NUMERO: 1652A - ANNO: 2015

A P P U N T I

STUDENTE: Salis

MATERIA: Geometry. Prof.Carlini

NOTES 1 - Geometric vectors in the plane and in the space

Applied vectors and free vectors

The typical treasure map gives information to you using applied vectors: in the first part of our treasure hunt we start from a point O and then we move along the west-est direction, in the verse towards est, walking a distance of magnitude 30 steps. This task can be encoded by an arrow → (vector) starting at the point O (application point). This is an example of an applied vector.

- In Physics applied vectors are widely used. Think of a force applied to a body, of the speed of a point and so on.
- In Mathematics we want to separate the information encoded in a applied vectors in two pieces: application point + vector.

Definition

A free vector $\vec{v}$ is completely determined by
(i) the direction of the line containing $\vec{v}$
(ii) the verse following which we move along the line according to $\vec{v}$
(iii) the magnitude (or length) of $\vec{v}$ denoted as $\|\vec{v}\|$

Definition

An applied vector is a free vector with an application point.
Note that for each free vector there are infinitely applied vectors, one for each possible application point. This means that different applied vectors can correspond to the same free vector. (= magn., direct, verse)

Exercise. When is it the case that different applied vectors correspond to the same free vector?

Example. A common way to produce applied vectors is to take two points, say A and B. Then we write $\vec{AB}$ or $(B-A)$, to represent the applied vector going from A to B. Notice that $\vec{BA}$ has the same direction and magnitude of $\vec{AB}$ but opposite verse.

Exercise. What kind of vector is $\vec{AA}$?

There are different ways to visualize the set of free vectors of the plane; one is the following. Fix a point O and apply all free vectors to this point. Then, the free vectors of magnitude R are in bijection with the points of the circumference of center O and radius R.

Exercise. How can we visualize free vectors in 3D-space?

Operation with vectors:

Natural numbers arose naturally to count objects, and for a long while there was no place for the number zero (no objects). However, if we want not only to count, but also to perform operations, zero is of crucial importance.

Definition

The zero vector $\vec{0}$ is the only vector of null magnitude. It does not have direction or verse.
We first see how to multiply a vector by a real number $c \in \mathbb{R}$ which we call a scalar.

Definition - Multiplication by a scalar

Let $c \in \mathbb{R}$ and $\vec{v}$ a free vector. If $c=0$ then $c\vec{v} = 0\vec{v} = \vec{0}$. If $c \neq 0$ then $c\vec{v}$ is the vector:
(i) having the same direction of $\vec{v}$,
(ii) having the same verse of $\vec{v}$ if $c>0$ and opposite verse if $c<0$
(iii) having magnitude $\|c\vec{v}\| = |c|\,\|\vec{v}\|$

Exercise. Define the multiplication by a scalar for applied vectors.

Note that $c\vec{v}$ is obtained by contracting or dilating the vector $\vec{v}$. Also, take note of the absolute value of c appearing in the definition. What would be the problem with $c\|\vec{v}\|$?
Let's go back to our treasure hunt. To reach X, we have to start from O, walk 30 steps est and 40 steps north. Or, we can move 50 steps north-est from O and then reach X. In other words, we can move along the diagonal $\vec{OX}$ instead of following the two sides.

Definition

Parallelogram Rule. Given free vectors $\vec{u}$ and $\vec{v}$, their sum $\vec{u+v}$ is obtained in the following way. Apply $\vec{u}$ and $\vec{v}$ at the same point O and consider the parallelogram having $\vec{u}$ and $\vec{v}$ as two consecutive sides. Then, $\vec{u+v}$ is the diagonal of the parallelogram starting from O.

Exercise. What can we say about the sum of applied vectors?

The multiple by a scalar and the addition of vectors have many useful properties. These properties make computing expressions with vectors very similar to computing expressions with real numbers.

Proposition

Basic properties. Let $c, d \in \mathbb{R}$ and $\vec{u}$, $\vec{v}$ and $\vec{z}$ be free vectors, then the following hold:
(i) $c(d\vec{u}) = (cd)\vec{u}$
(ii) $(c+d)\vec{u} = c\vec{u} + d\vec{u}$

components of $b\vec{w}$. As $b\vec{w}$ is a dilation, or contraction, of $\vec{w}$ by the factor b the result on the components follows.
QED

Exercise. Why it is enough to prove the Proposition for $a=0$ and then the case $a=b=1$?

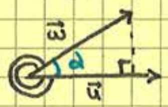
NOTES 2 – More operations with vectors: scalar product, vector product and mixed product.

Dot Product / Scalar product

Definition
Angles. Consider vectors $\vec{u}$ and $\vec{w}$ with a common tail. The angle between the vectors $\vec{u}$ and $\vec{w}$ is the smallest angle that the first vector, $\vec{u}$, makes while rotating onto the second vector, $\vec{w}$.

We now see how to use the angle between two vectors

Definition
Dot product or scalar product. The dot or scalar product of the vectors $\vec{u}$ and $\vec{w}$ is $\vec{u}\cdot\vec{w} = \|\vec{u}\|\|\vec{w}\|\cos(\alpha)$
where α = angle between $\vec{u}$ and $\vec{w}$.



$\alpha \le \pi$

Example Consider the unit vectors $\vec{i}, \vec{j}$ and $\vec{k}$. We want to compute $\vec{i}\cdot\vec{i}$. In order to do this, recall that $\vec{i}$ thus $\|\vec{i}\|=1$. Moreover, $\vec{i}$ forms a zero angle with itself. Hence we have
$\vec{i}\cdot\vec{i} = \|\vec{i}\|\|\vec{i}\|\cos(\alpha) = 1$
Similarly we can compute $\vec{j}\cdot\vec{k} = \|\vec{j}\|\|\vec{k}\|\cos(\pi/2) = 0$

Exercise. Compute the dot product of all possible pairs chosen among the unit vectors $\vec{i}, \vec{j}$ and $\vec{k}$. Does the result depend upon the order in which we choose the vectors?

The dot product is strictly related with projections

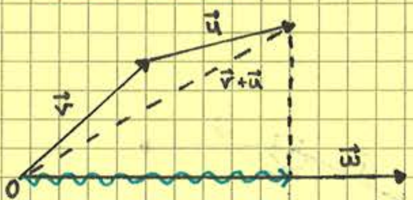
Proposition
Orthogonal projection. Let $\vec{u}$ and $\vec{w}$ be non-zero vectors, then $(\vec{u}\cdot\vec{w})\dfrac{\vec{w}}{\|\vec{w}\|^2} = \vec{z}$ is the orthogonal projection of $\vec{u}$ along the direction of $\vec{w}$.

PROOF The proof follows by the standard trigonometry, noticing that the scalar $\dfrac{\vec{u}\cdot\vec{w}}{\|\vec{w}\|} = \|\vec{u}\|\cos(\alpha)$ is the length of the orthogonal projection.

Proposition
Dot product properties. Consider vectors $\vec{u}, \vec{w}, \vec{z}$ and a scalar $a \in \mathbb{R}$. Then the following holds:
(i) Linearity 1 $(a\vec{u})\cdot\vec{w} = a(\vec{u}\cdot\vec{w})$
(ii) Linearity 2 $(\vec{u}+\vec{w})\cdot\vec{z} = \vec{u}\cdot\vec{z} + \vec{w}\cdot\vec{z}$
(iii) Symmetry $\vec{u}\cdot\vec{w} = \vec{w}\cdot\vec{u}$
(iv) Positivity $\vec{u}\cdot\vec{u} = \|\vec{u}\|^2 \ge 0$

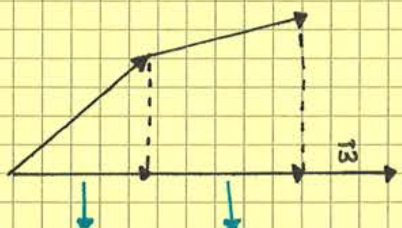
PROOF (Hint) The first two properties are proved by using the projection interpretation of the dot product. While the remaining properties follow by the definition. QED

PROOF Linearity 2



Assume $\|\vec{w}\|=1$
is the orthogonal projection of $\vec{v}+\vec{u}$ along $\vec{w}$

$[(\vec{u}+\vec{v})\cdot\vec{w}]\cdot\vec{w}$



$[(\vec{v}\cdot\vec{w})\cdot\vec{w}]$ $[(\vec{u}\cdot\vec{w})\cdot\vec{w}]$

* $[(\vec{v}\cdot\vec{w})\cdot\vec{w}] + [(\vec{u}\cdot\vec{w})\cdot\vec{w}] = [(\vec{u}+\vec{v})\cdot\vec{w}]\cdot\vec{w}$

$\underbrace{\boxed{(\vec{u}\cdot\vec{w}+\vec{v}\cdot\vec{w})}}_{a}\,\vec{w} = \underbrace{\boxed{[(\vec{u}+\vec{v})\cdot\vec{w}]}}_{b}\,\vec{w}$

$\|a\vec{w}\| = \|b\vec{w}\| \rightarrow |a|=|b|$ magnitude, direction and verse are the same, so the two vectors are equal.
It happens when:
$a\vec{w} = b\vec{w} \iff (a-b)\vec{w} = \vec{0} \iff |a-b|=0 \iff a=b$

We said that the vector product can detect parallelism between two vectors. This can be easily done noticing the following

*** COROLLARY**

If $\vec{u}, \vec{w} \neq \vec{0}$, then $\vec{u} \parallel \vec{w} \iff \vec{u} \times \vec{w} = \vec{0}$

Proposition
Vector product properties
Consider vectors $\vec{u}, \vec{w}, \vec{z}$ and a scalar $a \in \mathbb{R}$. Then the following holds:
(i) Linearity 1 $\quad (a\vec{u}) \times \vec{w} = a(\vec{u} \times \vec{w})$
(ii) Linearity 2 $\quad (\vec{u} + \vec{w}) \times \vec{z} = \vec{u} \times \vec{z} + \vec{w} \times \vec{z}$
(iii) Skew symmetry $\quad \vec{u} \times \vec{w} = -\vec{w} \times \vec{u}$
(iv) $\vec{u} \times \vec{u} = \vec{0}$

PROOF The proof of (ii) will require the notion of mixed product and it will be given in the next section.
We give here a proof of (i). If one among $a, \vec{u}$ and $\vec{w}$ is zero, then the equality is clear. Hence we may assume that no one of them is zero. Let
$\vec{z} = (a\vec{u}) \times \vec{w}$ and $\vec{t} = a(\vec{u} \times \vec{w})$
We want to show that $\vec{z} = \vec{t}$. It's easy to see that $\vec{z}$ and $\vec{t}$ have the same direction as this is the common perpendicular to $\vec{u}$ and $\vec{w}$. Also the magnitudes coincide, in fact
$\|\vec{z}\| = \|a\vec{u}\| \|\vec{w}\| \sin\alpha = |a| \|\vec{u}\| \|\vec{w}\| \sin\alpha = \|\vec{t}\|$
Finally, we deal with verse. First consider the case $a > 0$. Notice that the angle between $a\vec{u}$ and $\vec{w}$ and the angle between $\vec{u}$ and $\vec{w}$ coincide. Thus $a\vec{z}$ and $\vec{t}$ have the same verse.
Now consider the case $a < 0$. In this situation, $(a\vec{u}) \times \vec{w}$ and $\vec{u} \times \vec{w}$ have opposite verse. But, as $a < 0$, the vector $\vec{z}$ and $\vec{t}$ have, again, the same verse. QED

This properties allows us to compute the vector product of any pairs of vectors.

Example
$(\vec{i} + 2\vec{j}) \times (2\vec{j} - \vec{k}) =$
$= \vec{i} \times (2\vec{j} - \vec{k}) + 2\vec{j} \times (2\vec{j} - \vec{k}) = \vec{i} \times 2\vec{j} - \vec{i} \times \vec{k} + 2\vec{j} \times 2\vec{j} - 2\vec{j} \times \vec{k} = -2\vec{i} + \vec{j} + 2\vec{k}$

Extensively using the properties of the vector product we can find the following formula to compute the vector product.

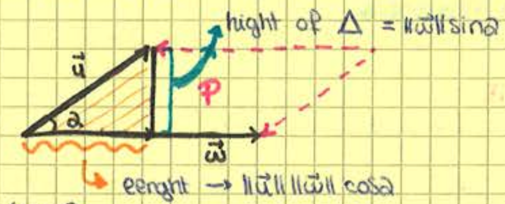
Proposition
Formula for $\vec{u} \times \vec{v}$ using components
Consider the vectors $\vec{v} = v_x\vec{i} + v_y\vec{j} + v_z\vec{k}$ and $\vec{w} = w_x\vec{i} + w_y\vec{j} + w_z\vec{k}$, then

$$\vec{u} \times \vec{v} = \begin{pmatrix} \vec{i} & \vec{j} & \vec{k} \\ u_x & u_y & u_z \\ w_x & w_y & w_z \end{pmatrix} \longrightarrow (v_y w_z - w_z w_y)\vec{i} - (v_x w_z - v_z w_x)\vec{j} + (v_x w_y - v_y w_x)\vec{k}$$

Exercise. Prove the formula!
We conclude describing how the vector product relates to computing areas.

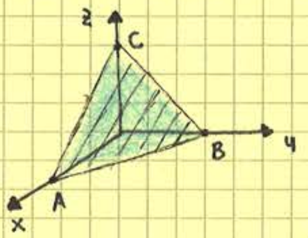
Proposition - Areas
Consider vectors $\vec{u}$ and $\vec{w}$ having common tail, then $\|\vec{u} \times \vec{w}\| = 2\,\text{area}(T) = \text{area}(P)$, where T is the triangle, and P is the parallelogram, of sides $\vec{u}$ and $\vec{v}$.

PROOF We use standard trigonometry. Recall that
$\|\vec{u} \times \vec{w}\| = \|\vec{u}\| \|\vec{w}\| \sin\alpha$, where α = angle between $\vec{u}$ and $\vec{w}$. Now it is enough to notice that
$\|\vec{w}\| \sin\alpha$ = length of the height of the triangle T w.r.t. the base $\vec{u}$. QED

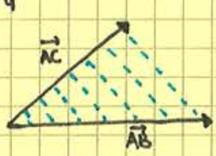


$$\|\vec{u} \times \vec{w}\| = \underbrace{\|\vec{u}\| \sin\alpha}_{\text{HEIGHT}} \|\vec{w}\| = \text{Area } P$$

Example
Find the area of the triangle ABC.



A(1,0,0)
B(0,1,0)
C(0,0,1)



For each couple of points, we construct a vector.
$\vec{AB} = (B-A) = -\vec{i} + \vec{j}$
$\vec{AC} = (C-A) = -\vec{i} + \vec{k}$

$2\,\text{Area } T = \|\vec{AC} \times \vec{AB}\|$

$$\vec{AC} \times \vec{AB} = \begin{pmatrix} \vec{i} & \vec{j} & \vec{k} \\ -1 & 0 & 1 \\ -1 & 1 & 0 \end{pmatrix} = -(\vec{i} + \vec{j} + \vec{k})$$

$\text{Area } T = \frac{1}{2} \|\vec{i} + \vec{j} + \vec{k}\| = \sqrt{3}/2$

* check if it's orthogonal to both the vectors (vector and dot product = 0)

NOTES 3 - Matrix addition and multiplication

Matrices and their entries

A matrix is a rectangular array of numbers.

Examples $\begin{pmatrix} 1 & 2 & 3 \\ 0 & 6 & 9 \end{pmatrix}$ $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ $\begin{pmatrix} 0&0&0&0&1 \\ 1&0&0&0&0 \\ 0&1&0&0&0 \\ 0&0&1&0&0 \\ 0&0&0&1&0 \end{pmatrix}$ $\begin{pmatrix} 1 & 2 \\ 1 & 4 \end{pmatrix}$

The individual numbers are called the entries, elements, or components of the matrix. If the matrix has m rows and n columns, we say that it has size "m by n" or $m \times n$. The above examples have respective size 2×3, 4×1, 5×5, 2×2.
If $m = n$ (the last two cases) the matrix is obviously square (same no. of columns and rows)
The set of matrices of size $m\times n$ whose entries are real numbers is denoted by $\mathbb{R}^{m,n}$; the first superscript is always the number of rows. Sometimes we use symbols to represent unspecified numbers, so the statement $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{R}^{2,2}$ is tantamount to saying that a,b,c,d are real numbers.

For matrices with more than about 4 entries, it's convenient to use subscripts to label the entries. Given a matrix A, we typically denote by a_{ij} the entry in the ith row and jth column (lower case to emphasize that the entry is a number)

Example

$A \in \mathbb{R}^{m,n}$ $\quad A = \begin{pmatrix} a_{1,1} & a_{1,2} \dots & a_{1,n} \\ a_{2,1} & a_{2,2} \dots & a_{2,n} \\ \vdots & \vdots & \vdots \\ a_{m,1} & a_{m,2} \dots & a_{m,n} \end{pmatrix} = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \dots & a_{m,n} \end{pmatrix} = (a_{i,j})_{1\le i\le m}$

* Mathematicians like to deal in generalities and will even write a matrix as $A = (a_{ij})$ without spec. its size.

Examples

$A \in \mathbb{R}^{2,2}$ $\quad A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ $\quad \begin{bmatrix} a_{11} = 1 & a_{12} = 2 \\ a_{21} = 3 & a_{22} = 4 \end{bmatrix}$

Vectors

Of special importance are matrices that have only one row or column; they are called row and column vectors. In writing a row vector with digits, it's useful to use commas to separate the entries. For example, both the matrices

$A = (1,2,-7) \in \mathbb{R}^{1,3}$, $\quad B = \begin{pmatrix} 1 \\ 2 \\ -7 \end{pmatrix} \in \mathbb{R}^{3,1}$ can be used to represent the point in space with Cartesian coordinates $x = 1$, $y = 2$, $z = -7$.

(Sometimes commas are used to distinguish between matrices and row vectors, but it is simpler to regard them as the same object).
One can switch between row and column vectors by observing that $A = B^T$ or $B = A^T$ (see TRANSPOSE).
For this reason, the distinction between a row vector and a column vector is often unimportant, and the sets $\mathbb{R}^{1,n}$ and $\mathbb{R}^{n,1}$ can be written more simply $\mathbb{R}^n$, and we can refer to both $A \in \mathbb{R}^3$ and $B \in \mathbb{R}^3$ as "vectors" of length 3. We shall use such vectors to study analytic geometry later in the course.

Whenever we write $\mathbb{R}^n$, the reader is free to use row or column vectors as he/she prefers; if such a choice is not possible, we shall use the other notation to specify either rows or columns.
Actually, vectors tend to be given lower-case names, and a vector of unspecified length n is more likely to be written

$u = (u_1, \dots, u_n)$ or $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$ or $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

Row and column vectors are not merely special cases of matrices. Any matrix can be regarded as an ordered list of both row and column vectors. Given a matrix $A \in \mathbb{R}^{m,n}$, we shall denote its rows (thought of as matrices in their own right) by $r_1, \dots, r_m \in \mathbb{R}^{1,n}$ and its columns $c_1, \dots, c_n \in \mathbb{R}^{m,1}$.

More informally (ignoring parentheses in a way that would be spell disaster in a computer program) we may write

$$\begin{pmatrix} \leftarrow r_1 \rightarrow \\ \cdots \\ \leftarrow r_m \rightarrow \end{pmatrix} = A = \begin{pmatrix} \uparrow & \cdot & \cdot & \uparrow \\ c_1 & \cdot & \cdot & c_n \\ \downarrow & \cdot & \cdot & \downarrow \end{pmatrix}$$

Much of the study of matrices is ultimately be based on one or other of these two descriptions.

* $A \in \mathbb{R}^{1,1} \rightarrow A = (x)$

Addition of matrices

A matrix is much more than an array of data. It's an algebraic object that is subject to operations generalizing the more familiar ones applicable to numbers and vectors.

Proposition
Transposition properties

I) $(A^T)^T = A$, the operation is self-inverse (u can obtain the previous matrix)

II) $(A+B)^T = A^T + B^T$

III) Simmetry, $A = A^T$, they must be square matrices

- $A \in \mathbb{R}^{1,1}$ is always symmetric
- $A \in \mathbb{R}^{2,2} \rightarrow$ to have symmetry $a_{1,2}$ and $a_{2,1}$ must be equal

IV) Skew-symmetry, $A^T = -A$

- $A \in \mathbb{R}^{1,1}$ never skew-symmetric, only 0 matrix
- $A \in \mathbb{R}^{2,2} \rightarrow$ to have skew-sym. it must be $a_{1,2} = -a_{2,1}$ and $a_{11} = a_{22} = 0$

Matrix multiplication

First we define a numerical product between two vectors u, v of the same length. For this it does not really matter whether they are row or column vectors, but for egalitarian purposes we shall suppose that the first is a row vector and the second a column vector. Thus, we consider

$$u = (u_1, \dots, u_n) \in \mathbb{R}^{1,n}, \quad v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^{n,1}$$

Definition
The dot or scalar product of u and v, written $u \cdot v$ is the number $\sum_{i=1}^{n} u_i v_i = u_1 v_1 + \dots + u_n v_n$
The dot product provides the basis for multiplying matrices:

Definition
The product of two matrices A, B is only defined if the number of columns of A equals the number of rows of B. If $A \in \mathbb{R}^{m,n}$ has rows $r_1 \dots r_m$ and $B \in \mathbb{R}^{p,q}$ has columns $c_1 \dots c_q$, it must be $n = p$, then $AB = C_{ij} \in \mathbb{R}^{m,q}$, so it will be the matrix with entries $r_i \cdot c_j$ and size $m \times q$

More explicitly,
$$AB = \begin{pmatrix} \leftarrow r_1 \rightarrow \\ \dots \\ \leftarrow r_m \rightarrow \end{pmatrix} \begin{pmatrix} \uparrow & \cdot & \cdot & \uparrow \\ c_1 & \cdot & \cdot & c_q \\ \downarrow & \cdot & \cdot & \downarrow \end{pmatrix} = \begin{pmatrix} r_1 c_1 & \cdot & \cdot & r_1 c_q \\ \dots & & & \dots \\ r_m \cdot c_1 & \cdot & \cdot & r_m c_q \end{pmatrix}$$

* One should imagine taking each row of A, rotating it and placing it on top of each column of B in turn so as to perform the dot product.

Example
A very special case is the product $r_1 c_1 = (r_1 \cdot c_1)$ of a single row and a column. Strictly speaking, this is a 1×1 matrix, but (again ignoring parentheses) we shall regard it as a number, i.e. the dot product.
With this convention, if $v = (x, y, z)$ then

$$v v^T = (x, y, z) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = x^2 + y^2 + z^2$$

Later, we shall refer to the square root of this quantity as the norm of the vector v (it's the distance from the corresponding point to the origin) By contrast, note that

$$v^T v = \begin{pmatrix} x^2 & xy & xz \\ yx & y^2 & yz \\ zx & zy & z^2 \end{pmatrix} \text{ is a } 3 \times 3 \text{ matrix}$$

An intermediate case of the matrix product is that in which the second factor is a single column $v = (x_1, \dots, x_n)^T \in \mathbb{R}^{n,1}$, so that

$$Av = \begin{pmatrix} r_1 \cdot v \\ \vdots \\ r_m \cdot v \end{pmatrix}$$

In general we can say that $AB = \begin{pmatrix} \uparrow & & \uparrow \\ Ac_1 & \cdot \cdot & Ac_p \\ \downarrow & & \downarrow \end{pmatrix}$ which shows that each column of the product is obtained by premultiplying the corresponding column of B.
The rule for manipulating size can be remembered by the scheme $m \times \underbrace{n \; p}_{} \times q \rightsquigarrow m \times q$ and matrix multipl. defines a mapping $\mathbb{R}^{m,n} \times \mathbb{R}^{p,q} \rightarrow \mathbb{R}^{m,q}$

Proposition
Product properties

I) Matrix product is not commutative, in general $AB \neq BA$

II) $AB = 0 \not\Rightarrow A = 0$ or $B = 0$

Example

$A, B \in \mathbb{R}^{2,2}$

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

$$AB = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \text{* zero matrix}$$
$$\neq$$
$$BA = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}$$

Exercise.
(I) If A is invertible, then so is A^T, and $(A^T)^{-1} = (A^{-1})^T$.
(II) If A, B are invertible then $(AB)^{-1} = B^{-1}A^{-1}$
(III) If A is invertible and $m \in \mathbb{N}$ then $(A^m)^{-1} = (A^{-1})^m$

The inverse can be used to help solve equations involving square matrices. For example, suppose that $AB = C$, where A is an invertible square matrix. Then $B = I_m B = (A^{-1}A)B = A^{-1}(AB) = A^{-1}C$, and we've solved for B in terms of C.

Example
Let $A \in \mathbb{R}^{2,2}$. A direct calculation shows that $A^2 - (a+d)A + (ad-bc)I_2 = 0$.
Assuming there exists A^{-1} such that $AA^{-1} = I_2$ we obtain
$A - (a+d)I_2 + (ad-bc)A^{-1} = 0 \implies (ad-bc)A^{-1} = (a+d)I_2 - A$
we get exactly the same expression for A^{-1} by assuming that $A^{-1}A = I_2$

INVERSE MATRIX
$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \longrightarrow A^{-1} = \begin{pmatrix} -d & b \\ c & -a \end{pmatrix}$$

Determinants

Definition
The quantity $ad-bc$ is called the determinant of the 2×2 matrix. A. It turns out that it is possible to associate to any square matrix $A \in \mathbb{R}^{m,m}$ a number called its determinant, written det A or $|A|$. This number is a function of the components of A and satisfies

Theorem
A is invertible $\iff |A| = ad - bc \neq 0$

We shall explain this result in Part II of the course, but here we give two ways of computing the determin. when $n = 3$ (3×3 matrix)
I) Sarrus rule
II) Expansion

Sarrus (only for 3x3)

Let $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$, then you obtain $|A| = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{matrix}$ by copying down the first 2 columns

The formula of Sarrus asserts that the determinant of A is the sum of the products of entries on the three downward diagonals ↘ minus those on the three upward diagonals ↗.

$|A| = \diagdown + \diagdown + \diagdown - / - / - /$

Expansion

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \longrightarrow |A| = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} *$$

here the 3 mini-det. are constructed from the last two rows of A.

Exercise. Use either of these formulae to prove the following properties for the det. of a 3x3 matrix A:
(i) If one row is multiplied by c so is det A
(ii) $\det(cA) = c^3 \det A$
(iii) If two rows are swapped then det A changes sign
(iv) If one row is a multiple of another then $\det A = 0$
(v) $\det A = \det(A^T)$, so the above statements apply equally to columns.

Determinant of 3x3 matrices can be used to compute mixed products.

Proposition
For vectors
$\vec{u} = u_x\vec{i} + u_y\vec{j} + u_z\vec{k}$
$\vec{v} = v_x\vec{i} + v_y\vec{j} + v_z\vec{k}$
$\vec{w} = w_x\vec{i} + w_y\vec{j} + w_z\vec{k}$ we have ↗ $\vec{u} \times \vec{v} \cdot \vec{w} = \det \begin{pmatrix} u_x & u_y & u_z \\ v_x & v_y & v_z \\ w_x & w_y & w_z \end{pmatrix}$

Exercise. Prove the previous proposition by direct computation.

Example
$$\det \begin{pmatrix} \vec{u} \\ \vec{v} \\ \vec{w} \end{pmatrix} = \begin{matrix} \vec{u} \cdot \vec{v} \times \vec{w} \\ i \cdot j \times k \end{matrix}$$

$$\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1 \cdot \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} - 0 \cancel{\begin{vmatrix} 0 & 0 \\ 0 & 1 \end{vmatrix}} + 0 \cancel{\begin{vmatrix} 0 & 1 \\ 0 & 0 \end{vmatrix}} = 1 \cdot \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1$$

$* \longrightarrow$

Matrix form

Let us begin with an arbitrary linear system of the form

$$\begin{cases} a_{11}x_1 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + \dots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + \dots + a_{mn}x_n = b_m \end{cases}$$

We shall always use
m to denote the number of equations,
n to denote the number of unknowns or variables.
We now see that the notation is tailored to that of matrices; indeed the system can be rewritten in the succinct matrix form

$\boxed{Ax = b}$ where

$A \in \mathbb{R}^{m,n}$ = matrix of coefficient or left-hand coefficient

x = column vector / matrix of unknowns $\in \mathbb{R}^{n,1}$

$b \in \mathbb{R}^{m,1}$ = matrix of right-hand coeff. or matrix of known terms (result)

$$\Longrightarrow A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b^m \end{pmatrix}$$

* The system is homogeneous $\iff b = 0$ is the null vector. A solution of the system is now understood as meaning a column vector x of length n that satisfies $Ax = b$. The problem is to find all such vectors.
Observe that in such matrix form, the left-hand side of each equation is translated into a row of A.
We shall normally solve such a system by operating on the rows of A, but first we show how the inverse matrix can sometimes be used.

Example

Consider the linear system $Ax = b$ and suppose that $m = n$ (SQUARE SYSTEM) and that A is invertible. This means that we can find a matrix A^{-1} s.t. $A^{-1}A = I_n$. Then

$$A^{-1}(Ax) = A^{-1}b \implies \underbrace{(A^{-1}A)}_{I_n}x = A^{-1}b \implies x = A^{-1}b$$ and the system is solved uniquely.

Thus, a linear system with the same number of equations and variables whose associated matrix is invertible has a unique solution. Applying this method to the generic 2×2 system

$$\begin{cases} ax + by = p \\ cx + dy = q \end{cases} \quad \text{gives} \quad \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}\begin{pmatrix} p \\ q \end{pmatrix} = \frac{1}{ad-bc}\begin{pmatrix} dp - bq \\ -cp + aq \end{pmatrix}$$

The solution is neatly expressed as

$$x = \frac{\left|\begin{pmatrix} \boxed{\begin{matrix} p \\ q \end{matrix}} & \begin{matrix} b \\ d \end{matrix} \end{pmatrix}\right|}{\left|\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right|}, \quad y = \frac{\left|\begin{pmatrix} \begin{matrix} a \\ c \end{matrix} & \boxed{\begin{matrix} p \\ q \end{matrix}} \end{pmatrix}\right|}{\left|\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right|}$$

It is a special case of Cramer's rule, whereby each unknows is obtained by substituting a column of A by b, taking the determinant, and then dividing by $\det A$.

Proposition

Cramer's rule

$Ax = b$ • square system • $|A| \neq 0$, then we can find the solution without computing the inverse A^{-1}

Examples

- $\begin{cases} x + y = 0 \\ 2x + y = 0 \end{cases}$ $\quad \overset{A}{\begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix}} \overset{x}{\begin{pmatrix} x \\ y \end{pmatrix}} = \overset{b}{\begin{pmatrix} 0 \\ 0 \end{pmatrix}}$ $\quad |A| = -1 \neq 0$, A is invertible and there's only one sol. $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ we don't need to compute A^{-1}

- $\overset{A}{\begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix}} \overset{x}{\begin{pmatrix} x \\ y \end{pmatrix}} = \overset{b}{\begin{pmatrix} 1 \\ 0 \end{pmatrix}} \rightsquigarrow A^{-1} = (-1)\begin{pmatrix} 1 & -1 \\ -2 & 1 \end{pmatrix} = -\begin{pmatrix} 1 & -1 \\ -2 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 2 & -1 \end{pmatrix}$

  $\rightarrow x = A^{-1}b = A^{-1}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$ unique solution

Solving a homogeneous system

Let us show how ERO's can be used to solve the linear system written before.

$$* \begin{cases} x_1+x_2+2x_3+3x_4=0 \\ 5x_1+8x_2+13x_3+21x_4=0 \\ 34x_1+55x_2+89x_3+144x_4=0 \end{cases}$$

(the choice of these coefficients will keep the arithmetic manageable)

We shall apply ERO's to convert A into a matrix that is roughly triangular, and then solve the resulting system

$$A=\begin{pmatrix}1&1&2&3\\5&8&13&21\\34&55&89&144\end{pmatrix} \overset{r_2-5r_1}{\sim} \begin{pmatrix}1&1&2&3\\0&3&3&6\\34&55&89&144\end{pmatrix} \overset{1/3\, r_2}{\sim} \begin{pmatrix}1&1&2&3\\0&1&1&2\\34&55&89&144\end{pmatrix} \overset{r_3-34r_1}{\sim} \begin{pmatrix}1&1&2&3\\0&1&1&2\\0&21&21&42\end{pmatrix} \overset{1/21\, r_3}{\sim} \begin{pmatrix}1&1&2&3\\0&1&1&2\\0&1&1&2\end{pmatrix} \overset{r_3-r_2}{\sim} \begin{pmatrix}1&1&2&3\\0&1&1&2\\0&0&0&0\end{pmatrix}$$

On the left, we jot down (in abbreviated form) the operations used. It is not essential to do this, provided the operations are carried out one at time; errors occur when one tries to be too ambitious!
It follows from the last matrix that * has the same solutions as the system

$$\begin{cases} x_1+x_2+2x_3+3x_4=0 \\ x_2+x_3+2x_4=0 \end{cases}$$

But one can see at a glance how to solve this; we can assign any values to x_3 and x_4 which will then determine x_2 (from the second equation) and then x_1 (from the first). Suppose that we set $x_3=s$ and $x_4=t$ (it's a good idea to use different letters to indicate free variables); then

$x_2=-s-2t \implies x_1=-(-s-2t)-2s-3t=-s-t$

The general solution is $(x_1,x_2,x_3,x_4)=(-s-t,-s-2t,s,t)$, or in column form

$$x=\begin{pmatrix}-s-t\\-s-2t\\s\\t\end{pmatrix}=s\begin{pmatrix}-1\\-1\\1\\0\end{pmatrix}+t\begin{pmatrix}-1\\-2\\0\\1\end{pmatrix}$$ where s,t are arbitrary.

The set of solutions is therefore

$\mathcal{L}\{u,v\}$, where $u=\begin{pmatrix}-1\\-1\\1\\0\end{pmatrix}$, $v=\begin{pmatrix}-1\\-2\\0\\1\end{pmatrix}$

- we shall see that the solution set of any homogeneous system is always a linear combination of this type.

Exercise. Compute $u+v$, and explain why this is also a solution.

Equivalence relations.

Applying ERO's produces a natural relation on the set of matrices of any fixed size.

Definition

A relationship $\sim$ between elements of a set is called an **equivalence relation** if

(E1) $A\sim A$ is always true $\longrightarrow$ $r_i \to 1\cdot r_i$
(E2) $A\sim B$ always implies $B\sim A$ $\longrightarrow$ we use a ERO to pass from A to B and viceversa
(E3) $A\sim B$ and $B\sim C$ implies $A\sim C$ $\longrightarrow$ we use a ERO to pass from A to B and another from B to C

Observe that these three conditions are satisfied by equality =. On the set of real numbers, "having the same absolute value" is an equivalence relation, but $\leq$ is not. But we are more interested in sets of matrices.

Definition

From now on, we write $A\sim B$ to mean that B is a matrix obtained by applying one or more ERO's in succession to A, i.e. $A\sim B \iff$ B is obtained by using a sequence of ERO's (by A), so el. row operations.

Proposition

This does define an equivalence relation, and if $A\sim B$ we say that A and B are row equivalent.
$A\sim B$ then $Ax=0 \iff Bx=0$

PROOF

By definition $A\sim B$ and $B\sim C$ imply that $A\sim C$. Obviously $A\sim A$ (take (i) with $a=0$ or (ii) with $c=1$, both of which are permitted).
The condition (E2) is less obvious. But each of the three operations is invertible; it can be undone by the same type of operation. For example, $r_1\to r_1-ar_2$ by $r_1\to r_1+ar_2$. So if $A\sim B$ then we can undo each ERO in the succession one at time, and $B\sim A$. QED

Example

If B is a matrix which has the same rows as A but in a different order, then $A\sim B$. This is because any permutation of the rows can be obtained by a succession of transpositions, i.e. ERO's of type (iii).
For example, let

$$A=\begin{pmatrix}1&2&3\\4&5&6\\7&8&9\end{pmatrix},\quad B=\begin{pmatrix}4&5&6\\1&2&3\\7&8&9\end{pmatrix},\quad C=\begin{pmatrix}4&5&6\\7&8&9\\1&2&3\end{pmatrix}$$

Example

Consider again

$$A=\begin{pmatrix} 1 & 1 & 2 & 3 \\ 5 & 8 & 13 & 21 \\ 34 & 55 & 89 & 144 \end{pmatrix}$$

we have already step-reduced it to

$$B=\begin{pmatrix} \boxed{1} & 1 & 2 & 3 \\ 0 & \boxed{1} & 1 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

To super-reduce B, we merely perform the operation $x_1 \mapsto x_1 - x_2$ so as to obtain

$$C=\begin{pmatrix} \boxed{1} & 0 & 1 & 1 \\ 0 & \boxed{1} & 1 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

When a matrix is superreduced, it is possible to read off immediately the solution to the original system.
In the above example, we get the equations

$$\begin{cases} x_1 + x_3 + x_4 = 0 \\ x_2 + x_3 + 2x_4 = 0 \end{cases}$$ whence $x_1 = -s-t$, $x_2 = -s-2t$ slightly more effectively than before.

Linear independence

Converting a matrix to row echelon form has the effect of making the transformed equations "independent", we formulate this notion.

Definition

Let $\{u_1, \dots, u_k\}$ be a finite subset of $\mathbb{R}^m$ (meaning either in $\mathbb{R}^{1,m}$ or $\mathbb{R}^{m,1}$). The set is called **linearly independent** (LI) if the equation $x_1u_1 + \dots + x_ku_k = 0$ admits only the trivial solution $x_1 = \dots = x_k = 0$

One often says $u_1, \dots, u_k$ are linearly independent, though strictly speaking being LI is a property of a set or list and not its individual elements. The order of the elements is immaterial, and any duplication prevents the list from being LI.

- A singleton set $\{v\}$ is LI $\iff v \neq 0$, and no set that contains 0 can be LI. A set $\{u, v\}$ is LI $\iff$ neither vector is a multiple (including zero times) the other. More generally,

LEMMA

A set $\{u_1 \dots u_k\}$ is LI $\iff$ no one element in it can be expressed as a linear combination of the others.

PROOF

Suppose that the set is LI, but that one of the elements is a LC of the others. For sake of argument, suppose that $u_k \in \mathcal{L}\{u_1, \dots, u_{k-1}\}$, so that $u_k = a_1u_1 + \dots + a_{k-1}u_{k-1}$ for some $a_i \in \mathbb{R}$. But then $a_1u_1 + \dots + a_{k-1}u_{k-1} + (-1)u_k = 0$ contradicting *. The converse is similar. QED

Proposition

Let B be a step-reduced matrix. Then its nonzero rows are LI.

PROOF we refer to the example

$$\begin{array}{c} a_1 \\ a_2 \\ a_3 \\ a_4 \end{array}\begin{pmatrix} 0 & 2 & 3 & 0 & 4 & 6 \\ 0 & 0 & 7 & 3 & 8 & 1 \\ 0 & 0 & 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} = B$$

assuming a linear relation $a_1r_1 + \dots + a_kr_k = 0$ between the nonzero rows, perform the addition column by column, high school fashion. The marker of r_1 is the only nonzero entry in its column, so $2a_1 = 0$. Passing to the next marker, $3a_1 + 7a_2 = 0$, so $a_2 = 0$ and so on. QED

Exercise. The same conclusion holds if B satisfies just the first condition (H1), namely that there is at most one marker in each column. Indeed, once B satisfies (H1) we can make it step-reduced without changing the unordered set of rows.

The columns of the matrix in (2) are not LI. Even if we forget about c_1 and c_4, we have $c_6 = 2c_5 - \frac{15}{7}c_3 - kc_2$, for some $k \in \mathbb{R}$. This illustrates the

Proposition

Let B be a step-reduced matrix. A column c_j of B is unmarked $\iff$ it is a LC of the previous marked columns (or null if $j = 1$).
The point is that we can always express an unmarked column as a LC of the previous marked ones by finding the coefficients one at a time starting from the bottom

* **COROLLARY**

If B is step-reduced then its marked columns are LI.

Thus, the markers of a step-reduced matrix "mark out" an independent set of both rows and columns. Whilst there may be unmarked columns in any position, row reduction ensures that all the unmarked rows are null. If every column of a step reduced matrix C is marked, then the set of columns is LI

Proposition

$A \sim B$

1) If $A \xrightarrow{(i)} B$ then $|B| = |A|$
2) If $A \xrightarrow{(ii)} B$ then $|B| = c|A|$
3) If $A \xrightarrow{(iii)} B$ then $|B| = -|A|$

Determinant $\longleftrightarrow$ mixed product

$$A = \begin{pmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix}_{3\times 3} = \begin{pmatrix} \vec{u} \\ \vec{v} \\ \vec{w} \end{pmatrix} \qquad |A| = \vec{u}\cdot\vec{v}\times\vec{w}$$

PROOF

Elementary matrices + Binet's theorem for $m=3$ mixed product

1) $x_1 \to x_1 + 2x_2$

$$A = \begin{pmatrix} \vec{u} \\ \vec{v} \\ \vec{w} \end{pmatrix} \rightsquigarrow B = \begin{pmatrix} \vec{u}+2\vec{v} \\ \vec{v} \\ \vec{w} \end{pmatrix} \to |B| = (\vec{u}+2\vec{v})\cdot(\vec{v}\times\vec{w}) = \underbrace{\vec{u}\cdot(\vec{v}\times\vec{w})}_{|A|} + \underbrace{2\vec{v}\cdot(\vec{v}\times\vec{w})}_{0} = |A| + 0 = |A|$$

2) $x_2 \to 3x_2$

$$A = \begin{pmatrix} \vec{u} \\ \vec{v} \\ \vec{w} \end{pmatrix} \rightsquigarrow B = \begin{pmatrix} \vec{u} \\ 3\vec{v} \\ \vec{w} \end{pmatrix} \to |B| = \vec{u}\cdot((3\vec{v})\times\vec{w}) = \vec{u}\cdot(3(\vec{v}\times\vec{w})) = 3(\vec{u}\cdot\vec{v}\times\vec{w}) = 3|A|$$

3) $x_2 \longleftrightarrow x_3$

$$A = \begin{pmatrix} \vec{u} \\ \vec{v} \\ \vec{w} \end{pmatrix} \rightsquigarrow B = \begin{pmatrix} \vec{u} \\ \vec{w} \\ \vec{v} \end{pmatrix} \to |B| = \vec{u}\cdot\vec{w}\times\vec{v} = -(\vec{u}\cdot\vec{v}\times\vec{w}) = -|A|$$

Definitions

- A matrix is said **diagonal** if the nonzero elements are on the main diagonal

diagonal $\iff a_{ij} = 0 \quad i \neq j \quad ; \quad A = (a_{ij})_{1\le i,j\le m}$

$$A = \begin{pmatrix} c_1 & & 0 \\ & c_2 \ddots & \\ 0 & & c_n \end{pmatrix} \quad D \text{ or } \Delta$$

$m=2$

$$D = \begin{pmatrix} c_1 & 0 \\ 0 & c_2 \end{pmatrix} \quad |D| = c_1 \cdot c_2$$

$m=3$

$$D = \begin{pmatrix} c_1 & 0 & 0 \\ 0 & c_2 & 0 \\ 0 & 0 & c_3 \end{pmatrix} \quad |D| = c_1 \begin{vmatrix} c_2 & 0 \\ 0 & c_3 \end{vmatrix} = c_1 \cdot c_2 \cdot c_3$$

for any m, $D \in \mathbb{R}^{m,m}$

$$D = \begin{pmatrix} c_1 & & 0 \\ & c_2 \ddots & \\ 0 & & c_n \end{pmatrix} \to |D| = c_1 \begin{vmatrix} c_2 \ddots & 0 \\ 0 & c_n \end{vmatrix} = c_1 \cdot c_2 \cdot c_3 \cdot \ldots \cdot c_n$$

- A matrix of the type $T = (t_{ij})_{1\le i,j\le m}$ is said **triangular**

$$\begin{pmatrix} \diagdown & * \\ 0 & \diagdown \end{pmatrix}$$ UPPER TRIANGULAR $t_{ij} = 0 \quad i > j$

$m=2$

$$T = \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \to |T| = ac$$

$m=3$

$$T = \begin{pmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{pmatrix} \to |T| = f \begin{vmatrix} a & b \\ 0 & d \end{vmatrix} = adf$$

for any m

$$T = \begin{pmatrix} c_1 & & * \\ & c_2 \ddots & \\ 0 & & c_n \end{pmatrix} \to |T| = c_1 \cdot c_2 \cdot \ldots \cdot c_n$$

Elementary matrices

(i) $\begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix} \quad x_1 \to x_1 + cx_2$

(ii) $\begin{pmatrix} c & 0 \\ 0 & 1 \end{pmatrix} \quad x_1 \to cx_1$

(iii) $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad x_1 \longleftrightarrow x_2$

Example

$$\begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} (1+4c) & (2+5c) & (3+6c) \\ 4 & 5 & 6 \end{pmatrix}$$

* Multiply to the left $\to$ elementary row operation
Multiply to the right $\to$ elementary column operation

It follows that we can assign the unmarked variables arbitrarily and solve uniquely for each of the marked variables in terms of them. QED

In the light of the procedure above, the unmarked variables are called **free variables**, and in the solution it is good practice to give them new names such as $s, t, u \dots$ or $t_1, t_2, t_3 \dots$. The conclusion is traditionally expressed by the statement

* If $r(A) = r(A|b) = r$, then the linear system has ∞^{n-r} solutions.
This is a useful way of recording the result that can be understood as follows. The actual number m of equations is irrelevant; what is important is the number of LI or effective equations, and this is the rank r. Each effective equation allows us to express one of the n variables in terms of the others, so we end up with $n-r$ free variables or parameters.

Inversion by reduction

Having introduced the augmented matrix, we can apply similar techniques to solve matrix equations of the type $AX = B$ where X and B are matrices rather than just column vectors.
* A special case is $AX = I_n$ $\quad A, X \in \mathbb{R}^{n,n}$, whose solution X (if it exists) is necessarily A^{-1}.
As a consequence,

Proposition
If $A \in \mathbb{R}^{n,n}$ is invertible then the unique super-reduced form of $(A|I_n)$ is $(I_n|A^{-1})$

Here is an example

$$(A|I_5) = \left(\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ 3 & 5 & 8 & 0 & 1 & 0 \\ 13 & 21 & 35 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 2 & 2 & -3 & 1 & 0 \\ 13 & 21 & 35 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 2 & 2 & -3 & 1 & 0 \\ 0 & 8 & 9 & -13 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 2 & 2 & -3 & 1 & 0 \\ 0 & 0 & 1 & -1 & -4 & 1 \end{array}\right)$$

$$\sim \left(\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & -1 & 9 & -2 \\ 0 & 0 & 1 & -1 & -4 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 1 & 0 & 3 & 8 & -2 \\ 0 & 2 & 0 & -1 & 9 & -2 \\ 0 & 0 & 1 & -1 & -4 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 1 & 0 & 3 & 8 & -2 \\ 0 & 1 & 0 & -1/2 & 9/2 & -1 \\ 0 & 0 & 1 & -1 & -4 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 7/2 & 7/2 & -1 \\ 0 & 1 & 0 & -1/2 & 9/2 & -1 \\ 0 & 0 & 1 & -1 & -4 & 1 \end{array}\right)$$

confirming the inverse found in L2. The matrices on the right (even) act as a book-keeping of the ERO's which there's no need for us to record separately.

The reason the method works is that each of the three types of ERO's is actually achieved by pre-multiplying A by a suitable invertible matrix E_i. For example, the first two are achieved by

$$E_1 = \begin{pmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad ; \quad E_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -13 & 0 & 1 \end{pmatrix} \quad \text{respectively.}$$

By the time we have finished, the last line tells us that
$E_7 E_6 E_5 E_4 E_3 E_2 E_1 A = I_5 \implies A^{-1} = E_7 E_6 E_5 E_4 E_3 E_2 E_1 I_5$

* COROLLARY
If $A \sim B$ there exists an invertible matrix $E \in \mathbb{R}^{m,m}$ such that $EA = B$.

• SEE FURTHER EXERCISES!

Here, x_0 is a particular solution of the inhomogeneous equation $Ax=b$; the difference of any two such solutions x_0, x_1 belongs to $\ker A$ because $A(x_1-x_0)=0$. → important to find the kernel of a non-hom system

Subspaces defined by a matrix

Given a matrix A, two separate collections of vectors are starting us in the face: the rows $r_1, \dots, r_m \in \mathbb{R}^{1,n}$ of A and the columns $c_1, \dots, c_n \in \mathbb{R}^{m,1}$ of A.
These give rise to two respective subspaces that complement the one already defined in (1).

Definition

With the notation above
(i) the row space of A, denoted $\mathrm{Row}\,A$, is $\mathcal{L}\{r_1, \dots, r_m\} \subset \mathbb{R}^{1,n}$ and
(ii) the column space of A, denoted $\mathrm{Col}\,A$, is $\mathcal{L}\{c_1, \dots, c_n\} \subset \mathbb{R}^{m,1}$
More informally, $\mathrm{Row}\,A$ is a subspace of $\mathbb{R}^n$, whereas $\mathrm{Col}\,A$ is a subspace of $\mathbb{R}^m$.

Each row of A corresponds to an equation of the linear system with augmented matrix $(A|0)$. We already know that there are many ways to transform this system into an equivalent one with the same solutions. The next result formalizes that it is the row space $\mathrm{Row}\,A$ (rather than the individual rows of A) that determines the solution space $\ker A$.

LEMMA

$\ker A = \{x \in \mathbb{R}^{n,1} : rx = 0 \text{ for all } r \in \mathrm{Row}\,A\}$

PROOF

Since

$Ax = \begin{pmatrix} r_1 \cdot x \\ \vdots \\ r_m \cdot x \end{pmatrix}$, x belongs to $\ker A \iff r_i x = 0$ for all i. This implies that $rx=0$ for any $r \in \mathrm{Row}\,A$ since such an r is a LC of the rows $r_1, \dots, r_m$. Conversely, if $rx=0$ for all $r \in \mathrm{Row}\,A$ then certainly $r_i x = 0$ for all i, and so $Ax=0$. QED

Recall the notion of row equivalence. It is easy to see that
1) $A \sim B \Rightarrow \mathrm{Row}\,A = \mathrm{Row}\,B$
For if $A \sim B$, each row of B is obtained from A using ERO's, and $\mathrm{Row}\,B \subseteq \mathrm{Row}\,A$. But the process is reversible: $B \sim A$ and $\mathrm{Row}\,A \subseteq \mathrm{Row}\,B$. It follows from the Lemma that
2) $A \sim B \Rightarrow \ker A = \ker B$,
confirming something we already know: if two matrices, A and B, are row equivalent then the associated homogeneous systems have the same solutions.

We can complete these observations by the next result, which is easily memorized

Theorem

Let A, B be two matrices of the size $m \times n$. The following are equivalent:
(i) $A \sim B$, i.e. A and B are related by ERO's. *
(ii) $\mathrm{Row}\,A = \mathrm{Row}\,B$.
(iii) $\ker A = \ker B$.

This is especially relevant in the case in which B is a step-reduced matrix obtained by applying ERO's to A. Notice that the statement $A \sim B$ forces the matrices to have the same size - one could relax this requirement (and retain the Theorem's validity) by introducing a fourth ERO, that of deleting null rows.

Warning $A \sim B$ does not imply that $\mathrm{Col}\,A = \mathrm{Col}\,B$; to see this reduce the matrix (8)

PROOF

Assume (iii). We already know that this implies (ii). Applying ERO's does not change the row space (2), so we may assume that A and B are step-reduced. Since the systems $Ax=0$, $Bx=0$ have the same solutions, we know that the markers of A and B occur in the same positions, and we may as well suppose that neither has a null row.
To deduce (i), one uses an exchange technique that we shall describe by means of an example with $n=5$. Let a_i be the rows of A and b_i those of B; the idea is to slowly replace the former by the latter by the process illustrated:

$$A' = \begin{pmatrix} \leftarrow b_1 \rightarrow \\ \leftarrow b_2 \rightarrow \\ \hline \leftarrow a_1 \rightarrow \\ \leftarrow a_3 \rightarrow \\ \leftarrow a_4 \rightarrow \end{pmatrix}, \quad A'' = \begin{pmatrix} \leftarrow b_1 \rightarrow \\ \leftarrow b_2 \rightarrow \\ \leftarrow b_3 \rightarrow \\ \hline \leftarrow a_1 \rightarrow \\ \leftarrow a_3 \rightarrow \end{pmatrix}$$

* $A \sim B$

$$\begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} \sim \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \qquad \begin{array}{l} b_1 \in \mathrm{Row}(A) \\ b_2 \in \mathrm{Row}(A) \end{array} \qquad \underset{\mathcal{L}\{b_1 \dots b_m\}}{\overset{\mathrm{Row}(B) \le \mathrm{Row}(A)}{\shortparallel}}$$

Suppose that we have already shown that A is row equivalent to A', in which a_2, a_5 have been replaced by b_1, b_2. Since $b_3 \in \mathrm{Row}\,B = \mathrm{Row}\,A = \mathrm{Row}\,A'$ (2), b_3 is a LC of the rows of A'. In this LC, one of a_1, a_3, a_4 must figure with a nonzero coefficient, otherwise b_1, b_2, b_3 would not be LI. If (say) a_4 features then A'' can be obtained from A by a sequence of ERO's, and $A \sim A' \sim A''$. We can repeat the process and conclude at the end that $A \sim B$.
To see that (iii) implies (ii), we need the formula $\mathrm{Row}\,A = \{r \in \mathbb{R}^{1,n} : rx = 0 \text{ for all } x \in \ker A\}$ that is the counterpart of the lemma. It is obvious that $\mathrm{Row}\,A$ is contained in the right-hand side, though the inequality is best proved using the notion of dimension that will be discussed in the next lecture. QED

$$W = \mathcal{L}\left\{\begin{pmatrix}1\\2\\3\\4\\5\end{pmatrix},\begin{pmatrix}6\\7\\8\\9\\10\end{pmatrix},\begin{pmatrix}11\\12\\13\\14\\15\end{pmatrix},\begin{pmatrix}16\\17\\18\\19\\20\end{pmatrix}\right\} \subset \mathbb{R}^{5,1}$$

We convert the columns into rows and reduce the 5x4 matrix:

$$\begin{pmatrix}1&2&3&4&5\\6&7&8&9&10\\11&12&13&14&15\\16&17&18&19&20\end{pmatrix} \sim \begin{pmatrix}1&2&3&4&5\\0&-5&-10&-15&-20\\0&-10&-20&-30&-40\\-15&-30&-45&-60&-75\end{pmatrix} \sim \begin{pmatrix}\boxed{1}&2&3&4&5\\0&\boxed{-5}&-10&-15&-20\\0&0&0&0&0\\0&0&0&0&0\end{pmatrix} \sim \begin{pmatrix}\boxed{1}&2&3&4&5\\0&\boxed{1}&2&3&4\\0&0&0&0&0\\0&0&0&0&0\end{pmatrix} \sim \begin{pmatrix}\boxed{1}&0&-1&-2&-3\\0&\boxed{1}&2&3&4\\0&0&0&0&0\\0&0&0&0&0\end{pmatrix}$$

The fact that W has basis of two elements is already clear after two steps, and the super-reduction was unnecessary. But we can now give three useful bases:

$$W = \mathcal{L}\left\{\begin{pmatrix}1\\2\\3\\4\\5\end{pmatrix},\begin{pmatrix}0\\1\\2\\3\\4\end{pmatrix}\right\} = \mathcal{L}\left\{\begin{pmatrix}-1\\0\\1\\2\\3\end{pmatrix},\begin{pmatrix}0\\1\\2\\3\\4\end{pmatrix}\right\} = \mathcal{L}\left\{\begin{pmatrix}1\\2\\3\\4\\5\end{pmatrix},\begin{pmatrix}6\\7\\8\\9\\10\end{pmatrix}\right\}$$

Dimension

First we reassure ourselves that any subspace of $\mathbb{R}^n$ has a basis.

Theorem

Let V be a subspace of $\mathbb{R}^n$ which is not null. Then V has a basis consisting of at most n elements.

PROOF

By assumption, V contains a nonzero vector u_1. Then either $V = \mathcal{L}\{u_1\}$ or we can choose $u_2 \in V \setminus \mathcal{L}\{u_1\}$.
In the latter case u_2 cannot be a multiple of u_1, and either $V = \mathcal{L}\{u_1, u_2\}$ or we can choose $u_3 \in V \setminus \mathcal{L}\{u_1, u_2\}$.
In the latter case $\{u_1, u_2, u_3\}$ is LI since any linear relation could be used to express u_3 as a LC of u_1, u_2 and the procedure can be continued. The set of elements $u_1, u_2, \ldots$ that we have selected at each stage is LI for the same reason. But we know that it is impossible to have more than n elements of $\mathbb{R}^n$ that are LI. So the process must stop. QED

Recall that if the rows of a matrix $A \in \mathbb{R}^{m,n}$ are LI then $r(A) = m$. A deeper fact we have seen is that if A, B are two matrices of the same size with $\text{Row}A = \text{Row}B$, then $A \sim B$.

*** COROLLARY**

Let V be a subspace of $\mathbb{R}^n$ that is not null. Any two bases of V have the same number of elements.

PROOF

Given two bases $\{a_1, \ldots, a_\ell\}$ and $\{b_1, \ldots, b_m\}$ with $\ell < m$, make them the rows of matrices A and B, adding null rows to A so that A, B both have size $m \times n$:

$$A = \begin{pmatrix}\leftarrow a_1 \rightarrow\\ \leftarrow a_2 \rightarrow\\ \cdots\\ \leftarrow a_\ell \rightarrow\\ \leftarrow 0 \rightarrow\end{pmatrix}, \quad B = \begin{pmatrix}\leftarrow b_1 \rightarrow\\ \leftarrow b_\ell \rightarrow\\ \cdots\\ \leftarrow b_m \rightarrow\end{pmatrix}$$

The ranks of A, B are ℓ, m respectively. But since both sets of rows generate the same subspace V, we must have $\text{Row}A = \text{Row}B$. From the earlier fact $A \sim B$, and so $r(A) = r(B)$, a contradiction. QED

The number of elements in a basis is called the dimension of V, and we write it as $\dim V$. For a subspace of $\mathbb{R}^n$, we know that $\dim V \leq n$. Thus,

$$V = \mathcal{L}\underbrace{\{u_1, \ldots, u_k\}}_{LI}, \quad k = \dim V$$

If $V = \{0\}$ then we set $\dim V = 0$, and (by convention) declare $\emptyset$ to be a basis.

Our final big result before turning to more geometrical applications is

Theorem

For any matrix A of size $m \times n$, $\dim(\text{Row}A) = r(A)$, $\dim(\text{Ker}A) = n - r(A)$

PROOF

We may suppose that A is step-reduced as none of $\text{Row}A$, $\text{Ker}A$, $r(A)$ changes under ERO's. The nonzero rows of A then form a basis of $\text{Row}A$, whence the first equality.
The second equality is then a restatement of (RC2), whereby the solutions of the homogeneous system $Ax = 0$ depend on $n - r(A)$ free parameters. More precisely, a basis of the $\text{Ker}A$ is given by the solutions of the form $x = (a_1, \ldots, a_{j-1}, -1, 0, \ldots, 0)^T$, of which there's one for each unmarked column. QED

We have also stated that the marked columns of B (of which there are $r(B)$ in number) form a basis of $\text{Col}B$, though the latter does in general change with ERO's. But the columns of two row equivalent matrices satisfy the same linear relations, so the same columns will form a basis of $\text{Col}A$. Thus,
$\dim(\text{Col}A) = r(A)$
→ BASES OF THE IMAGE

*** COROLLARY**

$r(A) = r(A^T)$ for any matrix A.

Since the matrices

$A=\begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{pmatrix}$, $A^+=\left(\begin{array}{ccc|c} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \end{array}\right)$ associated to the prev. system both have rank 2, it follows from

(RC2) that the general solution of the system will depend on one parameter, and can be written in one of the equivalent ways:

cause we have only two equations and three unknowns

$(x,y,z)=(x_0+tp,\ y_0+tq,\ z_0+tr)$,

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix}=\begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix}+t\begin{pmatrix} p \\ q \\ r \end{pmatrix} \bullet$$

$$v = v_0 + tp \qquad p\neq 0$$

The last equation asserts that $v-v_0$ is parallel to the fixed vector p. The equations therefore determine a straight line ℓ that passes through the point (x_0,y_0,z_0) with position vector v_0 and direction p. Any one of $\bullet$ is called the parametric equation of the line.
Because ℓ lies in both planes, it is perpendicular to both n_1, n_2. Thus,

LEMMA
p is a multiple of $n_1\times n_2$ (it's $\perp$ to both n_1, n_2)

* There are therefore two ways to find the equation of ℓ:
(a) Solve the system by super-reducing A^+;
(b) Compute $n_1\times n_2$ and then find a particular solution of the system perhaps by setting $z=0$.

We illustrate these two approaches next.

IMPORTANTE
Example. We shall find the parametric equation of the line ℓ: $\begin{cases} x+y+z=1 \\ x+2y+3z=4 \end{cases}$

Method (a) - The augmented matrix of the system is

$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{array}\right)\sim\left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{array}\right)\sim\left(\begin{array}{ccc|c} \boxed{1} & 0 & -1 & -2 \\ 0 & \boxed{1} & 2 & 3 \end{array}\right)$ whence $(x,y,z)=(-2+t,\ 3-2t,\ t)^\bullet$

Method (b) - The direction of ℓ is given by $p=(1,1,1)\times(1,2,3)=(1,-2,1)$. To find one point v_0 on the line, we set $z=0$ so that
$x+y=1$, $x+2y=4 \Rightarrow y=3$, $x=-2$
Therefore the equation is
$x=v_0+tp=(-2,3,0)+t(1,-2,1)$
By luck, this is the same as $\bullet$, though the parametric equation of a line is not unique! so the different methods will not necessarily give identical equations.

SEE FURTHER EXERCISES!

NOTES 12 - Vector Spaces

The theory of linear combinations, linear independence, bases and subspaces that we have studied in relation to $\mathbb{R}^n$ can be generalized to the more general study of vector spaces. Any subspace of $\mathbb{R}^n$ (including $\mathbb{R}^n$ itself) is an example of vector space, but there are many others including sets of matrices, polynomials and functions.

Motivation
A subspace of $\mathbb{R}^n$ is the prime example of a vector space, but there are a number of reasons for discussing the general definition, namely
(i) to emphasize aspects of the theory that do not depend upon the choice of a specific basis,
(ii) to allow the use of scalars that are different from real numbers,
(iii) to extend the theory to function spaces of infinite dimensions.
We shall explain each of these points in turn.

(i) The whole description of $\mathbb{R}^n$ is modelled on the existence of its canonical basis. To be specific, consider $\mathbb{R}^{n,1}$ and let e_j denote the jth column of the identity matrix I_n. Then a typical element of $\mathbb{R}^{n,1}$ is given by

$$v=\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}=x_1e_1+x_2e_2+\dots+x_ne_n,$$ and is represented by its coefficients relative to the basis $\{e_1,\dots,e_n\}$.

But when we wish to describe subspaces of $\mathbb{R}^{n,1}$ there's a need to work with other bases. In fact any subspace of $\mathbb{R}^n$ is a vector space in its own right. In general it is important to be able to change basis; in this way, the abstract concept of vector space comes into its own.

(ii) The 'scalars' that are used to multiply vectors in the definition of a vector space need not be real numbers. The set of scalars is required to be what is called a field, of which $\mathbb{R}$ is only one example.

PROOF

There's a way to define the basic operations, using a rule that works for any functions. Namely, we set
$(p_1+p_2)(x) = p_1(x)+p_2(x)$
$(ap)(x) = ap(x)$, $a \in F$
Using the polynomial, it is obvious that the sum of two pol. is a pol. and that the product of a polynomial with a scalar is a pol. In practice, it is just a matter of applying the operations coefficient-wise, as in the example
$(1+x)^2 + 3(1+x+\frac{1}{2}x^2+\frac{1}{6}x^3) = 4+5x+\frac{5}{2}x^2+\frac{1}{2}x^3$, representing a LC of two elements in $\mathbb{R}_3[x]$. QED

The previous proposition is a special case of the

Proposition

Let V be a vector space, and let A be any non-empty set. Then the prev. rules make the set of all mappings $f: A \to V$ into a vector space.

More about fields

We shall not give the formal definition of a field. But it is a set F that satisfies the rules of a vector space, in which we are allowed to take the set of scalars to be the same set F. Multiplication between scalars and vectors therefore becomes a multiplication between elements of F s.t.
$1a = a$, $a \in F$ and 1 is the multiplicative identity or unit. The multiplication is required to be commutative, so that $ab = ba$, $a,b \in F$•, and every nonzero element $a \in F$ must have a multiplicative inverse, written a^{-1}, satisfying $aa^{-1} = 1$

Every field must contain at least two elements: the additive identity (usually written 0) and the multiplicative identity (written 1). If there are no other elements, we obtain $B = \{0,1\}$ that is a field with the operations
$0+0=0$, $0+1=1=1+0$, $1+1=0$, $00=0$, $01=0=10$, $11=1$

Example Here is an example of a field F with 4 elements. It will be defined as a vector space over a simpler field, namely B. The set F consists of all linear combinations
$b_1f_1 + b_2f_2$, $b_1, b_2 \in B$ in which we decree that f_1, f_2 are independent. Although b_1, b_2 are arbitrary, there are only two choices for each. We can therefore list all four elements of F as row vectors
$(0,0) = 0f_1 + 0f_2 = 0$
$(1,0) = 1f_1 + 0f_2 = f_1$
$(0,1) = 0f_1 + 1f_2 = f_2$
$(1,1) = 1f_1 + 1f_2 = 1$
(On the right, we have avoided boldface to emphasize that the elements are to be treated like numbers, not vectors). Multiplication is carried out component-wise, using the operations of B.
The reason for also calling the last element 1 is that $(1,1)(a,b) = (1a,1b) = (a,b)$ for $a,b \in B$.
The full multiplication table for F is symmetric because of •

$\cdot$	0	f_1	f_2	1
0	0	0	0	0
f_1	0	f_1	1	f_1
f_2	0	1	f_2	f_2
1	0	f_1	f_2	1

If p is a prime number, the set $\{0,1,2,\dots,p-1\}$ with addition and multiplication modulo p (clockface arithmetic) becomes a field with exactly p elements. Applying the construction of the example with $f_1,\dots,f_k$ in place of f_1, f_2, shows that there is a field with p^k elements for any positive integer k. It turns out that this is essentially the only field with p^k elements. Moreover, any finite field has p^k elements for some prime number $p \geq 2$ and integer $k \geq 1$.

SEE FURTHER EXERCISES!

* **Definition**

The trace of a matrix $A = \sum_1^m a_{ij}$ is the sum of its diagonal elements

$A \in \mathbb{R}^{m \times m} \to \begin{pmatrix} a_{11} & & \\ & a_{22} & \\ & & \ddots \; a_{mm} \end{pmatrix}$

Consider two matrices A and B, we can say that
(i) $Tr(A+B) = Tr(A) + Tr(B)$
(ii) $Tr(aB) = a\,Tr(B)$
* Tr is linear

Example

Equation (3) effectively defines a linear mapping
$f: \mathbb{R}^{1,6} \to \mathbb{R}^{2,3}$, for which $f(a_1,\dots,a_6) = \begin{pmatrix} a_5 & a_3 & a_1 \\ a_6 & a_4 & a_2 \end{pmatrix}$

Here we've used the notation fv in place of $f(v)$ to avoid double parentheses. It is easy to check the conditions (LM1) and (LM2); the reason they hold is that each matrix component on the right is a linear comb. of the coordinates on the left. By contrast, neither of the following mapping is linear:

$g(a_1,\dots,a_6) = \begin{pmatrix} a_5+1 & a_3 & a_1 \\ a_6 & a_4 & a_2 \end{pmatrix}$, $h(a_1,\dots,a_6) = \begin{pmatrix} a_5 & (a_3)^2 & a_1 \\ a_6 & a_4 & a_2 \end{pmatrix}$

Let $f: A \to B$ be an arbitrary mapping between two sets. Recall that the image of f,
$\operatorname{Im} f = \{f(a) : a \in A\}$,
denotes the subset of B consisting of those elements 'gotten' from A. Also, given $b \in B$, its inverse image
$f^{-1}(b) = \{a \in A : f(a) = b\}$
is the subset of A consisting of all those elements that map to b. Then f is said to be
(i) surjective or onto if $\operatorname{Im} f = B$
(ii) injective or one-to-one if $f(a_1) = f(a_2) \Rightarrow a_1 = a_2$
Thus f is onto $\iff f^{-1}(b)$ is nonempty for all $b \in B$. If f is both surjective and injective then it is called bijective. This means that there exists a well defined inverse mapping $f^{-1}: B \to A$ so that $f^{-1} \circ f$ and $f \circ f^{-1}$ are identity mappings.

Example

Let n be a positive integer. Then $f(x) = x^n$ defines a bijection $\mathbb{R} \to \mathbb{R} \iff n$ is odd; if n is even f is neither inj nor surj.; f is linear only when $n=1$ (in which case it is the identity mapping). If $n=2$ then $f^{-1}(64) = \{8,-8\}$; if $n=3$ $f^{-1}(64)$ is ambiguous: it could mean the subset $\{4\}$, or the number 4 obtained by applying the inverse mapping to 64.

Here is an easy but important

LAZY LEMMA

Let $f: V \to W$ be a linear mapping. Then f is injective $\iff f(v) = 0 \Rightarrow v = 0$ • $\longleftrightarrow f^{-1}(0) = \{0\}$

PROOF

The equation $f(0) = 0$ tells us that • is a special case of the injectivity condition. So if f is injective and $f(v) = 0$, then $f(v) = f(0)$ and thus $v = 0$. Conversely, suppose that • holds. If $f(v_1) = f(v_2)$ then because f is linear,
$f(v_2 - v_1) = f(v_2) - f(v_1) = 0$ and by hypothesis, $v_2 - v_1 = 0$. Thus, f is injective. QED

Bases and linear mappings

We now use linear mappings to interpret the conditions that define a basis. Suppose that $v_1,\dots,v_n$ are any n elements of a vector space V. Define a mapping $f: \mathbb{R}^{1,m} \to V$ by
$(a_1,\dots,a_m) \mapsto a_1v_1 + \dots + a_nv_n$ *
It is easy to check that this mapping is linear. Then (B1)* asserts that it is surjective, and (B2)* implies that it is injective (with the help of the lazy lemma).
A bijective linear mapping is also called an isomorphism, so a bases of V defines a isomorphism f from $\mathbb{R}^m$ to V. Observe from * that f maps each element of the canonical basis of $\mathbb{R}^m$ ~~V~~ onto an element of the chosen basis of V. If $\{v_1,\dots,v_n\}$ is a basis, we may use f to identify $\mathbb{R}^m$ with V, and to transfer properties of $\mathbb{R}^m$ to V. This enables one to prove results such as the

* $B_1 = \{v_1 \dots v_n\}$ generates $V \Rightarrow f$ surj.
$B_2 = \{v_1 \dots v_n\}$ is LI $\Rightarrow f^{-1}(0_v) = (0,\dots,0) \Rightarrow f$ inj.

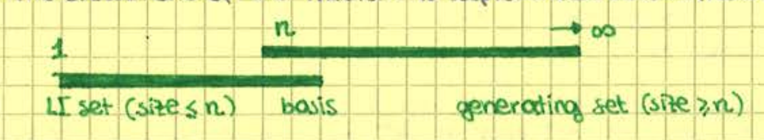
Theorem

Let V be a vector space with a basis of size m. We have
(i) if m vectors $v_1,\dots,v_m$ of V are LI then $m \le n$
(ii) if $V = \mathcal{L}\{v_1,\dots,v_p\}$ then $n \le p$
In particular, any basis of V has n elements, and V is said to have dimension n.

PROOF

We already know that this is true for $V = \mathbb{R}^m$. For, we represent vectors as rows of a matrix A with size $m \times n$ or $p \times n$, and use the theory of the rank $r(A)$ of A. Part (i) implies that $r(A) = m$, and so $m \le n$. Part (ii) implies that $r(A) = n$ and so $n \le p$.
To deduce (i) in general, suppose that $u_1,\dots,u_m$ are LI in V. Then $f^{-1}(u_1),\dots,f^{-1}(u_m)$ are LI in $\mathbb{R}^m$ cause $0 = a_1f^{-1}(u_1) + \dots + a_mf^{-1}(u_m) = f^{-1}(a_1u_1 + \dots + a_mu_m) \Rightarrow 0 = a_1u_1 + \dots + a_mu_m \Rightarrow a_1 = \dots = a_m = 0$
Hence $m \le n$. Part (ii) is similar. QED

The statement of the theorem is represented schematically by



* **LAZY COROLLARY**

Suppose that we already know that $\dim V = n$. Then in checking whether a set of n elements is a basis we only need bother to check one of (B1), (B2). (check (B1) ~ generates ⇒ (B2) holds too ~ L.I)
• $\mathcal{L}\{v_1,\dots,v_n\} = V$, if not L.I then $v_n \in \mathcal{L}\{v_1,\dots,v_{n-1}\} = V$

Example
We are perfectly at liberty to apply the Def. to the same vector space with different bases. Let $V = W = \mathbb{R}_2[x]$. Choose the basis $\{1, x+1, (x+1)^2\}$ for V and the basis $\{1, x, x^2\}$ for W. Let D be the linear mapping defined by differentiation: $Dp = p'$. Then the matrix of D wrt the chosen basis is

$$A = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}$$

the null column tells us that $D(1) = 0$ and the null row tells us that ImD consists of polynomials of degree no greater than 1.

Compositions and products

With the link between linear mappings and matrices now established, we shall see that composition of matrices corresponds to the product of matrices. Suppose that $B \in \mathbb{R}^{m,n}$, $A \in \mathbb{R}^{n,p}$ and consider the associated linear mappings
$\mathbb{R}^{m,1} \xleftarrow{g} \mathbb{R}^{n,1} \xleftarrow{f} \mathbb{R}^{p,1}$
defined by $f(u) = Au$ and $g(v) = Bv$ (It is easier to understand what follows by writing the mappings from right to left) The composition $g \circ f$ is obviously $BAu \leftarrow Au \leftarrow u$, and is therefore associated to the matrix BA.
More generally, given vector spaces U, V, W and linear mappings $W \xleftarrow{g} V \xleftarrow{f} U$, *
choose bases for each of U, V, W and let M_f, M_g be the associated matrices (the same basis of V being used for both matrices) Then we state without proof the

Proposition
Let $h = g \circ f$ be the composition *. Then $M_h = M_g M_f$.

This result is especially useful in the case of a single vector space V of dimension n, and a linear mapping $f: V \to V$. Such a linear mapping (between the same vector space) is called a linear transformation or endomorphism.
In these circumstances, we can fix the same basis $\{v_1, \dots, v_n\}$ of V, and consider compositions of f with itself:

Example - Define $f: \mathbb{R}^3 \to \mathbb{R}^3$ by $f(e_1) = e_2$, $f(e_2) = e_3$ and $f(e_3) = e_1$. Check that the matrix $A = M_f$ (taken wrt the canonic basis $\{e_1, e_2, e_3\}$) satisfies $A^3 = I_3$

Nullity and rank

Important examples of subspaces are provided by the
LEMMA
Let $g: V \to W$ be a linear mapping. Then
(i) $g^{-1}(0)$ is a subspace of V,
(ii) Im g is a subspace of W.

PROOF We shall only prove (ii). If $w_1, w_2 \in \text{Im}\, g$ then we may write $w_1 = g(v_1)$ and $w_2 = g(v_2)$ for some $v_1, v_2 \in V$. If $a \in F$ then
$aw_1 + w_2 = ag(v_1) + g(v_2) = g(av_1 + v_2)$, and so $aw_1 + w_2 \in \text{Im}\, g$. Part (i) is similar. QED

Example In the case of a linear mapping $f: \mathbb{R}^n \to \mathbb{R}^m$ defined by $f(v) = Av$ with $A \in \mathbb{R}^{m,n}$,
$f^{-1}(0) = \{v \in \mathbb{R}^{n,1} : Av = 0\}$ is the solution space of the hom. linear system $Ax = 0$. We already know that this is a subspace and we labelled it ker A. On the other hand, the image of f is generated by the vectors $f(e_i)$ that are columns of A:
$\text{Im} f = \mathcal{L}\{f(e_1) \dots f(e_n)\} = \mathcal{L}\{c_1, \dots, c_n\} = \text{Col}\, A$
It follows that the dimension of Im f equals the rank, $r(A)$, the common dimension of Row A and Col A.

In view of this key example, we state the
Definition (i) The Kernel of an arbitrary linear mapping $g: V \to W$ is the subspace $g^{-1}(0)$, more usually written Ker g or ker g. Its dimension is called the nullity of g.
(ii) The dimension of Im g is called the rank of g

Rank-Nullity Theorem
Given an arbitrary linear mapping $g: V \to W$, $[\dim V = \dim(\ker g) + \dim(\text{Im}\, g)]$
This important result can also be expressed in the form
nullity (g) + rank (g) = n , n = dimension of the space of "departure".

PROOF
By choosing bases for V and W, we may effectively replace g by a linear mapping $f: \mathbb{R}^n \to \mathbb{R}^m$. But in this case, the prev. example shows that $\ker f = \ker A$ and $\text{Im} f = \text{Col} A$. We know that $\dim(\text{Col} A) = \dim(\text{Row} A) = r(A)$, and (essentially by (RC2)),
The result follows $[\dim(\ker A) = n - r(A)]$ QED

The following result is for class discussion:
* **COROLLARY**
Given a linear mapping $g: V \to W$ with $\dim V = n$ and $\dim W = m$
g is injective <=> rank(g) = n (dim ker = 0)
g is surjective <=> rank(g) = m (dim Im = m oppure ker ≠ 0 e dim Im <)
g is bijective <=> rank(g) = m = n ⟶ ISOMORPHISM

KERNEL - note
Ker-f is the space of 2x2 skew-symmetric matrices.

SEE FURTHER EXERCISES!

Dimension counting

Any subspace U is a vector space in its own right and has a dimension: recall that this equals the number of elements inside any bases of U.

Obvious LEMMA
If U is a subspace of a vector space (or another subspace) V then $\dim U \leq \dim V$, with equality $\iff U = V$.

This is true because a basis of U can always be extended until it becomes one of V. To do this we can use the trick that if $v_1, \dots, v_k$ are LI and v_{k+1} is not a LC of $v_1, \dots, v_k$, then $v_1, \dots, v_k, v_{k+1}$ are LI. In the examples above, a subspace of $\mathbb{R}^2$ has dimension 2 only if it is $\mathbb{R}^2$. Similarly for dimension 3 in $\mathbb{R}^3$.
Much of the theory of bases and dimension was discovered by Hermann Grassmann, including the following result dating from around 1860:

Theorem — GRASSMANN'S FORMULA
Let U, V be two subspaces of a finite-dimensional vector space W. Then
$$[\dim U + \dim V = \dim(U \cap V) + \dim(U+V)] \ *$$

This result is illustrated by the following example (whose method is often used as a proof).

Example
Let $W = \mathbb{R}^5$. Consider the two subspaces
$U = \mathcal{L}\{(x_1, x_2, x_3, x_4, x_5) : 2x_1 - x_2 - x_3 = 0 = x_4 - 3x_5\}$,
$V = \mathcal{L}\{(x_1, x_2, x_3, x_4, x_5) : x_3 + x_4 = 0\}$
We are required to find a basis of $\mathbb{R}^5$ that contains both a basis of U and a basis of V. The trick is to start by finding a basis of $U \cap V$. It is easy to see that $\dim U = 3$ and $\dim V = 4$; this is because the homogeneous linear system have rank 2 and 1. Now, a vector $v \in \mathbb{R}^5$ belongs to $U \cap V \iff$ it satisfies all three equations.
Since the associated matrix
$$A = \begin{pmatrix} 2 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -3 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix} \sim \begin{pmatrix} \boxed{1} & -1/2 & 0 & 0 & 3/2 \\ 0 & 0 & \boxed{1} & 0 & 3 \\ 0 & 0 & 0 & \boxed{1} & -3 \end{pmatrix}$$
has rank 3, we deduce that $\dim(U \cap V) = 5 - 3 = 2$. Indeed, we may take $x_2 = s$ and $x_5 = t$ to be free variables and obtain (as a row) the general solution $v = (\tfrac{1}{2}s - \tfrac{3}{2}t, s, -3t, 3t, t)$
A basis of $U \cap V$ consists of
$w_1 = (1/2, 1, 0, 0, 0)$, $w_2 = (-3/2, 0, -3, 3, 1)$
(take first $s = 1, t = 0$ and second $s = 0, t = 1$) Extend this basis in any way to
a basis $\{w_1, w_2, w_3\}$ of U and
a basis $\{w_1, w_2, w_4, w_5\}$ of V
Then $\{w_1, w_2, w_3, w_4, w_5\}$ will always be LI and thus a basis of $\mathbb{R}^5$. There are lots of choices in this example, but we could take
$w_3 = (0, -1, 1, 0, 0)$ (this works since $w_3 \in U$ but $w_3 \notin \mathcal{L}\{w_1, w_2\}$)
$w_4 = (0, 0, 1, -1, 0)$, $w_5 = (0, 0, 0, 0, 1)$ (note that $w_5 \notin \mathcal{L}\{w_1, w_2, w_4\}$)
In conclusion $U + V = \mathbb{R}^5$ and the required basis is
$$\overbrace{w_5 \quad w_1 \quad w_2}^{U} \quad w_3 \quad w_4$$
$$\underbrace{w_1 \quad w_2 \quad w_3 \quad w_4}_{V}$$

BASIS
$V_1 = \mathcal{L}\{u_1 \dots u_r\}$ $\quad V_2 = \mathcal{L}\{w_1 \dots w_s\}$ $\implies V_1 + V_2 = \mathcal{L}\{u_1 \dots u_r, w_1 \dots w_s\}$

FANCY PROOF OF $*$
First consider the cartesian product $P = U \times V$ consisting of ordered pairs (u, v) with $u \in U$ and $v \in V$. This can be made into a vector space using the operations
$(u_1, v_1) + (u_2, v_2) = (u_1 + u_2, v_1 + v_2)$
$a(u, v) = (au, av)$
If $u_1, \dots, u_m$ is a basis of U and $v_1, \dots, v_n$ a basis of V, it's easy to verify that
$(u_1, 0) \dots (u_m, 0), (v_1, 0) \dots (v_n, 0)$ is a basis of P.
Hence, $\dim P = m + n$ (P is called the external direct sum of U and V).
Consider the mapping
$f : P \to W$, $f(u, v) = u + v$
One easily checks that (i) f is linear, (ii) the image of f equals $U + V$, and (iii) the kernel of f equals $\{(w, -w) : w \in U \cap V\}$. Since the last subspace has the same dimension as $U \cap V$, the Thm follows from the Rank-Nullity formula: $[\dim P = \dim \ker f + \dim \operatorname{Im} f]$ QED.

Example
Consider 2 subspaces $U = \mathcal{L}\{u_1, u_2\}$, $V = \mathcal{L}\{v_1, v_2\}$ of $\mathbb{R}^n$. There are two competing ways to decide mechanically whether $U = V$:
(i) super-reduce the $2 \times n$ matrix with rows u_1, u_2. Do the same for v_1, v_2. Then $U = V \iff$ the 2 super-red. matrices are identical (This method works because the super-reduced version of a matrix is unique).
(ii) Step-reduce the $4 \times n$ matrix with rows u_1, u_2, v_1, v_2 to find its rank ρ. Then $U = V \iff \rho = 2$. (This works because in general, $\rho = \dim(U+V)$, whereas $U = V \iff U + V = U = V$)

SEE FURTHER EXERCISES!

$*$ The sum $V_1 + V_2$ is the smallest subspace containing V_1 and V_2.

$V_1 + V_2 = \{v_1 + v_2 : v_1 \in V_1, v_2 \in V_2\}$ $\rightsquigarrow$ verify properties (S1) and (S2)
I) (S1) ✓
(S2) $a(v_1 + v_2) \in U \to av_1 + av_2 = v_1' + v_2'$ OK ✓
II) $V_1, V_2 \subseteq U \to$ SMALLEST ✓

has a right angle in Q by construction. Thus, PR clearly has bigger magnitude than PQ

iii) Determine $d(P,\alpha)$ the distance between a point and a plane. If we decide not to use the formula, we can proceed as follows. Take the line r such that $P \in r$, $r \perp \alpha$
Then consider the point $Q = \alpha \cap r$ and notice that
$d(P,\alpha) = d(P,Q)$

Exercise. Verify that this answer agrees with the definition.

iv) Determine $d(r,s)$ the distance between two lines. This is the more complicated case, and it depends on the respective position of the lines. If the line intersect, then the distance is zero. If the lines are parallel, then it is easy to see that
$d(r,s) = d(P,s) = d(r,Q)$ for any choice of $P \in r$ and $Q \in s$. If the lines are intersecting, or they are skew lines, the following formula will compute the distance

$$\left| \vec{PQ} \cdot \frac{\vec{v}_r \times \vec{v}_s}{\|\vec{v}_r \times \vec{v}_s\|} \right|$$

for any choice of $P \in r$ and $Q \in s$. Notice that, if the lines are intersecting, then they are coplanar, hence the dot product $\vec{PQ} \cdot \vec{v}_r \times \vec{v}_s$ is zero

Exercise. Verify that this answer agrees with the definition using the following geometric construction in the case of skew lines. There exists a (unique) plane α s.t. $\alpha \subset r$ and $\alpha \parallel s$. For this plane, the following holds $\rightarrow$ $d(r,s) = d(\alpha,s)$.

v) Determine $d(\alpha,\beta)$ the distance between two planes. If the plane are not parallel, then they are intersecting and the distance is $d(\alpha,\beta) = 0$; similarly if they coincide. If they are parallel, but distinct, then it is easy to see that
$d(\alpha,\beta) = d(P,\beta) = d(\alpha,Q)$ for any choice of $P \in \alpha$ and $Q \in \beta$

Exercise. Verify that this answer agrees with the definition.

NOTES 17 - Eigenvectors and Eigenvalues

Eigenvectors of a linear transformation

Consider a linear mapping $f: V \rightarrow V$ where V is a vector space with field of scalars F.

Definition
A nonzero element $v \in V$ is called an eigenvector of f if there exists λ (possibly 0) in F s.t. $[f(v) = \lambda v]$
In these circumstances, λ is called the eigenvalue associated to v.

Warning Since $f(0) = 0$, it is obvious that the null vector satisfies the prev. condition. But the null vector does NOT count as an eigenvector; for one thing its eigenvalue λ is undetermined. On the other hand, observe that if v is an eigenvector of f and $a \neq 0$ then av is also an eigenvector (with the same eigenvalue).

Example Here are two extreme cases:
(i) Suppose f is the identity mapping, so that $f(v) = v$ for all $v \in V$. This is obviously linear, and every nonnull vector $v \in V$ is an eigenvector with eigenvalue 1.
(ii) Define $g(v) = 0$ for every $v \in V$ (the 'null' transformation, linear by default). Once again, every nonnull $v \in V$ is an eigenvector, but this time with common eigenvalue 0.

More interesting examples of eigenvectors can easily be written down if there is a basis of the vector space V at one's disposal:

Example (i) Take $V = \mathbb{R}^2$ and define a linear mapping $f: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ by $f(e_1) = e_1$, $f(e_2) = -e_2$
By this very definition, e_1 and e_2 are eigenvectors with eigenvalues 1 and -1. Geometrically, f is a reflection in the x-axis, and any reflection in the plane will have two such eigenvectors.
(ii) Suppose that V has dimension 4 and a basis $\{v_1,v_2,v_3,v_4\}$. We are at liberty to define
$f(v_1) = v_2$, $f(v_2) = v_3$, $f(v_3) = v_4$, $f(v_4) = v_1$, and this uniquely determines the linear mapping $f: V \rightarrow V$ for which $v_1+v_2+v_3+v_4$, $v_1-v_2+v_3-v_4$ are eigenvectors with respective eigenvalues 1 and -1.

To show that bases are not essential for the existence of eigenvectors, here is an example in which the vector space is not finite dimensional:
Exercise. Let V be the vector space of functions $\phi: \mathbb{R} \rightarrow \mathbb{R}$ that admit derivatives of all orders.
Then the mapping $D: V \rightarrow V$ given by
$D(\phi) = \phi'$, where $\phi'(x) = \frac{d\phi}{dx}$, is linear. Find all the eigenvectors (or rather, "eigenfunctions") of D.

Eigenvectors of a square matrix

Suppose that V is a vector space of finite dimension n with $F = \mathbb{R}$ (we shall only consider this case from now on)
Once we have chosen a basis of V (any one will have n elements), we know that we can treat V as if it were $\mathbb{R}^n$. For this reason, we need only study linear mappings $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$, any one of which is given by $f(v) = Av$, where $A \in \mathbb{R}^{n,n}$ is a square matrix.
Accordingly, an eigenvector of the matrix A is a nonzero column vector v such that $Av = \lambda v$*

If $A \in \mathbb{R}^{n,n}$, then can have at most n roots, and there are at most n distinct eigenvalues. There may be less, as the roots of a polynomial can be repeated, and it is also possible that pairs of roots occur as complex numbers.

Given A, suppose that λ is a root of the characteristic polynomial, so that $p(\lambda)=0$. From above, we know that there must exist a nonnull column vector v satisfying (3). We can find therefore such a v by solving the homogeneous linear system associated to $A-\lambda I$. This we shall do in the next lecture.

Example We have not spoken much about 4x4 determinants, but these can be expanded along a row or column into a linear combination of 3x3 determinants. Let

$$A=\begin{pmatrix} x & 0 & 0 & s \\ -1 & x & 0 & r \\ 0 & -1 & x & q \\ 0 & 0 & -1 & x+p \end{pmatrix}$$

Expanding down the first column,

$$\det A = x \det\begin{pmatrix} x & 0 & r \\ -1 & x & q \\ 0 & -1 & x+p \end{pmatrix} + 1\cdot\det\begin{pmatrix} 0 & 0 & s \\ -1 & x & q \\ 0 & -1 & x+p \end{pmatrix} + 0 + 0 =$$

$$= x(x(x(x+p)+q)+1\cdot r)+s = x^4+px^3+qx^2+rx+s$$

This type of example can be used to show that any polynomial whose leading term is $(-x)^n$ is the characteristic polynomial of some $n\times n$ matrix.

SEE FURTHER EXERCISES!

NOTES 18 - Eigenspaces and Multiplicities

In this lecture, we shall explain how to compute methodically all the eigenvectors associated to a given square matrix.

Subspaces generated by eigenvectors

Given $A \in \mathbb{R}^{n,n}$, consider the characteristic polynomial $p(x)=\det(A-xI)$

We know that this polynomial has degree n, and its roots are precisely the eigenvalues of A: a real number λ satisfies $p(\lambda)=0 \iff$ there exists a nonnull column vector $v \in \mathbb{R}^{n,1}$ s.t. $Av=\lambda v$.

Definition

If λ is an eigenvalue, the subspace $E_\lambda = \ker(A-\lambda I) = \{v \in \mathbb{R}^{n,1} : Av=\lambda v\} = \{v \in \mathbb{R}^{n,1} : (A-\lambda I)v = 0\} = \mathrm{Null}(A-\lambda I)$ of $\mathbb{R}^n$ is called the eigenspace associated to λ.

Warning: Not quite all the elements of E_λ are eigenvectors, since (being a subspace) E_λ also includes the null vector 0 that is not counted as an eigenvector.

The dimension of E_λ satisfies

$[n - \dim E_\lambda = r(A-\lambda I)]$, by the Rank-Nullity thm or (RC2).

$E_\lambda \in \mathbb{R}^{n,1}$ – SUBSPACE
$E_\lambda = \{\text{eigenvectors of } \lambda\} \cup \{0_{\mathbb{R}^n}\}$

Example The matrix $A=\begin{pmatrix} -6 & 9 \\ -4 & 7 \end{pmatrix}$ has characteristic polynomial

$p(x) = (-6-x)(7-x)+36 = x^2-x-6 = (x+2)(x-3)$

The roots are $\lambda_1=-2$ and $\lambda_2=3$, and we consider one at time. Firstly,

$A-\lambda_1 I = A+2I = \begin{pmatrix} -4 & 9 \\ -4 & 9 \end{pmatrix} \sim \begin{pmatrix} \boxed{1} & -9/4 \\ 0 & 0 \end{pmatrix}$

As predicted (by the very fact that -2 is an eigenvalue), this matrix has rank less than 2. It is easy to find a nonnull vector in $\ker(A+2I)$, namely

$\begin{pmatrix} 9/4 \\ 1 \end{pmatrix}$ or $\begin{pmatrix} 9 \\ 4 \end{pmatrix}$, whence $E_{-2} = \mathcal{L}\left\{\begin{pmatrix} 9 \\ 4 \end{pmatrix}\right\}$

Similarly

$A-\lambda_2 I = A-3I = \begin{pmatrix} -9 & 9 \\ -4 & 4 \end{pmatrix} \sim \begin{pmatrix} \boxed{1} & -1 \\ 0 & 0 \end{pmatrix}$, and $E_3 = \mathcal{L}\left\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right\}$

Although the previous example only has $n=2$, it illustrates an important technique: that of selecting an eigenvector for each eigenvalue.

The next result is fairly obvious for $k=2$, and we prove it in another special case.

Proposition

Suppose that $v_1,\dots,v_k$ are eigenvectors of A associated to distinct eigenvalues $\lambda_1,\dots,\lambda_k$. Then $v_1,\dots,v_k$ are LI.

PROOF

To give the general idea, we take $k=3$ (which, for the conclusion to be valid, means that the vectors lie in $\mathbb{R}^n$ with $n\geq 3$ and $\lambda_1=1, \lambda_2=2, \lambda_3=3$. Suppose that

$0 = a_1v_1+a_2v_2+a_3v_3$ •

Applying A (or rather its associated linear mapping) gives

$0 = a_1Av_1+a_2Av_2+a_3Av_3 = a_1v_1+2a_2v_2+3a_3v_3$

and subtracting •,

$0 = a_2v_2+2a_3v_3$ •

Applying A again,

$0 = a_2Av_2+2a_3Av_3 = 2a_2v_2+6a_3v_3$.

$k=2$ $\lambda_1 \neq \lambda_2$ $\{v_1,v_2\} \iff x=y=0$
$xv_1+yv_2=0$, $x,y\in\mathbb{R}$
CONTRADICTION
$x\neq 0 \to v_1 = -y/x\, v_2$
$\lambda_1 v_1 = Av_1 = -y/x\, Av_2$
$[\lambda_1 v_1 = -\lambda_2\, y/x\, v_2]$
$-\frac{\lambda_2}{\lambda_1} w = -w$

(ii) If $\Delta = 0$ there is one eigenvalue with multiplicity $\text{mult}(\lambda) = 2$. In the subcase that $\dim E_\lambda = 2$, $E_\lambda = \mathbb{R}^2$ contains both e_1, e_2 and

$A = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} = \lambda I$ * which means that $b = c = 0$ and $a = d$.

(iii) If $\Delta < 0$ there are no real eigenvalues, though the theory still makes sense when one passes to the field $F = \mathbb{C}$ of complex numbers.

An important special case (of (i) or (ii)) is that in which $b = c$, and A is symmetric: this implies that $\Delta \geq 0$ with equality iff A is given by *. Actually, one can prove that the eigenvalues of any real symmetric $n \times n$ matrix are always themselves real.

Example An instance of (ii) with $\dim E_\lambda = 1$ is the matrix $N = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ that satisfies $N^2 = 0$.
The characteristic polynomial is x^2, so 0 is a repeated eigenvalue. A direct way of seeing that any eigenvalue must be 0 is to observe that

$Nv = \lambda v \implies 0 = N^2 v = N(\lambda v) = \lambda^2 v \implies \lambda^2 = 0$

Any eigenvector $\begin{pmatrix} x \\ y \end{pmatrix}$ has to satisfy $y = 0$, so $E_0 = \mathcal{L}\left\{\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right\}$ has dimension 1. Similar considerations apply to the matrix $N + aI$ where $a \neq 0$.

SEE FURTHER EXERCISES!

NOTES 19 - Diagonalizability

For an important class of square matrices (or linear transformations of a finite-dimensional vector space) it is possible to choose a basis of eigenvectors. We have already seen that this is possible if the character. polynomial of $A \in \mathbb{R}^{n,n}$ has distinct real roots. We shall now explain the significance of, and give a more general criterion for, the existence of a basis of eigenvectors. We include a well known application to the theory of Fibonacci numbers.

An example in detail

Consider the following matrix:

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

One's first observation upon setting eyes on this matrix is that $A - I$ has a row of 0's; it follows that 1 is an eigenvalue of A. Given that

$\det A = 5 \cdot 1 - (-3) \cdot 1 = 8 \qquad \text{tr} A = 5 + 1 + 1 = 7$

we may choose the roots $\lambda_1, \lambda_2, \lambda_3$ of $p(x) = \det(A - xI)$ so that $\lambda_1 = 1$, $1 \cdot \lambda_2 \lambda_3 = 8$, $1 + \lambda_2 + \lambda_3 = 7$, and $\lambda_2 = 2$ and $\lambda_3 = 4$. If required to do so, one can verify directly that

$p(x) = -(x-1)(x-2)(x-4)$

Eigenvectors can be found by picking particular solutions of the corresponding linear system:

$$A - I = \begin{pmatrix} 4 & 3 & -3 \\ 0 & 0 & 0 \\ 1 & 2 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 2 & 0 \\ 0 & 5 & 3 \\ 0 & 0 & 0 \end{pmatrix} \implies v_1 = \begin{pmatrix} -6 \\ 3 \\ -5 \end{pmatrix}$$

$$A - 2I = \begin{pmatrix} 3 & 3 & -3 \\ 0 & -1 & 0 \\ 1 & 2 & -1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \implies v_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

$$A - 4I = \begin{pmatrix} 1 & 3 & -3 \\ 0 & -3 & 0 \\ 1 & 2 & -3 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & -3 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \implies v_3 = \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}$$

Let $f: \mathbb{R}^3 \to \mathbb{R}^3$ denote the linear transformation defined by $f(v) = Av$. Instead of using the canonical basis e_1, e_2, e_3 of $\mathbb{R}^3$, we are at liberty to use the basis v_1, v_2, v_3 (these 3 vectors are obviously LI, though this is also a theoretical consequence of the fact that the corresponding eigenvalues are distinct). Whilst the matrix of f wrt the canonical basis is A, its matrix wrt the new basis is

$D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix}$ *, reflecting the equations $Av_1 = 1v_1$, $Av_2 = 2v_2$, $Av_3 = 4v_3$ •

We shall now make the relationship between the two matrices A and D more explicit.

Let P denote the matrix whose columns are the chosen eigenvectors:

$P = \begin{pmatrix} \uparrow & \uparrow & \uparrow \\ v_1 & v_2 & v_3 \\ \downarrow & \downarrow & \downarrow \end{pmatrix} = \left(\begin{array}{c|c|c} -6 & 1 & 3 \\ 3 & 0 & 0 \\ -5 & 1 & 1 \end{array}\right)$ ⇝ vertical lines to emphasize the column structure of this matrix.

In view of •

$AP = \begin{pmatrix} \uparrow & \uparrow & \uparrow \\ Av_1 & Av_2 & Av_3 \\ \downarrow & \downarrow & \downarrow \end{pmatrix} = \left(\begin{array}{c|c|c} -6 & 2 & 12 \\ 3 & 0 & 0 \\ -5 & 2 & 4 \end{array}\right)$

We get exactly the same result by multiplying P by the diagonal matrix D on the right:

$PD = \left(\begin{array}{c|c|c} -6 & 1 & 3 \\ 3 & 0 & 0 \\ -5 & 1 & 1 \end{array}\right) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix} = \left(\begin{array}{c|c|c} -6 & 2 & 12 \\ 3 & 0 & 0 \\ 5 & 2 & 4 \end{array}\right)$ In conclusion $AP = PD$

Since the columns of P are LI, P has rank 3 and is invertible. One can therefore assert (without the need to actually compute P^{-1}) that

$P^{-1}AP = D$ or $A = PDP^{-1}$

SEE FURTHER EXERCISES!

NOTES 20 - Symmetric and Orthogonal matrices

In this lecture, we focus attention on symmetric matrices, whose eigenvectors can be used to construct orthogonal matrices. Determinants will then help us to distinguish those orthogonal matrices that define rotations

Orthogonal eigenvectors

Recall the definition of the dot or scalar product of two column vectors $v, w \in \mathbb{R}^{n,1}$. Without writing out their components, we can nonetheless assert that

$[v \cdot w = v^T w]$*

Recall too that a matrix S is symmetric if $S^T = S$ (this implies of course that it is square)

LEMMA

Let v_1, v_2 be eigenvectors of a symmetric matrix S corresponding to distinct eigenvalues λ_1, λ_2. Then $v_1 \cdot v_2 = 0$

PROOF

First note that

$(Sv_1)\cdot v_2 = (Sv_1)^T v_2 = v_1^T S^T v_2 = v_1^T(Sv_2) = v_1 \cdot (Sv_2)$

This is true for any vectors v_1, v_2, but the assumptions $Sv_1 = \lambda_1 v_1$ and $Sv_2 = \lambda_2 v_2$ tells us that $\lambda_1 v_1 \cdot v_2 = \lambda_2 v_1 \cdot v_2$ and the result follows. QED

Example To begin with the 2x2 case, consider the symmetric matrix

$A = \begin{pmatrix} 5 & 2 \\ 2 & 8 \end{pmatrix}$. It's easy to check that its eigenvalues are 9 and 4, and that respective eigenvectors are $v_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, $v_2 = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$. As predicted by the LEMMA, $v_1 \cdot v_2 = 0$. Given this fact, we can normalize v_1, v_2 to manifacture an orthonormal basis $f_1 = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 2 \end{pmatrix}$, $f_2 = \frac{1}{\sqrt{5}}\begin{pmatrix} -2 \\ 1 \end{pmatrix}$ of eigenvectors, and use these to define the matrix $P = \frac{1}{\sqrt{5}}\left(\begin{array}{c|c} 1 & -2 \\ 2 & 1 \end{array}\right)$. With this choice $P^T P = \frac{1}{5}\begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}\left(\begin{array}{c|c} 1 & -2 \\ 2 & 1 \end{array}\right) = \frac{1}{5}\begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix} = I_2$*

Another way of expressing these relationship is $P^{-1} = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix} = P^T$, and also $PP^T = I$.

It is easy to verify that $P^{-1}AP = \frac{1}{5}\begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} 5 & 2 \\ 2 & 8 \end{pmatrix}\begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 9 & 0 \\ 0 & 4 \end{pmatrix}$

Definition

A matrix $P \in \mathbb{R}^{n,n}$ is called orthogonal if it satisfies one of the equivalent conditions:

(i) $P^T P = I_n$

(ii) $PP^T = I_n$

(iii) P is invertible and $P^{-1} = P^T$

Let us explain why the three conditions are indeed equivalent. As * and * make clear, condition (i) asserts that the columns of P are orthonormal. Condition (ii) asserts that the rows are orthonormal. A set $\{v_1, \dots, v_n\}$ of orthonormal vectors is necessarily LI since $a_1v_1 + \dots + a_nv_n = 0$ implies (by taking the dot product with each v_i in turn) that $a_i = 0$; thus either (i) or (ii) implies that P is invertible. It follows that both (i) and (ii) are equivalent to (iii)

The relationship between symmetric and orthogonal matrices is cemented by the

Theorem

Let $S \in \mathbb{R}^{n,n}$ be a symmetric matrix. Then

(i) the eigenvalues (or roots of the characteristic polynomial $p(x)$) of S are all real.

(ii) there exists an orthogonal matrix P s.t. $P^{-1}SP = P^T SP = D$.

PROOF

Suppose that $\lambda \in \mathbb{C}$ is a root of $p(x)$. Working over the field $\mathbb{C}$, we can assert that there exists a complex eigenvector $v \in \mathbb{C}^{n,1}$ satisfying $Sv = \lambda v$. If $v^T = (z_1, \dots, z_n)$ then the complex conjugate of this vector is $\bar{v}^T = (\bar{z}_1, \dots, \bar{z}_n)$ and $\bar{v}^T v = |z_1|^2 + \dots + |z_n|^2 > 0$ since $v \neq 0$.

Thus $\lambda \bar{v}^T v = \bar{v}^T(Sv) = \overline{Sv}^T v = \bar{\lambda}\bar{v}^T v$, and necessarily $\lambda = \bar{\lambda}$ and $\lambda \in \mathbb{R}$.

In the light of (i), part (ii) follows immediately if all the roots of $p(x)$ are distinct. For each repeated root λ, one needs to know that $\text{mult}(\lambda) = \dim E_\lambda$; for if this is true the LEMMA permits us to built up an orthonormal basis of eigenvectors. We shall not prove the multiplicity statement (that is always true for a symmetric matrix), but a convincing exercise follows. QED

Exercise. Consider again the symmetric matrix

$A = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix}$ and its eigenvectors $v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, $v_2 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$, $v_3 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$ with respective eigenvalues 0 (mult = 1) and -3 (mult = 2). As predicted by the LEMMA, $v_1 \cdot v_2 = 0 = v_1 \cdot v_3$. Observe however that $v_2 \cdot v_3 \neq 0$; show nonetheless that there exists an eigenvector v_3' with eigenvalue 3 s.t. $v_2 \cdot v_3' = 0$. Normalize the vectors v_1, v_2, v_3' so as to obtain an orthogonal matrix P for which $P^{-1}AP$ is diagonal. Compute the determinant of P; can the latter be chosen so that $\det P = 1$?

NOTES 21 - Quadratic forms and conics

After some general definitions, we shall study quadratic forms in 2 variables x, y. Such a form corresponds to a 2x2 symmetric matrix whose diagonalization enables us to simplify the quadratic form. The resulting equations typically describe ellipses or hyperbolas, although degenerate cases are subject of some of the exerc.

Homogeneous polynomials.

We already know that any linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$ is effectively multiplication by an $m\times n$ matrix. In particular, setting $m=1$, a linear mapping $\ell: \mathbb{R}^n \to \mathbb{R}$ is defined by

$$\ell(v) = (a_1 \dots a_n)\, v = a_1x_1 + \dots + a_nx_n, \text{ where } v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

The sum $a_1x_1 + \dots + a_nx_n$ is called a linear form in the n variables $x_1, \dots, x_n$; each term in this sum has degree exactly 1 (no constants are allowed on their own as these would have degree 0). Such a sum is also called a homogeneous polynomial of degree 1 in $x_1, \dots, x_n$.

* A quadratic form in $x_1, \dots, x_n$ is a linear combination of terms of degree exactly 2, such as x_1^2, x_1x_2 and so on. We can construct a quadratic form from a square matrix by setting

$$q(v) = v^TAv = (x_1, \dots, x_n)\begin{pmatrix} a_{11} & a_{12} & \dots & \\ a_{21} & a_{22} & \dots & \\ \vdots & \vdots & & a_{nn} \end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

Using a calculation (best done in the 3x3 case), or the summation formula for matrix multiplication, we obtain

$$q(v) = a_{11}x_1^2 + a_{12}x_1x_2 + \dots + a_{21}x_2x_1 + \dots + a_{nn}x_n^2 = \sum_{i,j=1}^{n} a_{ij}x_ix_j$$

This function is also called a homogeneous polynomial of degree 2 since the total power of each term equals exactly 2.
Since the coefficient of x_ix_j in the prev. function equals $a_{ij} + a_{ji}$ if $i \neq j$, we may as well suppose that A is a symmetric matrix. With this assumption, we can recover A from the quadratic form:

Example Given the quadratic form $q(x,y,z) = 2x^2 + y^2 + 5z^2 + 6xy + 2yz$, we need to half the "mixed" coefficients to find the off-diagonal entries of the associated symmetric matrix

$$A = \begin{pmatrix} 2 & 3 & 0 \\ 3 & 1 & 1 \\ 0 & 1 & 5 \end{pmatrix} \quad \text{so that} \quad q(x,y,z) = (x,y,z)\,A\begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

A 2x2 example revisited

Recall the matrix

$A = \begin{pmatrix} 5 & 2 \\ 2 & 8 \end{pmatrix}$ It is associated to the ~~matrix~~ quadratic form $q(x,y) = 5x^2 + 4xy + 8y^2$ and we now pose the problem: describe the set of points (x,y) in $\mathbb{R}^2$ s.t. $q(x,y) = 1$. We shall answer this question by diagonalizing A.

The eigenvalues of A are 9 and 4. This time, we take the smallest first: $\lambda_1 = 4$ and $\lambda_2 = 9$.
A first eigenvector lies in $\ker(A-4I) = \ker(1\ 2)$, and a unit one is $\frac{1}{\sqrt5}\begin{pmatrix} 2 \\ -1 \end{pmatrix}$. It follows that A is diagonalized by means of the orthogonal matrix

$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \frac{1}{\sqrt5}\begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}$; there's no need to compute separately the 2nd eigenvector. Thus,

$R_{-\theta}AR_\theta = D = \begin{pmatrix} 4 & 0 \\ 0 & 9 \end{pmatrix}$*, since $(R_\theta)^T = R_{-\theta} = (R_\theta)^{-1}$

We now define new coordinates X, Y by setting

$\begin{pmatrix} X \\ Y \end{pmatrix} = R_{-\theta}\begin{pmatrix} x \\ y \end{pmatrix}$ or $\begin{cases} X = x\cos\theta + y\sin\theta \\ Y = -x\sin\theta + y\cos\theta \end{cases}$

For practical purposes, it is however best to express the old coordinates in terms of the new ones:

$\begin{pmatrix} x \\ y \end{pmatrix} = R_\theta\begin{pmatrix} X \\ Y \end{pmatrix}$ or $\begin{cases} x = X\cos\theta - Y\sin\theta \\ y = X\sin\theta + Y\sin\theta \end{cases}$ *

For this enables us to substitute the new for old, and the coefficients on the right of * are the entries of R_θ placed in the correct position. The linear mapping associated to R_θ is a rotation through an angle of θ, where

$$\cos\theta = 2/\sqrt5, \quad \sin\theta = -1/\sqrt5 \qquad \text{which implies that } \theta \text{ is about } -25.6°.$$

The transpose of the left equation in * is $(x,y) = (X\ Y)R_{-\theta}$. It follows from * that

$$(x\ y)\,A\begin{pmatrix} x \\ y \end{pmatrix} = (X\ Y)R_{-\theta}AR_\theta\begin{pmatrix} X \\ Y \end{pmatrix} = (X\ Y)\begin{pmatrix} 4 & 0 \\ 0 & 9 \end{pmatrix}\begin{pmatrix} X \\ Y \end{pmatrix}$$

Hence $5x^2 + 4xy + 8y^2 = 4X^2 + 9Y^2$ and our equation becomes $4X^2 + 9Y^2 = 1$ or $\frac{X^2}{(1/2)^2} + \frac{Y^2}{(1/3)^2} = 1$ which is the equation of an ellipse.

* QUADRATIC FORMS APPLICATION
- Geometry: conics and quadrics
- Calculus: study of extremal values (MAX, MIN) in functions with more than 1 variable

To eliminate all this, we need to solve the linear system with unknowns u, v and matrix
$$\left(\begin{array}{cc|c} A & B & -D \\ B & C & -E \end{array}\right) \bullet$$

Since the left hand side of this matrix is twice (2), a solution may not be possible if $\det S = 0$.

Definition
The conic $\mathscr{C}$ is central if there is a translation * that converts its equation into the form $A'X^2 + 2B'XY + C'Y^2 + F' = 0$.
In this case, $(X,Y) \in \mathscr{C} \iff (-X,-Y) \in \mathscr{C}$, and the centre of symmetry is the point $(X,Y) = (0,0)$ or $(x,y) = (u,v)$. (also the symmetric of the point considered is a point of the conic)*

From the analysis above, we know that there is only one case in which • is incompatible and $\mathscr{C}$ is not central, namely (viii)

*COROLLARY The conic (1) can only be a parabola if $B^2 = 4AC$.

Example Given the conic $x^2 + 4y^2 - 6x + 8y = 3$, we can locate a centre by completing the squares:
$(x-3)^2 - 9 + 4(y+1)^2 - 4 = 3$
Thus $u = 3$, $v = -1$, and the equation becomes

$X^2 + 4Y^2 = 16$, or $\frac{X^2}{4^2} + \frac{Y^2}{2^2} = 1$, which is an ellipse with width twice its height. In general, if the original equation has a term in xy, one needs to find the centre by solving (4).•

Central quadrics

One can carry out a parallel discussion in space by adding a third variable.

Definition
A quadric $\mathscr{Q}$ is the locus of points (x,y,z) in $\mathbb{R}^3$ satisfying an equation of the form

$Ax^2 + By^2 + Cz^2 + 2Dyz + 2Ezx + 2Fxy + 2Gx + 2Hy + 2Iz + J = 0$ •
The word quadric implies that • has order 2, so not all of A, B, C, D, E, F are zero.

One can list all the different types of quadrics; whilst there were 8 types of conics there are 15 types of quadrics. However, we shall only consider the more interesting cases in this course.
Let us start with an obvious example. The equation
$|x^2 + y^2 + z^2 = r^2|$
fits the definition (with $A = B = C = 1$, $J = -r^2$, and all other coeff. $= 0$). It is of course a sphere of radius r with centre the origin. Indeed if $v = (x,y,z)^T$, then the equation becomes $|v|^2 = r^2$ or $|v| = r$ and asserts that the distance of (x,y,z) from the origin is r.
In the light of the discussion of ellipses, it should now come as no surprise that the equation
$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$ represents an ellipsoid that fits snugly into a box centred at the origin of dimension $2a \times 2b \times 2c$.

Definition
A central quadric is the locus of points (x,y,z) satisfying an equation
$Ax^2 + By^2 + Cz^2 + 2Dyz + 2Ezx + 2Fxy + J = 0$ •
or equivalently

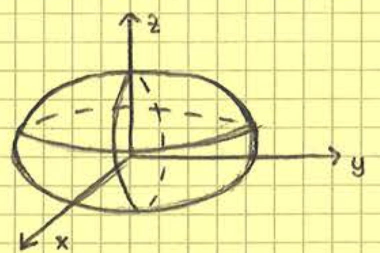
* the same for the quadric

$$(x\ y\ z)\begin{pmatrix} A & F & E \\ F & B & D \\ E & D & C \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = -J$$

The 3x3 matrix here is symmetric, and we can rewrite the equation as $[v^T S v = -J]$, where $\mu = -J$
we know that there exists a 3x3 orthogonal matrix P so that $P^{-1}SP = P^TSP$ is diagonal.
we may also suppose that $\det P = 1$ (for if not, $\det P = -1$ and we merely replace P by $-P$ and note that $\det(-P) = 1$). It follows that P represents a rotation; thus we have the

Theorem
Given a central quadric, it is possible to rotate the coordinate system about the origin in space so that in the new system the equation becomes
$[\lambda_1 X^2 + \lambda_2 Y^2 + \lambda_3 Z^2 = \mu.]$ •
The numbers $\lambda_1, \lambda_2, \lambda_3$ are (in no particular order) the eigenvalues of S.

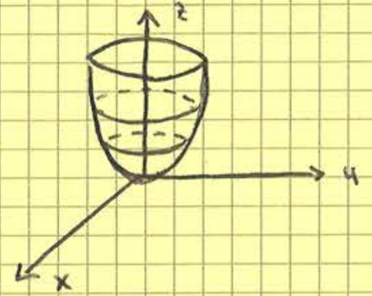
Here are some examples of central quadrics in which the eigenvalues are all nonzero:
(i) an ellipsoid (if $\lambda_1, \lambda_2, \lambda_3, \mu$ all have the same sign);



λ_1	λ_2	λ_3	μ
+	+	+	+
−	−	−	−

$$\boxed{\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1}$$

If a,b have opposite signs, it is again a hyperbolic paraboloid. If a,b have the same sign, the quadric is easier to draw and is called an elliptic ~~hyperboloid~~ paraboloid (circular if a=b). Its intersection with the plane z=a is an ellipse (circle)



$\lambda_1 \ \lambda_2 \ \lambda_3 \ \mu$
$+ \ + \ 0$

$$\boxed{z = x^2 + y^2}$$

SEE FURTHER EXERCISES!

✳ Definition
A matrix A is symmetric if $A = A^T$

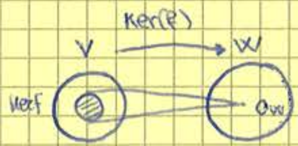
Consider $V \subseteq \mathbb{R}^{3,3} = \{A : A \text{ symm.}\}$ then
1 - Prove that V is a vector space ⇝ proving properties (SE1), (SE2)
2 - Find a basis of V

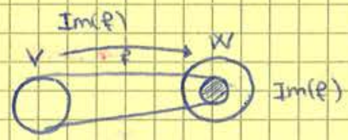
$v_1 = A \in \mathbb{R}^{3,3}$ symm $(A = A^T)$; $v_2 = B \in \mathbb{R}^{3,3}$ symm $(B = B^T)$
$\Rightarrow [A+B = A^T + B^T = (A+B)^T]$

✳ About Kernel and Image

$f : V \to W$ (linear mapping)

- $\ker(f) = \{v \in V : f(v) = 0\}$
- $\mathrm{Im}(f) = \{w \in W : w = f(v)\}$





Kernel → subspace of V
Image → subspace of W

- Ker(f) is a subspace

(S1) $v_1, v_2 \in \ker(f) \Rightarrow v_1 + v_2 \in \ker(f)$
$v_1 + v_2 \in \ker(f) \Leftrightarrow f(v_1+v_2) = 0_W = f(v_1) + f(v_2) = 0_W + 0_W = 0_W$
(S2) $a \in \mathbb{R}$, $v \in \ker(f) \Rightarrow av \in \ker(f)$
$f(av) = a f(v) = 0_W$

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

$\ker(f) = \{(x,y,z) : f(x,y,z) = (0,0)\}$

$$\begin{cases} x+y=0 \\ x+y=0 \end{cases}$$

$$\begin{array}{ccc} x & y & z \end{array}$$
$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{rank} = 1 \rightsquigarrow \infty^2 \text{ sol } \begin{cases} x = -t \\ y = t \\ z = u \end{cases}$$

$\ker(f) = \{(-t, t, u) : t, u \in \mathbb{R}\}$ DIM 2 $\{(0,0,1), (-1,1,0)\}$ basis

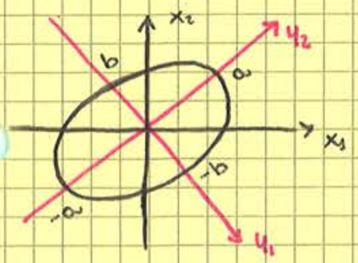
Definition - Image
$\mathrm{Im}(f) = \{f(x,y,z) \in \mathbb{R}^2 : (x,y,z) \in \mathbb{R}^3\}$
$f(x,y,z) = f(xe_1 + ye_2 + ze_3)$, $(x,y,z) = xe_1 + ye_2 + ze_3$
$f(xe_1 + ye_2 + ze_3) = xf(e_1) + yf(e_2) + zf(e_3)$
$\mathrm{Im}(f) = \mathcal{L}\{f(e_1), f(e_2), f(e_3)\}$

$e_1 = (1,0,0)$ → $f(e_1) = (1,1)$
$e_2 = (0,1,0)$ → $f(e_2) = (1,1)$
$e_3 = (0,0,1)$ → $f(e_3) = (0,0)$

→ $\mathrm{Im}(f) = \mathcal{L}\{(1,1), \cancel{(1,1)}, \cancel{(0,0)}\} = \{(1,1)\}$ BASIS of Im(f)
$\dim \mathrm{Im}(f) = 1 = r(A)$
$\ker(f) = \mathcal{L}\{(0,0,1), (-1,1,0)\}$
$\dim \ker(f) = 2 = 3 - r(A)$

$f : \mathbb{R}^3 \to \mathbb{R}^2 \quad \dim \mathrm{Im} + \dim \ker = 3$
$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ f(e_1) & f(e_2) & f(e_3) \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = f(x,y,z)$$

Drawing conics



ELLIPSE - plane curve

$$\boxed{\frac{x^2}{a^2}+\frac{y^2}{b^2}=1}$$

HOW TO CLASSIFY CONICS

$\gamma: Ax^2+Bxy+Cy^2+2Dx+2Ey+F=0$

$$\Delta=\begin{vmatrix} A & B & D \\ B & C & E \\ D & E & F \end{vmatrix} \qquad \delta=\begin{vmatrix} A & B \\ B & C \end{vmatrix}$$

$\Delta=0 \rightarrow \gamma$ degenere

$\delta<0$ iperbole, $\delta=0$ parabola, $\delta>0$ ellisse

MATH I

Curves in space

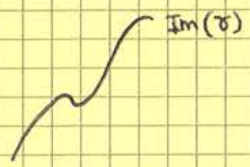
Definition

A curve is a function $\gamma: \mathbb{R}\rightarrow\mathbb{R}^3$, $t \mapsto (x(t),y(t),z(t)) = \gamma(t)$

$x,y,z: \mathbb{R}\rightarrow\mathbb{R}$ CONTINUOUS

In general a curve is $\gamma: I \subseteq \mathbb{R} \rightarrow \mathbb{R}^3$

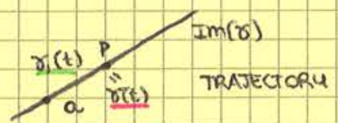
CURVE $\longleftrightarrow$ FUNCTION



Example (Physics)

$\gamma: [0,+\infty) \rightarrow \mathbb{R}^3$

• $\gamma(t) = (t,t,t)$



• $\gamma_1(t) = (2t,2t,2t)$

Differentiable curves - same image and trajectory

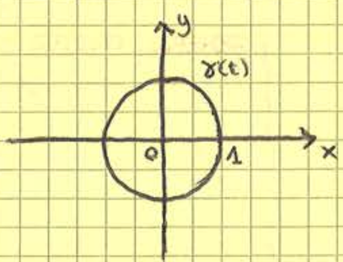
Definition

A regular curve is a differentiable curve $\gamma: I \subseteq \mathbb{R}\rightarrow\mathbb{R}^3$ such that $\gamma'(t) = (x'(t),y'(t),z'(t)) \neq 0$ for $t \subseteq I$ in 0.

Example - lines are regular curves, moreover NOTE $\gamma'(t) = \vec{v}$, DIRECTION VECTOR

Example

$\gamma: [0,2\pi] \rightarrow \mathbb{R}^3$, $\gamma(t) = (\sin t, \cos t, 0)$ $\rightarrow$ we are in the x,y plane ($z=0$)



$x^2(t) + y^2(t) = 1$

$\gamma'(t) = (\cos t, -\sin t, 0) \rightarrow$ NEVER 0, regular curve

Topology in $\mathbb{R}$

Intervals: (a,b) $\longleftrightarrow$; $(a,+\infty)$ $\longleftarrow$; $(-\infty,a]$ $\longrightarrow]$

Definition $P \in I$ is called interior point if $\exists$ J interval s.t. $P \in J \subset I$, $J=(a,b)$ open interval

Integral along a curve

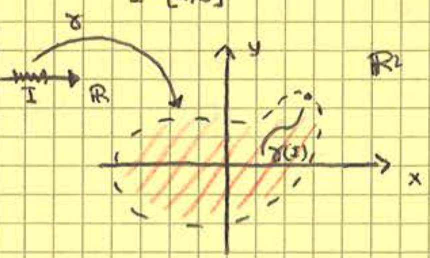
Definition

$\gamma : I \subseteq \mathbb{R} \to \mathbb{R}^d$

$f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$
$(x,y) \mapsto f(x,y)$
$\gamma(I) \subseteq D$

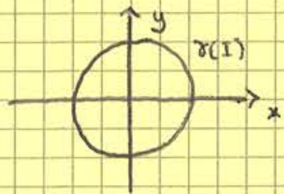
$$\left[\int_\gamma f = \int f(\gamma(t)) \, \|\gamma'(t)\| \, dt \right]$$

$I = [a,b]$



Exercise

$\gamma(t) = (\cos t, \sin t)$
$\gamma : I \to \mathbb{R}$, $I = [0, 2\pi]$



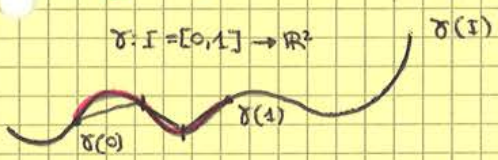
$f : \mathbb{R}^2 \to \mathbb{R}$
$(x,y) \mapsto x+y$

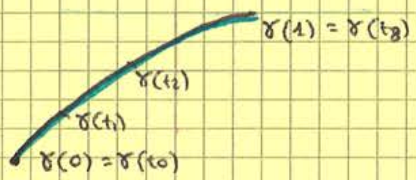
$g : \mathbb{R}^2 \to \mathbb{R}$
$(x,y) \mapsto 1$

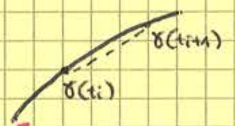
$$\int_\gamma f = \int_a^b f(\gamma(t)) \|\gamma'(t)\| \, dt = \int_0^{2\pi} (\cos t + \sin t) \, dt = 0$$

$$\int_\gamma g = \int_0^{2\pi} g(\gamma(t)) \|\gamma'(t)\| \, dt = \int_0^{2\pi} 1 \cdot 1 \, dt = 2\pi = \text{LENGTH}$$

Length of a curve





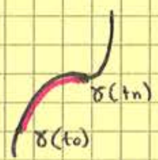


$t_{i+1} - t_i = \Delta t_i$

$$\left[\sum_{i=0}^{n-1} \|\gamma(t_{i+1}) - \gamma(t_i)\| \right] \to \text{length of } \gamma \text{ between } \gamma(t_0) \text{ and } \gamma(t_n)$$

$$\| \quad \sqrt{\underbrace{\frac{(\gamma(t_{i+1}) - \gamma(t_i))}{\Delta t_i}}_{\Delta t_i \to 0 \;\; = \gamma'(t)} \cdot \frac{(\gamma(t_{i+1}) - \gamma(t_i))}{\Delta t_i}}$$

$$\sum_{i=0}^{n-1} \|\gamma(t_{i+1}) - \gamma(t_i)\| \xrightarrow{\Delta t_i} \int_{t_0}^{t_i} \|\gamma'(t)\| \, dt$$

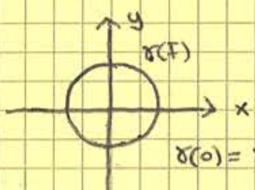


$\int_\gamma 1$ length
$\| $
$\int_0^{t_n} \|\gamma'(t)\| \, dt$

Exercise

$\gamma(t) = (\cos 2t, \sin 2t)$
$\gamma : [0, 2\pi] \to \mathbb{R}^2$

$$4 = \int_\gamma 1 = \int_0^{2\pi} \|\gamma'(t)\| \, dt = \int_0^{2\pi} 2 \, dt$$



$\gamma(0) = \gamma(\pi) = \gamma(2\pi)$

ARC LENGTH

$\boxed{t_1 - t_0}$

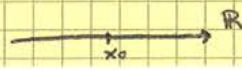
if $\|\gamma'(t_0)\| = 1 \;\; \forall t$
then $\int_\gamma 1 = t_1 - t_0$

t is called **arc length**

NORM

$\|\vec{v}\|$ magnitude or **norm**

Ex $x \in \mathbb{R}$: $\|x\| = |x|$



$I_r(x_0) = \{x \in \mathbb{R} : |x - x_0| < r\}$

Chain Rules

$\mathbb{R} \xrightarrow{f} \mathbb{R} \xrightarrow{g} \mathbb{R}$

$g \circ f(x)$

$\gamma'(t) \longrightarrow \frac{d}{dx} g \circ f(x) = \frac{d}{dx} g(f(x)) \cdot \frac{d}{dx} f(x)$ ⇝ partial derivative of a composition of functions

consider a curve $\mathbb{R} \xrightarrow{\gamma} \mathbb{R}^m$

$m = 2,3$

and a function $\mathbb{R}^m \xrightarrow{f} \mathbb{R}$

the composition will be $\mathbb{R} \xrightarrow{\gamma} \mathbb{R}^m \xrightarrow{f} \mathbb{R}$

* $\frac{d}{dt} f \circ \gamma$?

DOT PRODUCT

$$\left[\frac{d}{dt} f \circ \gamma(t) = \underbrace{\nabla f(\gamma(t))}_{\mathbb{R}^{1,m}} \cdot \underbrace{\gamma'(t)}_{\mathbb{R}^{1,m}}\right]$$ CHAIN RULE

m = 2

$\mathbb{R} \xrightarrow{\gamma} \mathbb{R}^2 \xrightarrow{f} \mathbb{R}$

$t \longmapsto (x(t), y(t)) \longmapsto f(x(t), y(t))$

$\gamma'(t) = (x'(t), y'(t))$

$\nabla f(P) = \left(\frac{\partial}{\partial x} f(P), \frac{\partial}{\partial y} f(P)\right)$ with $P = (x(t), y(t))$

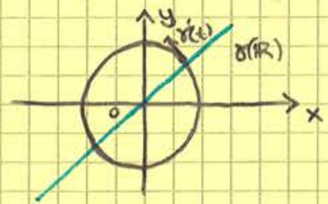
$$\frac{d}{dt} f \circ \gamma(t) = \frac{d}{dt} f(x(t), y(t)) = \boxed{\nabla f(x(t), y(t)) \cdot \gamma'(t)} = \left(\frac{\partial f}{\partial x}(x(t), y(t)), \frac{\partial f}{\partial y}(x(t), y(t))\right) \cdot (x'(t), y'(t)) =$$

$$= x'(t) \cdot \frac{\partial f}{\partial x}(x(t), y(t)) + y'(t) \cdot \frac{\partial f}{\partial x}(x(t), y(t))$$

Exercises

- $\mathbb{R} \xrightarrow{\gamma} \mathbb{R}^2$

$t \longmapsto (\cos(t), \sin(t))$



$\mathbb{R}^2 \xrightarrow{f} \mathbb{R}$

$(x,y) \longmapsto x^2 + y^2 - 1$ ⇝ $\cos^2 x + \sin^2 x - 1 = 0$

$t \longmapsto 0$

$f \circ \gamma(t) = 0$ constant

$\left[\frac{d}{dt} f \circ \gamma(t) = 0\right] \Rightarrow \nabla f(\gamma(t)) \cdot \underbrace{\gamma'(t)}_{\text{tangent vector}} = 0$

$\gamma'(t) = (-\sin t, \cos t)$

$\nabla f(x,y) = (2x, 2y)$

$\Rightarrow \nabla f$ orthogonal to $\gamma'(t)$

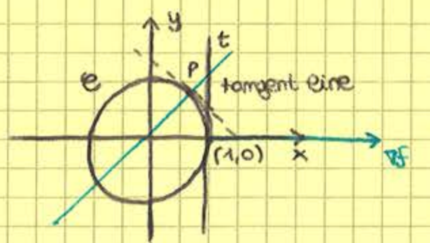
- $q(x,y) = x^2 + y^2$

$q(x,y) = 1$ CONIC

$x^2 + y^2 = 1 \iff x^2 + y^2 - 1 = 0 = f(x,y)$

$\nabla f(x,y) = (2x, 2y)$

$\nabla f(1,0) = (2,0) \Rightarrow \vec{j}$ is the tangent vector



$P = (\sqrt{2}/2, \sqrt{2}/2)$ $\quad f(x,y) = x^2 + y^2 - 1$ $\quad e: f(x,y) = 0$

$\nabla f(x,y) = (\sqrt{2}, \sqrt{2})$

$(1,-1) \cdot \nabla f(P) = 0$

$(1,-1)$ tangent direction

NOTE

$\nabla f(x,y) = (2x, 2y)$

BLUE LINE → radius

radius ⊥ tangent line

MATH II

Scalar functions of 2 and 3 variables

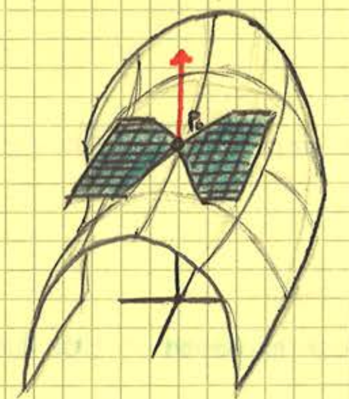
Consider $F: \mathbb{R}^3 \supseteq D \to \mathbb{R}$. Suppose the partial derivatives $F_x = \frac{\partial F}{\partial x}$, $F_y = \frac{\partial F}{\partial y}$, $F_z = \frac{\partial F}{\partial z}$ exist, and that $\nabla F = (F_x, F_y, F_z)$ is nonzero at all points of D.

Theorem
Fix $c \in f(D)$. Then $S = F^{-1}(c) = \{(x,y,z) \in \mathbb{R}^3 : F(x,y,z) = c\}$ is a smooth surface.
Moreover if $P = (x_0, y_0, z_0) \in S$ then the vector $(\nabla F)(x_0, y_0, z_0)$ is orthogonal to S at P.

Examples
If $F(x,y,z) = x^2 + 4y^2 + 16z^2$ and $c = 16$ then S is an ellipsoid. If $F(x,y,z) = x^3 - y^2$ and $c = 0$ then S contains the z-axis $(x = 0 = y)$ as a sharp edge.

Suppose that $\gamma(t)$ is a straight line such that $P_0 = \gamma(0) = (x_0, y_0, z_0) \in S$ and consider the values $F(t) = (F \circ \gamma)(t) = F(x_0 + tA, y_0 + tB, z_0 + tC)$ of F along this line.
The line is tangent to S $\iff$ $F'(0) = (\nabla F) \cdot (A, B, C)$ is zero. These tangent lines generate the tangent plane
$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0$ to S at P_0 whose normal vector
$n = (a, b, c) = (F_x(P_0), F_y(P_0), F_z(P_0))$ can therefore be taken to be the gradient computed at P_0.

Suppose that $f: \mathbb{R}^2 \supseteq D \to \mathbb{R}$. The graph of f is the surface $z = f(x,y)$. If we set $F(x,y,z) = f(x,y) - z$, it is the level surface $F(x,y,z) = 0$ and $-\nabla F = (-f_x, -f_y, 1)$ points upwards, everywhere orthogonal to the graph. (In the picture, the red arrow represents $-\nabla F$ computed at a point $P_0 = (x_0, y_0, z_0)$ on the graph.)



Fix (x_0, y_0) and set $a = f_x(x_0, y_0)$ and $b = f_y(x_0, y_0)$. The tangent plane to the graph at (x_0, y_0) has equation
$a(x - x_0) + b(y - y_0) - (z - z_0) = 0$
where $z_0 = f(x_0, y_0)$. This plane (green) approximates the graph of f at P_0 and is itself the graph

$z = z_0 + a(x - x_0) + b(y - y_0)$
of a simpler function (a linear mapping plus a constant).

Second order partial derivatives

We have seen that the first partial derivatives of a function $f: \mathbb{R}^2 \supseteq D \to \mathbb{R}$ are used to define the tangent plane to its graph at a given point P_0. This gives a first approximation to f at P_0.
One can define various higher order partial derivatives, that can be used in a similar way.
We shall concentrate on

- $\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = (f_x)_x = f_{xx}$
- $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = (f_y)_x = f_{yx}$
- $\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = (f_x)_y = f_{xy}$
- $\frac{\partial^2 f}{\partial y^2} = f_{yy}$

These are all functions, whose values at a given point are defined by limits such as

$$\left[f_{xx}(x_0, y_0) = \lim_{h \to 0} \frac{f_x(x_0 + h, y_0) - f_x(x_0, y_0)}{h} \right]$$

But it is not normally necessary to revert to this definition.

Example
If $r = \sqrt{x^2 + y^2}$ then $f(x,y) = \log r = \frac{1}{2}\log(x^2 + y^2)$ and $f_x = \frac{r_x}{r} = \frac{x}{r^2}$, $f_y = \frac{y}{r^2}$