies

$$f\left(\sum_{i=r+1}^{n} \alpha_i v_i\right) = 0$$

so the argument lies in the kernel of f, and may be expressed as a linear combination of kernel basis elements:

$$\sum_{i=r+1}^{n} \alpha_i v_i = \sum_{i=1}^{r} \beta_i v_i.$$

But these vectors are linearly independent, since they are a basis for V, so we have a contradiction. QED (proposition).

Hence $dim(Im\ f) = n - r$. QED (theorem).

Definition 16

$$rank(f) = dim(Im\ f)$$

$$nullity(f) = dim(Ker\ f).$$

Our definition of linearity does not depend on any coordinatisation of (choosing of bases for) the vector spaces involved. However, linearity interacts very usefully with the theory of bases. For example, a linear function is completely determined by its action on a basis of the domain. (See below.) Indeed, we will see shortly that

Theorem 9 *For K a field (as usual we will only be concerned here with $\mathbb{R}$ and $\mathbb{C}$), any K–vector space of dimension n is isomorphic to K^n.*

Example: Let V be an arbitrary $\mathbb{R}$–vector space with basis $\{e_1, ..., e_n\}$, and for $v \in V$ put

$$v = \sum_i \alpha_i e_i.$$

Recall from our Theorem 2 that the coefficients α_i are unique. Thus there is a well defined map from V to $\mathbb{R}^n$ which takes v to the (column) vector of its coefficients α_i. This is clearly linear, and using our Theorem 2 again we can show that it is an isomorphism. The image of v here is called the coordinate vector of v with respect to the given basis.

In particular, a new choice of basis gives rise to an isomorphism from $\mathbb{R}^n$ to itself in this way. For example, with basis $\{(1,1),(1,0)\}$ we have $(x,y) = y.(1,1)+(x-y)(1,0)$, so $(x,y) \mapsto (y, x-y)$. (Readers attaching more structure to such spaces than we have yet covered in this course — such as angles between vectors — will note that isomorphism does not have to preserve angles.)

For each $m \times n$ (real) matrix A we may define a function

$$f_A : \mathbb{R}^n \to \mathbb{R}^m$$

by giving the vectors in each space their natural realisation as column vectors (vertically oriented tuples) and using matrix multiplication

$$f_A(X) = AX.$$

This function is LINEAR. (Exercise.) In these terms the system of simultaneous equations $AX = Y$ gives a linear transformation A and a target vector Y and asks what set of vectors are transformed to that target by A (in particular, the solution of the homogeneous form $AX = 0$ is the kernel of the transformation).

3.2 Matrix representation

More generally, an $m \times n$ matrix (with elements in field K) can be viewed as a (representation of a) linear mapping $T : K^n \to K^m$ in the following way:

Pick bases for K^n, K^m (e.g. the SOB in each case, say $\{e_i\}, \{\epsilon_i\}$ respectively). Then certainly $T(e_1) = \sum_{i=1}^m \alpha_i^1 \epsilon_i$ for some set of coeficients α_i^1, and similarly $T(e_j) = \sum_{i=1}^m \alpha_i^j \epsilon_i$ for $j = 2, ..., n$. (NB, fixing T, the coefficients still depend on the choice of bases.) Thus by linearity an arbitrary element $v = \sum_{k=1}^n \kappa_k e_k$ transforms as $T(\sum_{k=1}^n \kappa_k e_k) = \sum_{k=1}^n \kappa_k \sum_{i=1}^m \alpha_i^k \epsilon_i$. The coefficient of ϵ_i in $T(v)$ is $\sum_{k=1}^n \kappa_k \alpha_i^k$.

Now consider the matrix multiplication $AX = Y$:

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & \dots & & \vdots \\ \vdots & & & \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

giving, by definition,

$$y_i = \sum_{k=1}^n a_{ik} x_k.$$

If we represent v as a column vector $\mathcal{K} = (\kappa_1, ..., \kappa_n)^t$ of coefficients for the given basis (note that this is the *defining* representation in our case, using SOBs — more generally it is the *coordinate* *vector* of v with respect to the given basis) then the corresponding column vector of coefficients for $T(v)$ may be determined by matrix multiplication $A_e^\epsilon \mathcal{K}$, where the matrix A_e^ϵ is given by $a_{ij} = \alpha_i^j$.

For example $T : \mathbb{R}^3 \to \mathbb{R}^3$ specified by

$$T((x,y,z)) = (3x+y, x+y+z, x+2z)$$

has

$$T(e_1) = T((1,0,0)) = (3,1,1) = 3.(1,0,0) + 1.(0,1,0) + 1.(0,0,1) = 3\epsilon_1 + \epsilon_2 + \epsilon_3$$

$$T(e_2) = T((0,1,0)) = (1,1,0) = 1.(1,0,0) + 1.(0,1,0) + 0.(0,0,1) = \epsilon_1 + \epsilon_2$$

$$T(e_3) = T((0,0,1)) = (0,1,2) = 0.(1,0,0) + 1.(0,1,0) + 2.(0,0,1) = \epsilon_2 + 2\epsilon_3$$

so letting

$$A_e^\epsilon = \begin{pmatrix} 3 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 2 \end{pmatrix}$$

then

$$A_e^\epsilon \begin{pmatrix} x \\ y \\ z \end{pmatrix} = (T((x,y,z)))^t .$$

We will see later that it is often useful to change the bases we use to describe transformations (for now this may seem perverse, but it really can be useful). Suppose we want to use the (ordered) basis $\{(1,1,1),(1,1,0),(1,0,0)\}$ for the codomain (leaving the domain basis alone). Then our calculation above becomes

$$T(e_1) = T((1,0,0)) = (3,1,1) = 1.(1,1,1) + 0.(1,1,0) + 2.(1,0,0) = \epsilon_1' + 2\epsilon_3'$$

$$T(e_2) = T((0,1,0)) = (1,1,0) = 0.(1,1,1) + 1.(1,1,0) + 0.(1,0,0) = \epsilon_2'$$

$$T(e_3) = T((0,0,1)) = (0,1,2) = 2.(1,1,1) - 1.(1,1,0) - 1.(1,0,0) = 2\epsilon_1' - \epsilon_2' - \epsilon_3'$$

so letting

$$A_e^{\epsilon'} = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & -1 \\ 2 & 0 & -1 \end{pmatrix}$$

then

$$A_e^{\epsilon'} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x+2z \\ y-z \\ 2x-z \end{pmatrix} .$$

The right hand side is the *coordinate vector* for the transformed vector with respect to the new basis. Check:

$$(x+2z)\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + (y-z)\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + (2x-z)\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} x+2z+y-z+2x-z \\ x+2z+y-z \\ x+2z \end{pmatrix} = \begin{pmatrix} 3x+y \\ x+z+y \\ x+2z \end{pmatrix} .$$

Note that

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} A_e^{\epsilon'} = A_e^\epsilon .$$

There is an analogous change on changing the domain basis. We will return to this shortly. Note that $rank(A) = rank(T)$ always (hence we use the same name).

Definition 17 (Hom) *Let V, U be vector spaces over K, then $Hom(V,U)$ is a vector space over K where the set is all linear maps $f : V \to U$ and $f + g$ is defined by*

$$(f+g) : V \to U$$

$$(f+g) : v \mapsto f(v) + g(v)$$

and αf is

$$(\alpha f) : v \mapsto \alpha f(v).$$

Exercise 15 *Check closure and the other axioms.*

Definition 18 $End(V) = Hom(V,V)$.

Definition 19 *A linear map $f : V \to U$ is SINGULAR if there exists $v \in V$ such that $v \neq 0$ but $f(v) = 0$.*

Otherwise, i.e. if $Ker\ f = \{0\}$, then NON-SINGULAR.

4 Linear functions and systems of linear equations

The matrix equation $Ax = B$ of a systems of linear equations may be seen as a linear map

$$A : K^n \to K^m$$

($K = \mathbb{R}$). Note in particular that the solution space to $Ax = 0$ is $ker\ A$. We then have

$$dim(ker\ A) = dim(K^n) - dim(Im\ A) = n - rank(A)$$

as required.

4.1 Invertible functions

A linear function $T : V \to V$ is invertible if there exists $T^{-1} : V \to V$ such that $TT^{-1} = T^{-1}T = I$ (I is the identity map).

Recall that T is invertible iff a bijection. The 1-to-1 property ensures that $ker(T) = \{0\}$, so then T is non-singular. Conversely if T is non-singular then it is 1-to-1 (exercise), and furthermore

$$dim(V) = dim(Im\ T) + dim(ker\ T) = dim(Im\ T) + 0$$

so it is also onto. Thus we have

Theorem 10 *A linear map $T : V \to V$ is invertible iff non-singular.*

Now consider the matrix equation $Ax = 0$ for A an $n \times n$ matrix. If A non-singular then A^{-1} exists and this implies $x = 0$, and $Ax = B$ has a unique solution ($A^{-1}B$) for each B.

If A is singular $Ax = 0$ has other solutions as well as $x = 0$, thus A is not onto and there exist B's such that $Ax = B$ has NO solution; and if a solution exists it is not unique.

4.2 Determinants

Recall that a bijection from a finite set to itself is called a permutation.

Proposition 9 *The set Σ_n of all possible bijections of a set S of order n, has order $n!$.*

Proof: Exercise!

The set Σ_n forms a group under composition of functions (exercise), called the SYMMETRIC GROUP.

Any permutation $p \in \Sigma_n$ can be written as a product of permutations of the form

$$(ij) = \begin{pmatrix} 1 & 2 & 3 & 4 & \dots & i & i+1 & \dots & j & \dots & n \\ 1 & 2 & 3 & 4 & \dots & j & i+1 & \dots & i & \dots & n \end{pmatrix}$$

that is as a product of permutations in each of which exactly 2 elements are changed.

For example

$$\begin{pmatrix} 123 \\ 312 \end{pmatrix} = \begin{pmatrix} 123 \\ 321 \end{pmatrix} \cdot \begin{pmatrix} 123 \\ 213 \end{pmatrix}$$

and

$$\begin{pmatrix} 123 \\ 123 \end{pmatrix} = \begin{pmatrix} 123 \\ 321 \end{pmatrix} \cdot \begin{pmatrix} 123 \\ 321 \end{pmatrix}$$

Obviously this expansion is not unique, but the number of factors required in the product is unique mod.2 (exercise).

If this number is odd (even) the perm. is called "odd" ("even").

Definition 20 (Parity Function)

$$S : \Sigma_n \to \{+1, -1\}$$

$$S : p \mapsto \begin{cases} +1 & p \text{ even} \\ -1 & p \text{ odd} \end{cases} .$$

Let A be an n-square matrix with entries a^{ij}. Then

$$det\ A = |A| = \sum_{f \in \Sigma_n} \left(S(f) \prod_{i=1}^{n} a_{if(i)} \right) .$$

For example,

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

Then again, for $n = 4$ there will be 24 terms in the sum, like $a_{13}a_{21}a_{34}a_{42}$. This is the term for

$$f = \begin{pmatrix} 1234 \\ 3142 \end{pmatrix}$$

(compare the first and second index on each a_{ij} with the top and bottom row of this perm.). In this case $S(f) = -1$, because f is "odd".

Exercise 16 *Find a product of three pair interchange perms which gives this f.*

4.3 Properties of determinants

(1) $|A| = |A^t|$ (exercise);
(2) Interchanging any 2 rows or columns of A changes the sign of every term in the determinant (i.e. changes the determinant by an overall factor -1.

Exercise 17 *Check an example.*

Note that

$$det(H^{ij}A) = -det(A)$$

and that $det(H^{ij} = -1$.

(3) Multiplying a row or column of A by $\alpha \in \mathbb{R}$ multiplies $det(A)$ by α. That is

$$det(H^i(\alpha)\ A) = \alpha A$$

(note that $det(H^i(\alpha)) = \alpha$).
(4) If matrix A has 2 rows or columns equal then $det(A) = 0$.
(5) The sum of two matrices A and B which are identical in all but one row or one column vector ($a_i \neq b_i$, say) then

$$det(A) + det(B) = det(C)$$

where C is also identical but for $c_i = a_i + b_i$.
(6) Properties 3,4 and 5 imply that

$$det(M^{ij}\ A) = det(A)$$

(note that $det(M^{ij}) = 1$).

Theorem 11 (A) *For A an n-square matrix the following are equivalent:*
i) A is invertible;
(ii) A is non-singular;
(iii) $det(A) \neq 0$.

Theorem 12 (B)

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

Exercise 18 *Prove theorem B for $n = 3$.*

4.4 Determinants, minors and cofactors

Consider A an n-square matrix. Let A_{ij} be the $(n-1)$-square matrix obtained from A be deleting the i^{th} row and j^{th} column. Then $det(A_{ij})$ is called the MINOR of a_{ij} and $a'_{ij} = (-1)^{i+j} det(A_{ij})$ is called the COFACTOR of a_{ij}, and for any i

$$|A| = \sum_{j=1}^{n} a_{ij} a'_{ij}$$

and

$$|A| = \sum_{i=1}^{n} a_{ij} a'_{ij}$$

for any j. This is the "Laplace expansion" for $|A|$, and comes directly from the above definitions. For example

$$\begin{vmatrix} 5 & 4 & 2 & 1 \\ 2 & 3 & 1 & -2 \\ -5 & -7 & -3 & 9 \\ 1 & -2 & -1 & 4 \end{vmatrix} = \begin{vmatrix} 1 & 4 & 2 & 1 \\ -1 & 3 & 1 & -2 \\ 2 & -7 & -3 & 9 \\ 3 & -2 & -1 & 4 \end{vmatrix}$$

since subtracting the second column from the first does not change the determinant. Then after subtracting appropriate multiples of the first column from the second, third and fourth columns this becomes (again with no change - exercise: check it!)

$$= \begin{vmatrix} 1 & 0 & 0 & 0 \\ -1 & 7 & 3 & -1 \\ -2 & -15 & -7 & 7 \\ 3 & -14 & -7 & 1 \end{vmatrix}$$

At this point we can use the Laplace expansion to good effect, expanding with respect to the first row. We get

$$= \begin{vmatrix} 7 & 3 & -1 \\ -15 & -7 & 7 \\ -14 & -7 & 1 \end{vmatrix} = \begin{vmatrix} 0 & 0 & -1 \\ 34 & 14 & 7 \\ -7 & -4 & 1 \end{vmatrix} = - \begin{vmatrix} 34 & 14 \\ -7 & -4 \end{vmatrix} = 34.4 - 14.7 = 136 - 98 = 38$$

(we added some copies of the last column to other columns at one point here, and then used "Laplace" again) therefore A^{-1} exists!

Definition 21 (CLASSICAL ADJOINT) *The (classical) adjoint of A (written $adj(A)$) is the transpose of the matrix formed by replacing each a_{ij} with its cofactor.*

Proposition 10

$$A.adj(A) = |A|I.$$

Note that this proves that theorem A part (iii) implies part (i). Also note that hence, provided $|A| \neq 0$, then

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

The proof is slightly messy, but well illustrated by examples.

Exercise 19 *Verify the theorem for the case*

$$A = \begin{pmatrix} 2 & 3 & 4 \\ 4 & 3 & 1 \\ 1 & 2 & 4 \end{pmatrix} \tag{12}$$

by obtaining

$$A^{-1} = \frac{1}{-5} \begin{pmatrix} 10 & -4 & -9 \\ -15 & 4 & 14 \\ 5 & -1 & -6 \end{pmatrix} \tag{13}$$

and then checking that $AA^{-1} = 1$.

5 Applications

In linear equations:

$$Ax = B$$

implies

$$A^{-1}Ax = A^{-1}B$$

so

$$x = A^{-1}B$$

and theorem A (i) implies (ii); and (iii) implies (ii).

Here is a concrete example: The system

$$2x + 3y + 4z = 1$$

$$4x + 3y + z = 3$$

$$x + 2y + 4z = 0$$

has coefficient matrix A as in equation 12 above, and

$$B = \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix}$$

so the solution is

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = A^{-1}. \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix} = \begin{pmatrix} 2/5 \\ 3/5 \\ -2/5 \end{pmatrix}$$

using A^{-1} from equation 13 (exercise). This may be directly verified by substitution into the original system.

Example 2:

$$2x + 3y + 4z = 1$$

$$4x + 3y + z = 3$$

$$6x + 6y + 5z = 0$$

has

$$A = \begin{pmatrix} 2 & 3 & 4 \\ 4 & 3 & 1 \\ 6 & 6 & 5 \end{pmatrix} \tag{14}$$

for which $|A| = 0$, so either there are many or NO solutions. The homogeneous system $Ax = 0$ has many solutions (rank A is 2), but this particular inhomogeneous system has no solution, so the solution set is EMPTY!

5.1 Another way to invert non-singular matrices

If we perform elementary row operations (by the matrix premultiplications we defined above) on A until we reach I (which is always possible for a non-singular A - exercise), then the matrix product of these operations is A^{-1}.

Example: – see lectures. See also [Schaum — Linear Algebra] §§3.36, 3.38.

5.2 Review - solution of n equations in n variables

$$Ax = B$$

(1) If $|A| \neq 0$ then $x = A^{-1}B$ is unique solution, so find A^{-1};
(2) If $|A| = 0$ then either we will have inconsistency (hence no solutions) or an infinite number of solutions.

Geometrically, $n = 2$ i.e. 2 equations in two unknowns, may be thought of as two lines in the plane. Either

a) they intersect at a unique point;
b) they coincide (infinite no. of solutions);
c) they are parallel (no solutions).

There is a similar picture for equations viewed as planes in $\mathbb{R}^3$, and so on.

5.3 Solution of m equations in n variables

$$Ax = B$$

Firstly $|A|$ is not defined unless $m = n$. Let us think of

$$A : \mathbb{R}^n \to \mathbb{R}^m$$

We have solutions only if $B \in Im(A)$, then find any u such that $Au = B$, whereupon the general solution is

$$\{u + v \ : \ v \in ker\ A\}.$$

Why? - well, $Au = B$ and $Av = 0$, so linearity gives

$$A(u+v) = Au + Av = B + 0 = B$$

so any such $(u+v)$ is a solution.

Example:

$$\begin{pmatrix} 6 & 4 & 2 \\ 1 & 3 & 9 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ 7 \end{pmatrix}$$

has *rank* $A = 2$, and the same rank for the augmented matrix (exercise!). Thus there are solutions, and the dimension of the solution space of the associated homogeneous problem $Ax = 0$ is 3-2=1.

Geometrically

$$6x + 4y + 2z = 2$$

is a plane in $\mathbb{R}^3$, as is

$$x + 3y + 9z = 7.$$

These two planes are neither parallel nor coincident, so they intersect in a LINE:

$$7y + 26z = 20$$

$$y = (20 - 26z)/7$$

$$7x = 15z - 11$$

thus one way to represent the solution set is

$$\{(\frac{15z-11}{7}, \frac{20-26z}{7}, z) \ : \ z \in \mathbb{R}\}.$$

6 Eigenvalues and Eigenvectors

Let $T : V \to V$ be a linear function on a vector space V (over K).

Definition 22 *A scalar $\lambda \in K$ is called an EIGENVALUE of T if there exists $v \in V - \{0\}$ such that*

$$Tv = \lambda\, v.$$

The vector v is called an EIGENVECTOR of T.

Proposition 11

$$\{v \in V \ : \ Tv = \lambda\, v\} \leq V$$

This subspace, written $S_T(\lambda)$, is called the eigenspace of λ.

Proof: (of closure):

For all $u, v \in S_T(\lambda)$

$$T(\alpha u + \beta v) = \alpha T u + \beta T v = \alpha \lambda u + \beta \lambda v = \lambda(\alpha u + \beta v).$$

QED.

Example: Find the eigenvalues and eigenvectors of $A = \begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix}$, i.e. find $\lambda \in \mathbb{R}$ and $v \in \mathbb{R}^2$ such that

$$\begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \end{pmatrix}.$$

This is the same thing as

$$\begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

in other words

$$\begin{pmatrix} 1-\lambda & 2 \\ 3 & 2-\lambda \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0.$$

Recall that such a homogeneous system has non-zero solution iff

$$det \begin{pmatrix} 1-\lambda & 2 \\ 3 & 2-\lambda \end{pmatrix} = 0$$

i.e.

$$(\lambda - 4)(\lambda + 1) = 0$$

therefore the only possible eigenvalues are $4, -1$. Putting these into the equation we can solve for v in each case:

Try $\lambda = 4$:

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

implies $3x = 2y$ so, for example,

$$\begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = 4. \begin{pmatrix} 2 \\ 3 \end{pmatrix}.$$

Obviously if $v = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$ is an eigenvector then kv is an eigenvector for any $k \in K$, so we don't usually explicitly mention all the other constant multiples of v.

For $\lambda = -1$ we have

$$\begin{pmatrix} 2 & 2 \\ 3 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

giving $x = -y$, so

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

This leads us to

Theorem 13

$$S_T(\lambda) = ker(\lambda I - T).$$

(Note that the kernel only exists if $|\lambda I - T| = 0$.)

Theorem 14 (Independence theorem) *If $u \in S_T(a)$ and $v \in S_T(b)$ and $a \neq b$ then u and v are linearly independent.*

Proof: suppose $\alpha u + \beta v = 0$, then $\alpha a u + \beta b v = T.0 = 0$ so

$$\alpha(b-a)u + \beta(b-b)v = 0$$

which implies $\alpha = 0$, and hence $\beta = 0$. QED.

Note, on the other hand, that u, v lin.indep. does not imply $a \neq b$.

6.1 Similar matrices and matrix diagonalisation

Definition 23 *Matrices A and C are SIMILAR if there exists a non-singular matrix B such that*

$$C = B^{-1}AB$$

Remarks:
(1) We have

Proposition 12 *Similar matrices have the same eigenvalues.*

Proof:

$$|C - \lambda I| = |B^{-1}AB - \lambda I| = |B^{-1}AB - \lambda I B^{-1}B|$$
$$= |B^{-1}(A - \lambda I)B| = |B^{-1}|.|A - \lambda I)|.|B| = |A - \lambda I|.$$

Definition 24 *The polynomial of order n in λ given by $|A - \lambda I|$ is called the Characteristic Polynomial of A, and*

$$|A - \lambda I| = 0$$

is the characteristic equation.

Theorem 15 (Cayley-Hamilton) *Let $P(\lambda) = |A - \lambda I|$. Then $P(A) = 0$ (which is to say that if we substitute A in for λ in the polynomial P it evaluates to the zero matrix).*

The roots of $|A - \lambda I| = 0$ are the eigenvalues of A. QED.
(2) A, A^t have the same eigenvalues (exercise);
(3) $Av = \lambda v$ implies $(kA)v = (k\lambda)v$ for any $k \in K$.
(4) $Av = \lambda v$ implies

$$A^2 v = A(Av) = A\lambda v = \lambda Av = \lambda^2 v$$

and similarly for A^n;
(5) A non-singular and $Av = \lambda v$ implies

$$(AA^{-1})v = Iv = v = (A^{-1}A)v = A^{-1}\lambda v$$

so that finally $A^{-1}v = \lambda^{-1}v$.

It is a corollary of the independence theorem that if $dim\ V = n$ and $T : V \to V$ has all n eigenvalues distinct then the n corresponding eigenvectors are linearly independent and hence form a basis for V. Let's call then $\{v_1, v_2, ..., v_n\}$ then the matrix representation of T in this basis comes from

$$T(v_1) = \lambda_1 v_1 \quad (+0.v_2 + 0.v_3 + ...)$$
$$T(v_2) = \lambda_2 v_2 \quad (+0.v_1 + 0.v_3 + ...)$$
$$T(v_3) = \lambda_3 v_3 \quad (+0.v_1 + 0.v_2 + ...)$$

$\vdots$

$$T(v_n) = \lambda_n v_n \quad (+0.v_1 + 0.v_2 + ...)$$

so then the matrix for T is diagonal:

$$\begin{pmatrix} \lambda_1 & 0 & ... & 0 & 0 \\ 0 & \lambda_2 & 0 & ... & 0 \\ 0 & 0 & \lambda_3 & & \\ ... & & & & \\ ... & & & & \\ 0 & 0 & \cdots & 0 & \lambda_n \end{pmatrix}$$

(all other entries zero) and the corresponding eigenvectors are the standard ordered basis!

Theorem 16 (Diagonalisation theorem) *If $\dim V = n$ and matrix $A : V \to V$ has n lin.indep. eigenvectors $\{v_1, v_2, ..., v_n\}$ corresponding to (not necessarily distinct) eigenvalues $\lambda_1, \lambda_2, ..\lambda_n$, and if P is a matrix with columns $\{v_1, v_2, ..., v_n\}$ then $|P| \neq 0$ and*

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & 0 & ... & 0 & 0 \\ 0 & \lambda_2 & 0 & ... & \\ & & & & \\ 0 & 0 & \vdots & & \\ 0 & 0 & \cdots & 0 & \lambda_n \end{pmatrix}$$

Proof: (i) P^t has n lin.indep. rows by construction, thus $|P^t| \neq 0$, but $|P| = |P^t|$ so P is non-singular;

(ii) Note that $(v_1 v_2 ... v_n)$ is an $n \times n$ matrix, then

$$AP = A(v_1 v_2 ... v_n) = (\lambda_1 v_1 \ \lambda_2 v_2 \ ... \ \lambda_n v_n)$$

$$= (v_1 v_2 ... v_n) \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 & 0 \\ 0 & \lambda_2 & 0 & \ldots & \\ 0 & 0 & \vdots & & \\ 0 & 0 & \cdots & 0 & \lambda_n \end{pmatrix} = P \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 & 0 \\ 0 & \lambda_2 & 0 & \ldots & \\ 0 & 0 & \vdots & & \\ 0 & 0 & \cdots & 0 & \lambda_n \end{pmatrix}$$

thus $P^{-1}AP$ gives the required diagonal matrix. QED.

Corollary: Every matrix with all distinct eigenvalues is similar to a diagonal matrix with diagonal entries given by the eigenvalues.

For example, suppose $T : \mathbb{R}^3 \to \mathbb{R}^3$ in some basis is given by

$$A = \begin{pmatrix} 1 & 0 & 0 \\ -8 & 4 & -6 \\ 8 & 1 & 9 \end{pmatrix}$$

Then

$$|A - \lambda I| = (1 - \lambda)(\lambda^2 - 13\lambda + 42) = 0$$

gives $\lambda = 1, 6, 7$ (all different).

Then for $\lambda = 1$ we have

$$\begin{pmatrix} 1 & 0 & 0 \\ -8 & 4 & -6 \\ 8 & 1 & 9 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0$$

giving

$$-8x + 3y - 6z = 0$$

and

$$8x + y + 8z = 0$$

so we can use eigenvector $\begin{pmatrix} 15 \\ 8 \\ -16 \end{pmatrix}$ (for example).

For $\lambda = 6$ we get eigenvector $\begin{pmatrix} 0 \\ -3 \\ 1 \end{pmatrix}$ by similar means.

Finally $\lambda = 7$ gives $\begin{pmatrix} 0 \\ -2 \\ 1 \end{pmatrix}$ and altogether we have

$$P = \begin{pmatrix} 15 & 0 & 0 \\ 8 & -3 & -2 \\ -16 & 1 & 1 \end{pmatrix}$$

giving

$$P^{-1} = \frac{-1}{15} \begin{pmatrix} -1 & 0 & 0 \\ 24 & 15 & 30 \\ -40 & -15 & -45 \end{pmatrix}.$$

Finally PAP^{-1} takes the required form $(diag(1, 6, 7))$ - exercise!

More generally, if A and B are matrix representations of the same transformation $T : V \to V$ obtained by working with different bases then there exists a matrix Q such that $Q^{-1}AQ = B$.

Exercise 20 *Find two eigenvalues and eigenvectors of*

$$M = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

Definition 25 *A real matrix is ORTHOGONAL if $A^t = A^{-1}$ (i.e. its rows are 'orthonormal').*

For example consider the vectors $v = (1,1)$ and $u = (1,-1)$. Then $u.v = 0$, $u.u = v.v = 2$ (the usual dot product), so

$$M = \frac{1}{\sqrt{2}} \begin{pmatrix} v \\ u \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

obeys $MM^t = 1$, so $M^t = M^{-1}$ ($= M$ in this case).

Definition 26 $A^\dagger = (A^t)^*$.

Definition 27 *A matrix A is HERMITIAN if $A = A^\dagger$.*

Proposition 13 *If A is Hermitian then the roots of its characteristic polynomial are all real, and there exists a matrix P such that PAP^{-1} is diagonal.*

Proof: Exercise.

Exercise 21 *Read [Schaum — Linear Algebra] Chapters 9 and 10.*

6.2 $n \times n$ Matrices with less than n eigenvalues

It is a consequence of the diagonalisation theorem that a matrix cannot have *more* than n distinct eigenvalues (or more than n linearly independent eigenvectors). What about less?

Strictly speaking, a matrix may not have any eigenvalues (just as an equation may have no solutions). Consider $T : \mathbb{R}^2 \to \mathbb{R}^2$ represented by

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

This has $|A - \lambda I| = \lambda^2 - 1 = 0$ so $\lambda = \pm i$, but these are NOT in $\mathbb{R}$ (but rather in $\mathbb{C}$). This is, then, just the same as to say that not all polynomials are completely factorizable over $\mathbb{R}$, (although they ARE over $\mathbb{C}$).

More importantly, what about, for example,

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

Then the characteristic equation is $\lambda^2 = 0$, i.e. the root is degenerate. Putting $\lambda = 0$ we can only find one linearly independent eigenvector (exercise).

Then again, suppose that I_n is the $n \times n$ unit matrix, and U_n is the $n \times n$ matrix with all entries zero except on the diagonal immediately above the main diagonal, where all the entries are 1. Then the matrix $J_n(\lambda) = \lambda I_n + U_n$ is called the *Jordan block matrix* with eigenvalue λ. (For example, from above, $A = J_2(0)$.)

Exercise 22 *How many independent eigenvectors does this matrix have in general?*

In fact every matrix is similar to a block diagonal matrix whose blocks are Jordan block matrices. Up to the order in which these blocks appear this is a canonical form for the matrix, called the *Jordan Canonical (or Jordan Normal) Form.*

6.3 The Trace of a matrix

Recall:

Definition 28 (Trace) *For a square matrix A we define $Trace(A)$ as the sum of the diagonal entries, that is*

$$Trace(A) = \sum_{i=1}^{n} a_{ii}.$$

Proposition 14

$$Trace(AB) = Trace(BA).$$

Exercise 23 *Check this for all 2×2 matrices!*

...therefore $Trace(B^{-1}AB) = Trace(A)$, which implies that if A has n distinct eigenvalues $\lambda_1, \lambda_2, ...$ then

$$Trace(A) = \sum_{i=1}^{n} \lambda_i.$$

For example, with

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

we know immediately that $\lambda_1 = -\lambda_2$. Now noting that

$$A^2 = \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix}$$

that is, with eigenvalue 10, we deduce that $\lambda_1 = \sqrt{10}$, and so on.

Exercise 24 *Find the 2 indep. eigenvectors here.*

7 Inner product spaces

(**7.1**) For V a real (complex) vector space, assign to each pair $u, v \in V$ a scalar $< u, v >$. This is an *inner product* if

(i) it is linear as a function of u;
(ii) $< u, v > = < v, u >^*$;
(iii) $< u, u > \geq 0$ and $< u, u > = 0$ if and only if $u = 0$.

The space V with inner product is called an inner product space.

Define $||u|| = \sqrt{< u, u >}$, called the *norm* of u.

A real inner product space is a *Euclidean space.*

(**7.2**) Examples:

(i) The usual dot product on $\mathbb{R}^n$ makes it a Euclidean space. Further, norms give *distances* in the usual way, making it a *metric space.*

(ii) For the vector space of $m \times n$ matrices over $\mathbb{R}$

$$< A, B > := trace(B^t A)$$

is an inner product.

(iii) For V the space of real continuous functions on interval $[a, b]$

$$< f, g > := \int_a^b f(t)g(t)dt$$

is an inner product.
(**7.3**) Cauchy-Schwarz inequality. For any $u, v \in V$

$$| < u, v > | \leq ||u||.||v||$$

(**7.4**) For V an inner product space we say u, v *orthogonal* if $< u, v >= 0$ (cf. the *angle* between vectors in $\mathbb{R}^n$). A set $\{v_i\}$ of vectors in V is *orthogonal* if they are pairwise orthogonal. If a set is orthogonal and each $||u_i|| = 1$ then the set is *orthonormal.*

Note that an orthonormal set $\{v_i\}$ is linearly independent, and for any $u \in V$ the vector

$$w(u) := u - \sum_i < u, v_i > v_i \tag{15}$$

is orthogonal to each v_i (although not necessarily non-zero!).

7.1 Gram–Schmidt Process

(**7.5**) Gram–Schmidt Orthogonalization. For $\{v_i\}$ a basis of V there exists a corresponding orthonormal basis $\{u_i\}$, with

$$u_i = \sum_{j=1}^{i} a_{ij} v_j$$

for some coefficients a_{ij}.
Proof: Exercise. Hint: consider equation 15!
(**7.6**) An Orthogonal matrix is a real matrix A such that $A^t = A^{-1}$.
Exercises.

(i) Obtain an orthonormal basis for $\mathbb{R}^3$ from

$$\{(1,1,1), (0,1,1), (0,0,1)\}$$

(ii) Show that the rows of an orthogonal matrix form an orthonormal set.