Geometry & Linear Algebra

Lectured in Autumn 2014 by Prof. A. N. Skorobogatov at Imperial College. Humbly typed by Karim Bacchus.

Caveat Lector: unofficial notes. Comments and corrections should be sent to kb514@ic.ac.uk. Other notes available at wwwf.imperial.ac.uk/∼kb514.

Syllabus

This course details how Euclidean geometry can be developed in the framework of vector algebra, and how results in algebra can be illuminated by geometrical interpretation.

Geometry in $\mathbb{R}^2$: Vectors, lines, triangles, Cauchy-Schwarz inequality, Triangle inequality.

Matrices and linear equations: Gaussian elimination, matrix algebra, inverses, determinants of 2×2 and 3×3 matrices. Eigenvalues and eigenvectors, diagonalisation of matrices and applications. Conics and quadrics: Matrix methods, orthogonal matrices and diagonalisation of symmetric matrices.

Geometry in $\mathbb{R}^3$: Lines, planes, vector product, relations with distances, areas and volumes.

Vector spaces: Axioms, examples, linear independence and span, bases and dimension, subspaces.

Appropriate books

J. Fraleigh and R. Beauregard, *Linear Algebra*
S. Lipschutz and M. Lipson, *Linear Algebra*

# Contents

0 Introduction

Geometry Greek approach

- Undefined objects (points, lines)

- Axioms, e.g. *through any two distant points in a plane there is exactly one line.*

- Using logic you deduce *theorems*

- Proof

Contractibility

Using only a ruler and a compass e.g. perpendicular bisector, $60°$ (... *can you construct* $20°$*?*)

What can we construct using a ruler and compass, once we are given a unit length?

- Rational numbers are constructable

- So are square roots
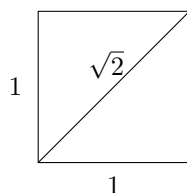
But can you construct $\sqrt[3]{2}$?

Modern Approach: Cartesian (Rene Descartes)

Plane $\longleftrightarrow$ the set of points (x_1, x_2), $x_1, x_2 \in \mathbb{R}$

Line $\longleftrightarrow$ a set of points satisfying $ax_1 + bx_2 = c$ for some given $a, b, c \in \mathbb{R}$

e.g. a circle $x_1^2 + x_2^2 = r$

Space $\longleftrightarrow \{(x_1, x_2, x_3)\, x_1, x_2, x_3 \in \mathbb{R}\}$

Planes $\longleftrightarrow \{ax_1 + bx_2 + cx_3 = d\}$
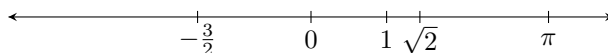
Lines $\longleftrightarrow \{ax_1 + bx_2 + cx_3 = d\}$

Advantages:

(i) Can use Analysis

(ii) Can do any dimension.

(iii) Can replace $\mathbb{R}$ by $\mathbb{C}$ or by binary fields, i.e. $\{0, 1\}$ with binary operations $1 + 1 = 0$.

1 Geometry in $\mathbb{R}^2$

$\mathbb{R}$ is the set of real numbers:

This has the properties of being:

- Commutative: $ab = ba$

- Associative: $a + (b + c) = (a + b) + c$

- Distributive: $c(a + b) = ca + cb$

- If $a < b,\ b > 0$ then $ca < cb$

> **Definition.** $\mathbb{R}^2 =$ the set of ordered pairs (a, b) where $a, b \in \mathbb{R}$ i.e. $(a, b) \neq (b, a)$ in general. Elements of $\mathbb{R}^2$ are *points*, or *vectors*. $\mathbf{0} = (0, 0) =$ origin of $\mathbb{R}^2$.
>
> A *scalar* is an element of $\mathbb{R}$.

Properties of $\mathbb{R}^2$:

(i) Addition (sum): $(a, b) + (a', b') = (a + a', b + b')$
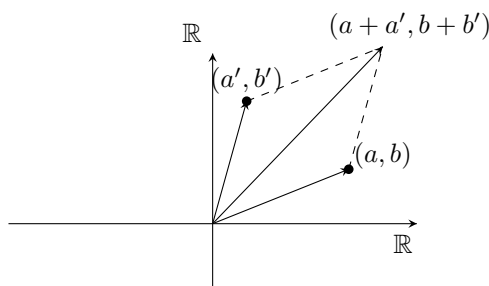
(ii) Scalar Multiplication: If $\lambda = \mathbb{R}$, $(a, b) \in \mathbb{R}^2$, then

$$\lambda \cdot (a, b) = (\lambda a, \lambda b)$$
$$v \in \mathbb{R}^2 \implies \lambda v \in \mathbb{R}^2$$

Exercise: Check $\lambda(v_1 + v_2) = \lambda v_1 + \lambda v_2$ where $\lambda \in \mathbb{R},\ v_1, v_2 \in \mathbb{R}^2$.

1.1 Lines in $\mathbb{R}^2$

> **Definition.** L is a *line in $\mathbb{R}^2$* if $\exists u, v \in \mathbb{R}^2$ $v \neq 0$ with
>
> $$L = \{u + \lambda v \mid \lambda \in \mathbb{R}\}$$

So if $u = (a_1, b_1),\ v = (a_2, b_2)$ then $L = \{(a_1 + \lambda b_1, a_2 + \lambda b_2) \mid \lambda \in \mathbb{R}\}$.

Example 1.1.
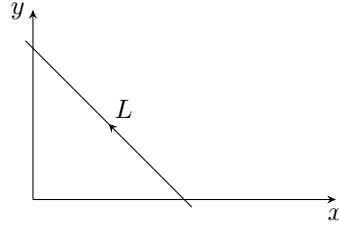
x line is $\{(0,0) + \lambda(1,0) \mid \lambda \in \mathbb{R}\}$

y line is $\{(0,0) + \lambda(0,1) \mid \lambda \in \mathbb{R}\}$

We draw the line

$$L = \{x + y = 1\}$$
$$= \{(1,0) + \lambda(-1,1) \mid \lambda \in \mathbb{R}\}.$$

Note that v is the vector for which L is parallel to.

Checking the components

$$\left.\begin{array}{l} x = 1 + (-1)\lambda = 1 - \lambda \\ y = 0 + 1 \cdot \lambda = \lambda \end{array}\right\} x + y = 1$$

Assume L, M are two lines in $\mathbb{R}^2$, with

$$L = \{u = \lambda v \mid \lambda \in \mathbb{R}\}, v \neq 0$$
$$M = \{a + \mu b \mid \mu \in \mathbb{R}\}, b \neq 0$$

Proposition 1.2. *The two lines are the same $(L = M)$ if and only if the following holds:*

(i) $v = \alpha b$ for some $\alpha \in \mathbb{R}$

(ii) $L \cap M \neq \emptyset$ (L and M have a point in common)

Proof. $(\Longrightarrow)$

(i) Assume that $L = M$. We know that

$$u \in L \implies u \in M$$
$$\implies u = a + \mu b \text{ for some } \mu \in \mathbb{R}$$

Also

$$u + v = L \ (\lambda = v) \implies u + v \in M$$
$$\implies u + v = a + \mu_1 b \text{ for some } \mu_1 \in \mathbb{R}$$
$$\implies v = (a + \mu_1 b) - (a - \mu b) = (\mu_1 - \mu)b = \alpha b.$$

(ii) Since $L = m$, surely $L \cap M = \emptyset$.

($\impliedby$) Assume $v = \alpha b \implies L = \{u + \lambda \alpha b \mid \lambda \in \mathbb{R}\}$. We also know that

$$\begin{aligned}
L \cap M \neq \emptyset &\implies \exists c \in L \cap M \\
&\implies c \in L \\
&\implies c = u + \lambda_0 \alpha b, \text{ for some } \lambda_0 \in \mathbb{R}.
\end{aligned}$$

Then also

$$\begin{aligned}
c \in M &\implies c = a + \mu b, \text{ for some } \mu \in \mathbb{R} \\
&\implies u + \lambda_0 \alpha b = a + \mu b \\
&\implies u = a + \mu b - \lambda_0 \alpha b = a + (\mu - \lambda_0 \alpha) b \qquad (*)
\end{aligned}$$

Let's consider a point inside L:

$$\begin{aligned}
L &= u + \lambda \alpha b \\
&= a + (\mu - \lambda_0 \alpha) b + \lambda \alpha b \quad \text{(using } *) \\
&= a + (\mu - \lambda_0 \alpha + \lambda \alpha) b \in M
\end{aligned}$$

So any point in L is inside M, so $L \subseteq M$. By symmetry $M \subseteq L \implies L = M$. $\blacksquare$

Parallel Lines

Let

$$\begin{aligned}
L &= \{u = \lambda v \mid \lambda \in \mathbb{R}\}, \, v \neq 0 \\
M &= \{a + \mu b \mid \mu \in \mathbb{R}\}, \, b \neq 0
\end{aligned}$$

> **Definition.** L and M are *parallel* if $\exists \alpha \in \mathbb{R}$ such that $b = \alpha v$.

Proposition 1.3. *Let x, y be two distinct points of $\mathbb{R}^2$. Then there exists a unique line which contains both x and y.*

Proof. (Existence) Choose $u - x$, $v = y - x$. Then let

$$L = \{x + \lambda(y - x) \mid \lambda \in \mathbb{R}\}$$

N.B. $v = y - x \neq 0$. We want to show that $x, y \in l$. Chose $\lambda = 0$, $x + \lambda(y = x) = x \in L$. Also choose $\lambda = 1$, $x + \lambda y - \lambda x = y \in L$. So x, y are contained in a line.

(Uniqueness) $L = \{u + lv \mid l \in \mathbb{R}\}$. L is a line through x and y. $x \in L \implies x = u + l_1 v$ for $l_1 \in \mathbb{R}$. $y \in L \implies y = u + l_2 v$ for $l_2 \in \mathbb{R}$.

$$y - x = (l_2 - l_1) v \implies v = \frac{y - x}{(l_2 - l_1)}$$

The vector that defined the direction of L is proportional to the vector $y - x$. By the question we answered before, it's enough to show that the two lines have a point in common. x, y are points in common $\implies$ Two lines are the same. $\blacksquare$

Problem: Construct $\sqrt{x}$.

$$|AD| = 1$$
$$|BD| = x$$
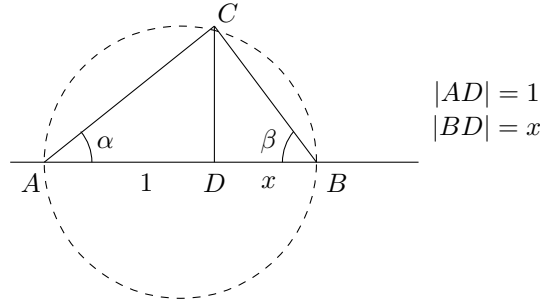
Claim: $|CD| = \sqrt{x}$.

Proof. The angle $ACD = 90°$, since AB is the diameter of a circle $\implies$ $\alpha + \beta = 90°$. Note $\triangle ABC$ is similar to $\triangle DBC$, since $CAD = DCB = \alpha$, $ACD = CBD = 90°$ and $CBA = CBD = \beta$. So we have a proportion:

$$\frac{|BD|}{|CD|} = \frac{|CD|}{|AD|}$$

$$\frac{x}{|CD|} = \frac{|CD|}{1} \implies |CD| = \sqrt{x}$$

$\blacksquare$

Recall: The plane $\mathbb{R}^2$ is the set of vectors (x_1, x_2), $x_i \in \mathbb{R}$.

For lines

$$L = \{u = \lambda v \mid \lambda \in \mathbb{R}\}, v \neq 0$$
$$M = \{a + \mu b \mid \mu \in \mathbb{R}\}, b \neq 0$$

$L \parallel M \iff v$ and b are proportional.

$L = M \iff v$ and b are proportional and they have a common point ($L \cap M \neq \emptyset$).

Proposition 1.4. *Through any two points, there is a unique line.*

Proposition 1.5. *Any two non-parallel lines meet in a unique point.*

Proof. See exercise sheet.

Example 1.6. Let

$$L = \{(0, 1) + \lambda(1, 1)\}$$
$$M = \{(4, 0) + \mu(2, 1)\}$$

The common point $= (0, 1) + \lambda(1, 1) = (4, 0) + \mu(2, 1)$.

Equating x and y's, we see from the first co-ordinate: $\lambda = 4 + 4\mu$. From

the second co-ordinate: $1 + \lambda = \mu$.

So

$$\lambda = 4 + 2(\lambda + 1)$$
$$\implies \lambda = -6$$

Therefore the common point is $(-6, -5)$.

1.2 Triangles

Definition. A triangle is a set of 3 points in $\mathbb{R}^2$ that are not on a common line.
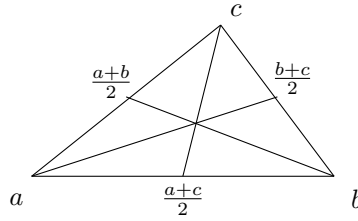
Definition (Midpoint)**.** The *midpoint* of ab is the vector $\frac{1}{2}(a + b)$.

Definition (Median)**.** A *median* is the line joining a corner to the midpoint of the opposite side.

Proposition 1.7. *The three medians of the triangle meet in a common point.*

Proof. The medians are

$$L_a = \{a + \lambda(\tfrac{b+c}{2} - a)\}$$
$$L_b = \{b + \lambda(\tfrac{a+c}{2} - b)\}$$
$$L_c = \{c + \lambda(\tfrac{a+b}{2} - c)\}$$

We can write them as

$$L_a = \{(1 - \lambda)a + \tfrac{\lambda}{2}b + \tfrac{\lambda}{2}c\}$$
$$L_b = \{(1 - \mu)b + \tfrac{\mu}{2}a + \tfrac{\mu}{2}c\}$$

Take $\lambda = \mu$, then $\frac{\mu}{2}c = \frac{\lambda}{2}c$. Also $\frac{\lambda}{2} = 1 - \mu = 1 - \lambda \implies \lambda = \frac{2}{3}$. Hence $L_a = L_b$.

Check: $\lambda = \frac{2}{3} \implies L_a = L_b$. So the common point is $\frac{1}{3}(a + b + c) \in L_a \cap L_b \cap L_c$. ∎

> **Definition** (Distance). The *length* of a vector $x = (x_1, x_2)$ is $\sqrt{x_1^2 + x_2^2}$.
> We denote the length of x by $||x|| = \sqrt{x_1^2 + x_2^2}$.
>
> The *distance* between two points $x, y \in \mathbb{R}^2$ is
>
> $$||x - y|| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

> **Definition** (Scalar product). The *scalar product* (or dot product) of
> two vectors $x, y \in \mathbb{R}^2$ is
>
> $$(x \cdot y) = x_1 y_1 + x_2 y_2$$

E.g. If $x = (1, -1)$, $y = (1, 2)$ then $(x \cdot y) = 1 + (-2) = -1$.

Easy properties: For any vectors $x, y, z \in \mathbb{R}^2$

- $x \cdot (y + z) = x \cdot y + x \cdot z$ (distribution over $+$)
- $||x||^2 = (x \cdot x) = x_1^2 + x_2^2$
- $(x \cdot y) = (y \cdot x)$ (commutativity)

Proposition 1.8 (Cauchy-Schwartz Inequality). *For $x, y \in \mathbb{R}^2$*   Lecture 3

$$|(x \cdot y)| \leq ||x|| \cdot ||y||$$

With equality iff x, y are multiples of the same vector.

Proof. w.l.o.g. $y \neq (0, 0)$. Consider all vectors of the form $x - \lambda y$. The length
$||x - \lambda y||$ squared is clearly ≥ 0, so

$$
\begin{aligned}
||x - \lambda y||^2 &= (x - \lambda y \cdot x - \lambda y) \\
&= (x \cdot x) - 2\lambda(x \cdot y) + \lambda^2(y \cdot y) \\
&= ||x||^2 - 2\lambda(x \cdot y) + \lambda^2||y|| \geq 0
\end{aligned}
$$

This is a quadratic polynomial in λ with real coefficients and at most one real
root ($= 0$). Now

$$
\begin{aligned}
\lambda^2 ||y||^2 - 2\lambda(x \cdot y) + ||x||^2 \geq 0 \implies \quad &\text{discriminant, } D \leq 0 \\
\implies &((2(x \cdot y))^2 \leq 4||x||^2 ||y||^2 \\
\implies &4(x \cdot y)^2 \leq 4||x||^2 ||y||^2 \\
\implies &|x \cdot y| \leq ||x|| ||y||
\end{aligned}
$$

The equality condition is clear for $y = 0$. Assume that y is not the zero vector.
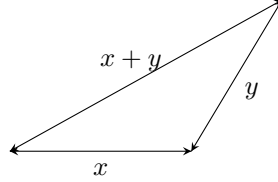If $|(x \cdot y)|^2 = ||x|| \cdot ||y||$, then $D = 0$.

$$D = 4(x \cdot y)^2 - 4||x||^2 \cdot ||y||^2 = 0$$

Then $a\lambda^2 + b\lambda + c$ has a double root, say λ_0. But $a\lambda_0^2 + b\lambda_0 + c = ||x - \lambda_0 y||^2 = 0$.
Therefore $x - \lambda_0 y$ is the zero vector. Thus $x = \lambda_0 y$.   ∎

Proposition 1.9 (Triangle Inequality)**.** *For any vectors $x, y \in \mathbb{R}^2$, we have*

$$||x + y|| \leq ||x|| + ||y||$$

If equality holds, then x and y are multiples of the same vector.

Proof. It is enough to prove that the square on LHS $\leq$ RHS, i.e.

$$||x + y||^2 \leq ||x||^2 + ||y||^2 + 2||x|| \cdot ||y||$$

We've already seen that

$$||x + y||^2 = (x + y \cdot x + y) = ||x||^2 + ||y||^2 + 2(x \cdot y)$$

By Cauchy-Schwartz $(x \cdot y) \leq ||x|| \cdot ||y||$, hence the inequality is proved. To prove the equality condition, we note that the triangle inequality is an equality precisely when $(x \cdot y) = ||x|| + ||y||$. By Theorem 1.4, we conclude that x and y are multiples of the same vector. ∎

Recall the scalar product of x as $||x||^2 = (x \cdot x)$. The scalar product can be reconstructed if we only know the lengths of vectors:

Proposition 1.10. *For any two vectors $x, y \in \mathbb{R}^2$, we have*

$$(x \cdot y) = \frac{1}{2}(||x||^2 + ||y||^2 - ||x - y||^2)$$

Proof.

$$RHS = \frac{1}{2}(||x||^2 + ||y||^2 - ||x||^2 - ||y||^2 + 2(x \cdot y))$$
$$= (x \cdot y)$$

∎

1.3 Angles

From the Cauchy-Schwartz inequality, we find that $-1 \leq \dfrac{(x \cdot y)}{||x|| \cdot ||y||} \leq 1$.
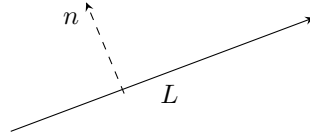
Definition. We define θ as the angle such that $\cos \theta = \dfrac{(x \cdot y)}{||x|| \cdot ||y||}$
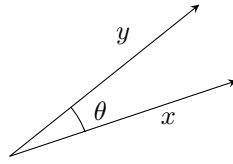
In particular, x and y are perpendicular iff $(x \cdot y) = 0$.

Lines in $\mathbb{R}^2$ can be written as $L = \{u + \lambda v \mid \lambda \in \mathbb{R}\}$. This will be referred to as *vector form*. Lines can also be described by their *Cartesian equation*: $px_1 + qx_2 + r = 0$, where $p_1, q_1 \in \mathbb{R}$.

> **Definition.** Any vector perpendicular to the direction vector of a line is called a *normal* vector to L.

If $L = \{u + \lambda v\}$, $v = (v_1, v_2)$, then $n = (-v_2, v_1)$ is a normal vector to L.

Remark. Let $n \in \mathbb{R}^2$ be any (non-zero) normal vector to $L = \{u + \lambda v\}$. Then all points $x \in L$ have the property that $(x \cdot n)$ is constant. Indeed, $x = u + \lambda v$, hence

$$(x \cdot n) = ((u + \lambda v) \cdot n) = (u \cdot n) + \underbrace{\lambda(v \cdot n)}_{= 0} = u \cdot n.$$

Therefore every point x of L satisfies the equation $n_1 x_1 + n_2 x_2 = c$, where c is a constant ($c = u_1 n_1 + u_2 n_2$). L can be given by $-v_2 x_1 + v_1 x_2 = c$ for some c.

From Cartesian to Vector form

Start with L which is the zero set of $px_1 + qx_2 + r = 0$. Now (p, q) is a normal vector to L, hence $v = (q, -p)$ is a possible direction vector of L. (($-q, p)$ is another possibility.) Then $L = \{u + \lambda v \mid \lambda \in \mathbb{R}\}$ for an appropriate vector of a. $px_1 + qx_2 + r = 0$. $((p, q) \cdot x) = $ constant. (this constant is $-r$).

The lines $L = \{u + \lambda v\}$ and $M = \{a + \mu b\}$ are perpendicular iff $(v \cdot b) = 0$. The lines $px_1 + qx_2 + r = 0$ and $ex_1 + fx_2 + g = 0$ are perpendicular iff the normal vectors to these lines e.g. (p, q) and (e, f) are perpendicular. This happens when $(p, q) \cdot (e, f) = 0$.

Claim: The vector (p, q) is a normal vector to L, given by $px_1 + qx_2 + r = 0$.   Lecture 4

Proof. We must show that (p, q) is perpendicular to $(x - y)$, where $x \in L$ and $y \in L$.

Consider
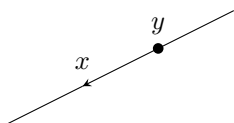
$$px_1 + qx_2 + r = 0 \qquad (1)$$

$$py_1 + qy_2 + r = 0 \qquad (2)$$

$$(1) - (2): p(x_1 - y_1) + q(x_2 - y_2) = 0$$
$$\implies (p, q) \cdot (x_1 - y_1, x_2 - y_2) = 0$$

Hence $(p, q) \cdot (x - y) = 0$. So that (p, q) is perpendicular to $(x - y)$. ∎

*"Could you read the part where Stephen
envisions life outside Dublin again?"*

2 Matrices

> **Definition.** Let n be a natural number, $n = 1, 2, 3, \ldots$. Then $\mathbb{R}^n$ is the set of all vectors with n co-ordinates $(x_1, \ldots, x_n)$.

E.g. $\mathbb{R}$ is the real line. $\mathbb{R}^2$ is the real plane. $\mathbb{R}^3$ is 3-dimensional space etc.

> **Definition** (Vector Operations)**.**
>
> Vector addition:
>
> $$(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n)$$
> $$= (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$$
>
> Multiplication by scalar $\alpha \in \mathbb{R}$:
>
> $$\alpha(x_1, \ldots, x_n) = (\alpha x_1, \ldots, \alpha x_n)$$
>
> Dot product of two vectors:
>
> $$(x_1, \ldots, x_n) \cdot (y_1, \ldots, y_n) = (x_1 y_1 + x_2 y_2 + \ldots x_n y_n)$$

2.1 System of Linear Equations

A linear equation of n vairables is of the form $a_1 x_1 + \ldots a_n x_n = b$, where $a_1, \ldots, a_n, b \in \mathbb{R}$. We want to solve systems of such linear equations.

> **Example 2.1.**
>
> $$x_1 + x_2 = 3 \tag{1}$$
> $$-2x_1 + x_2 = 1 \tag{2}$$
>
> $2 * (1) + (2)$ elimated x, leaving $3x_2 = 7 \implies x_2 = \frac{7}{3}$. Then $x_1 = 3 - x_2 = \frac{2}{3}$.

> **Example 2.2.**
>
> $$x_1 - 2x_2 = 2 \tag{1}$$
> $$3x_1 - 6x_2 = 4 \tag{2}$$
>
> $-3 \times (1) + 2$ is $0 = -2$. So this system has no solutions.

Example 2.3.

$$x_1 + x_2 = 0 \tag{1}$$
$$x_2 - x_3 = 0 \tag{2}$$
$$x_1 - x_2 + 2x_3 = 0 \tag{3}$$

Observe $(3) = (1) - 2 \times (2)$. So (1) and (2) are enough here, so (3) is automatically satisfied if (1) and (2) are satisfied. So $x_3 = a \in \mathbb{R}$, any real number. Then (2) says that $x_2 = a$. Then (1) says that $x_1 = -a$. So this system has infinitely many solutions, which are $(-a, a, a)$ for any $a \in \mathbb{R}$.

Consider a general system of linear equations of n variables $x_1, \ldots, x_n$:

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1$$
$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2$$
$$\vdots$$
$$a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m$$

A vector $(y_1, \ldots, y_n)$, $y_i \in \mathbb{R}$ is a solution of the system if all the equations are satisfied when we put $x_1 = y_1, \ldots, x_n = y_n$.

2.2 Gaussian Elimination

Example 2.4.

$$x_1 + 2x_2 + 3x_3 = 9 \tag{1}$$
$$4x_1 + 5x_2 + 6x_3 = 24 \tag{2}$$
$$3x_1 + x_2 - 2x_3 = 4 \tag{3}$$

STEP 1: Eliminate x_1 from (2) and (3). We replace (2) by $(2) - 4 \times (1)$ and replace (3) by $(3) - 3 \times (1)$. We obtain:

$$x_1 + 2x_2 + 3x_3 = 9$$
$$-3x_2 - 6x_3 = -12$$
$$-5x_2 - 11x_3 = -23$$

STEP 2: Make the coefficent of x_2 in (2) equal to 1.

$$x_1 + 2x_2 + 3x_3 = 9$$
$$x_2 + 2x_3 = 4$$
$$-5x_2 - 11x_3 = -23$$

STEP 3: Eliminate x_2 from equation (3) by adding $5 \times (2)$ to (3):

$$x_1 + 2x_2 + 3x_3 = 9$$
$$x_2 + 2x_3 = 4$$
$$-x_3 = -3$$

STEP 4: Make the coefficient of $x_3 = 1 \implies x_3 = 3$.

STEP 5: Work backwards to obtain first x_2 from (2) and then x_1 from (1). $\implies x_2 = 4 - 2x_3 = -2$ and $x_1 = 9 - 2x_2 - 3x_3 = 4$. Finally we've solved the system and we have found all the solutions of the system.

2.3 Matrices

Definition. A $m \times n$ matrix is a rectangular array of mn real numbers, arranged into m rows and n columns.

$$\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}$$

The matrix entry is written down as a_{ij} where i is the number of the row and j is the number of the column. Sometimes this matrix is denoted by (a_{ij}) or $(a_{ih}) 1 \leq 1 \leq m$, $1 \leq j \leq n$.

Basic examples of matrices

- Any n-dimensional vector is a matrix of size $1 \times n$. $(1, \ldots, n)$ is a matrix with one row.

- A column vector is $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}$. This is a matrix of size $m \times 1$.

Other examples

- $\begin{pmatrix} 1 & -1 & 5 \\ 2 & 0 & 1 \end{pmatrix}$ has size 2×3 and $\begin{pmatrix} 2 & 5 \\ 6 & 7 \\ 0 & 3 \end{pmatrix}$ has size 3×2.

- $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ is the *identity matrix* (of size 4×4).

A matrix such that $m = n$ is called a *square matrix.*

To our general linear system we attach its co-efficient matrices:

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \text{ and } \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$$

$$\implies \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & & & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{pmatrix}, \text{ the } \textit{augmented matrix.}$$

Definition (Elementary Row Operations)**.** The following operations are the *elementary row operations.*

   (i) Add a scalar multiple of one row to another: $r_i \longrightarrow r_i + \lambda r_j, \ \lambda \in \mathbb{R}$.

  (ii) Swap two rows $(r_i \leftrightarrow r_j)$

 (iii) Multiply a row by a non-zero number: $r_i \longrightarrow \lambda r_i, \ \lambda \in \mathbb{R}, \ \lambda \neq 0$.

Point: The new equations are combinations of the old ones, and vice-versa, so that the set of solutions is exactly the same.

Example 2.4 (Revisited)**.**

$$\begin{pmatrix} 1 & 2 & 3 & 9 \\ 4 & 5 & 6 & 24 \\ 3 & 1 & -2 & 4 \end{pmatrix}$$

STEP 1: $r_2 \longrightarrow r_2 - 4r_1$, $r_3 \longrightarrow r_3 - 3r_1$.

$$\begin{pmatrix} 1 & 2 & 3 & 9 \\ 0 & -3 & -6 & -12 \\ 0 & -5 & -11 & -23 \end{pmatrix}$$

STEP 2: $r_2 \longrightarrow -\frac{1}{3}r_2$.

$$\begin{pmatrix} 1 & 2 & 3 & 9 \\ 0 & 1 & 2 & 4 \\ 0 & -5 & -11 & -23 \end{pmatrix}$$

STEP 3: $r_3 \longrightarrow r_3 + 5r_2$.

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

STEP 4: $r_3 \longrightarrow -1r_3$

$$\underbrace{\begin{pmatrix} 1 & 2 & 3 & 9 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & 1 & 3 \end{pmatrix}}_{\text{Echelon form}} \implies \begin{aligned} x_1 + 2x_2 + 3x_3 &= 9 \\ x_2 + 2x_3 &= 4 \\ x_3 &= 3 \end{aligned}$$

Definition. A matrix is in *echelon form* if

(i) The first non-zero entry in each row is appears to the right on the non-zero entry in all the rows above

(ii) All rows consiting only of zeros $(0, 0, \ldots, 0)$, if any, appear at the bottom.

Examples of echelon form:

$$\begin{pmatrix} 0 & 1 & 5 & 0 \\ 0 & 0 & 5 & -3 \end{pmatrix} \quad \text{Yes} \qquad \begin{pmatrix} 6 & 2 & 4 & 0 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{Yes}$$

Non-examples

$$\begin{pmatrix} 0 & 1 & 17 \\ 1 & 0 & 2 \end{pmatrix} \quad \text{No} \qquad \begin{pmatrix} 0 & 1 & 17 \\ 0 & 1 & 5 \end{pmatrix} \quad \text{No} \qquad \begin{pmatrix} 0 & 0 & 0 \\ 1 & 2 & 5 \\ 0 & 1 & 2 \end{pmatrix} \quad \text{No}$$

Remark. If we reduce the augmented matrix of a system of linear equations to the echelon form, then we can easily solve it.

Example 2.5.

$$\begin{pmatrix} 1 & -1 & 3 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

If we have a row $(0, \ldots, 0, a)$ where $a \neq 0$ then there are no solutions. We say that the linear system is *inconsistent* since $0 \cdot x_1 + 0 \cdot x_2 + \ldots 0 \cdot x_n = a$.

Now ignore all zero-rows. To solve this linear system, we start at the bottom: $x_2 + x_3 = 1$. $x_3 = a$, where a is any real number. $x_1 = 1 - a$.

Express the variable corresponding to the first non-zero entry in this row in terms of the variables that come after it. $x_1 - x_2 + 3x_3 = 0$. Thus $x_1 = x_2 - 3x_3 = 1 - a - 3a = 1 - 4a$. The general solution is $(1 - 4a, 1 - a, a), a \in \mathbb{R}$. The variables that are assigned arbitrary values are called *free variables*, and all other variables are *dependent* variables.

Example 2.6.

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

$0 \cdot x_1 + 0 \cdot x_2 + 0 \cdot x_3 = 2$. This is inconsistent, so no solutions.

Example 2.7.

$$\begin{pmatrix} 1 & 1 & 1 & 3 & 2 & 4 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 3 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

$x_5 + x_6 = 0$, so let $x_6 = a$, $a \in \mathbb{R}$. Then $x_5 = -a$. $x_4 - x_5 = 3 \implies x_4 = 3 - a$.

$x_1 + x_2 + x_3 + 3x_4 + 2x_5 + 4x_6 = 0$. We get new free variables: $x_2 = b, b \in \mathbb{R}$ and $x_3 = c$, for any $c \in \mathbb{R}$. $\implies x_1 = -b - c + a - 9 \implies (a - b - c - 9, b, c, -a - 3, -a, a)$.

Theorem 2.8: Gaussian Elimination

Any matrix can be reduced to a matrix in echelon form by elementary row operations.

Example:

$$\begin{pmatrix} 1 & 0 & 3 & 2 & 5 & 1 \\ 1 & 0 & 3 & 5 & 8 & 1 \\ 1 & 1 & 4 & 5 & 7 & 0 \\ 1 & -1 & 2 & -1 & 3 & 2 \end{pmatrix}$$

Step 1: Clear the first column.

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

Step 2: Swap 2nd row with 4th row

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

Step 3: $r_3 \longrightarrow r_2 + r_3$

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

and we're done.

Now for the proof of the theorem:

Proof. Let $A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$ be an arbitrary $m \times n$ matrix.   Lecture 6

The basic procedure consists of two steps:

STEP 1: Find a non-zero entry furthest to the left in A. So A looks like this:

$$\begin{pmatrix} 0 & 0 & \ldots & 0 & * & \ldots \\ 0 & 0 & \ldots & 0 & * & \ldots \\ \vdots & \vdots & & \vdots & * & \\ 0 & 0 & \ldots & 0 & * & \ldots \end{pmatrix}$$

If this non-zero element is in the row with number j, then swap the first row with the j-th row ($r_i \leftrightarrow r_j$). Multiply this new row 1 to make this non-zero element equal to 1.

Outcome of Step 1:

$$\begin{pmatrix} 0 & \ldots & 0 & 1 & \ldots & * \\ 0 & \ldots & 0 & * & \ldots & * \\ \vdots & & \vdots & * & & \\ 0 & \ldots & 0 & * & \ldots & * \end{pmatrix}$$

STEP 2 Subtracting multiples of the first row from other rows to ensure that then entries underneath this 1 are all equal to 0.

Outcome of Step 2:

$$\begin{pmatrix} 0 & \ldots & 0 & 1 & * & \ldots & * \\ 0 & \ldots & 0 & 0 & * & \ldots & * \\ \vdots & & \vdots & \vdots & & A_1 & \\ 0 & \ldots & 0 & 0 & * & \ldots & * \end{pmatrix}$$

From now on we only use row operations involving rows from 2 to m. (never use row 1 anymore). We'll apply steps 1 and 2 to the "red" matrix A_1. (A_1 is A with the first j columns and 1st row removed). The outcome of steps 1 and 2 applied to A_1 is:

$$\begin{pmatrix} 0 & \ldots & 0 & 1 & * & \ldots & * & * \\ 0 & \ldots & 0 & 0 & 1 & \ldots & * & * \\ 0 & \ldots & 0 & 0 & 0 & * & \ldots & * \\ \vdots & & \vdots & \vdots & \vdots & & A_2 & \\ 0 & \ldots & 0 & 0 & 0 & * & \ldots & * \end{pmatrix}$$

Let us now consider A_2 and repeat steps 1 and 2 for A_2 and so on. The number of rows of A_1 is $m-1$, the number of rows of A_2 is $m-2, \ldots$. Hence after finitely many iterations, we'll stop. In the end the matrix will be in the echelon form! ∎

3 Matrix Algebra

Definition. The *zero matrix* of size $m \times n$ is

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

Addition

If A and B are both $m \times n$ matrices, with entries a_{ij} and b_{ij} respectively, then $A + B$ is the matrix of size $m \times n$ with entries $a_{ij} + b_{ij}$:

$$\begin{pmatrix} a_{11} + b_{11} & \ldots & a_{1n} + b_{1n} \\ \vdots & & \vdots \\ a_{m1} + b_{m1} & \ldots & a_{mn} + b_{mn} \end{pmatrix}$$

Multiply matrices by a scalar:

If $c \in \mathbb{R}$ and A is a matrix of size $m \times n$, with $A = (a_{ij})$ then cA is the matrix of size $m \times n$ with entries (ca_{ij}).

$$\text{If } A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \qquad cA = \begin{pmatrix} ca_{11} & ca_{12} \\ ca_{21} & ca_{22} \end{pmatrix}$$

Matrix Multiplication:

Let A be a matrix of size $m \times n$ and B a matrix of size $n \times k$. Then AB is a matrix of size $m \times k$ defined as follows:

$$B = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ c_1(B) & c_2(B) & \ldots & c_k(B) \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \qquad A = \begin{pmatrix} \leftarrow & r_1(A) & \rightarrow \\ \leftarrow & r_2(A) & \rightarrow \\ & \vdots & \\ \leftarrow & r_k(A) & \rightarrow \end{pmatrix}$$

Each if the row vectors $r_i(A)$ has n co-ordinates. Each of the column vectors $r_i(B)$ has n co-ordinates. Because each of $r_i(A)$ and $c_j(B)$ has exactly n co-ordinates, we can form the dot product $r_i(A) \cdot c_j(B)$ for any $1 \leq i \leq m$ $1 \leq j \leq k$. The $ij-$entry of AB is defined to be the dot prodcut $r_i(A) \cdot c_j(B)$.

Examples 3.1.

(i) 1×1 matrices are just a number, so if $A = (a)$ and $B = (b) \implies AB = (ab)$.

(ii) 2×2 matrices, $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, $B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$, then

$$AB = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{pmatrix}$$

(iii) $A = (a_{11}, a_{12}, \ldots, a_{1n})$ a $1 \times n$ matrix and $B = \begin{pmatrix} b_{11} \\ \vdots \\ b_{n1} \end{pmatrix}$ then

$$AB = (a_{11}b_{11} + a_{12}b_{21} + \cdots + a_{1n}b_{n1})$$

(which is a 1×1 matrix, i.e. a number)

(iv) BA will be a matrix of size $n \times m$:

$$\begin{pmatrix} b_{11} \\ b_{21} \\ \vdots \\ b_{n1} \end{pmatrix} (a_{11}, a_{12}, \ldots, a_{1n}) = \begin{pmatrix} b_{11}a_{111} & \ldots & b_{11}a_{1n} \\ b_{21}a_{11} & \ldots & b_{21}a_{1n} \\ \vdots & & \vdots \\ b_{n1}a_{11} & \ldots & b_{n1}a_{1n} \end{pmatrix}$$

Remark. In general multiplication of matrices is far from symmetric! $AB \neq BA$.

Recall: If A is an $m \times n$ matrix and B is an $n \times p$ matrix, then AB is an $m \times p$ matrix. The $ij-$entry is the dot product $r_i(A) \cdot c_j(B)$.

Examples 3.2.

(i)

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 4 & 3 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix}$$

So $AB \neq BA$.

(ii)

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} -1 & 0 & 1 \\ 1 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 5 \\ 1 & 4 & 11 \end{pmatrix}$$

(iii) If A has size $m \times n$ and v is a column vector with n co-ordinates, then Av is well defined. e.g.

$$\begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 1 \\ 1 & 0 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 + 2x_2 - x_3 \\ x_2 + x_3 \\ x_1 + 2x_3 \end{pmatrix}$$

Therefore the system of linear equations

$$x_1 + 2x_2 - x_3 = 1$$
$$x_2 + x_3 = -7$$
$$x_1 + 2x_3 = 0$$

can be written in matrix form. $Ax = B$, where A is the 3 matrix above and B is the vector $(1, -7, 0)^t$.

In general the system of equations

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1$$
$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2$$
$$\vdots$$
$$a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m$$

is equivalent to our matrix equation $Ax = b$, where A is the matrix with entries a_{ij}, $x = (x_1, \ldots, x_n)^t$ and $b = (b_1, \ldots, b_n)^t$.

Proposition 3.3 (Distributivity Law)**.**

(i) *Let A be an $m \times n$ matrix and let B and C be $n \times p$ matrices. Then*

$$A(B + C) = AB + AC$$

(ii) *Let D and E be $m \times n$ matrices and F be an $n \times p$ matrix. Then*

$$(D + E)F = DF + EF$$

Proof. (i) By definition, the ij-entry of $A(B + C)$ is

$$\begin{aligned} r_i(A) \cdot c_j(B + C) &= r_i(A) \cdot (c_j(B) + c_j(C)) \\ &= (r_i(A) \cdot c_j(B) + (r_i(A) \cdot c_j(C)) \\ &= ij - \text{entry of } AB + ij - \text{entry of } AC \\ &= ij - \text{entry of } AB + AC \qquad\qquad \blacksquare \end{aligned}$$

Exercise: Prove (ii).

Application to system of linear equations

Remark 1. Any system of linear equations can only have 0, 1 or infinitely many solutions.

Proof. We've seen cases when there is 0 or only 1 solution.

Suppose we have at least two distinct solutions, say p and q (column vectors with n co-ordinates). Let $\lambda \in \mathbb{R}$ be any real number. Then $A(p + \lambda(q - p)) = Ap + A(\lambda(q - p))$ by Prop. 3.3. This equals $Ap + \lambda A(q - p) = Ap + \lambda \cdot Aq - \lambda Ap$. Since p and q are solutions, we have $Ap = Aq = b$. So $A(p + \lambda(q - p)) = Ap = b$. Since $q - p$ is not the zero vector, we have have obtained infinitely many solutions in the form $p + \lambda(q - p) = x$. $\qquad \blacksquare$

Remark 2. If p is a solution to $Ax = b$, then any solution of $Ax = b$ is the sum of p and the general solution of $Ax = 0$.

Proof. Indeed if $Ax = b$, then $A(x - p) = Ax - Ap = b - b = 0$, so $x = p + (x - p)$ where $x - p$ is a solution of $Ay = 0$. Conversely if $Ay = 0$ then $A(p + y) = b$. $\quad \blacksquare$

Consider matrices A, B, C. Does $(AB)C = A(BC)$? Yes!

> **Theorem 3.4: Associativity law**
>
> Let A be a matrix of size $m \times n$, and let B be a matrix of $n \times p$ and let C be a matrix of size $p \times q$. Then
>
> $$(AB)C = A(BC)$$

Lemma 3.5. *Let $x = (x_1, \ldots, x_n)$ be a row vector and let $y = \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix}$ and B be an $n \times p$ matrix with entries b_{ij}. Then $x \cdot By = xB \cdot y$.*

Remark. We're multiplying a $(n \times p)$ matrix by a $(p \times 1)$ column vector, and also a $(1 \times n)$ row vector by a $(n \times p)$ matrix, so both products are well-defined.

Proof of Lemma. Idea: both sides are equal to $\sum b_{ij} x_i y_j$ over all values of i and j.

Consider LHS $= (x_1, \ldots, x_n) \cdot By$

$$
By = \begin{pmatrix} b_{11} & \ldots & b_{1p} \\ \vdots & & \vdots \\ b_{n1} & \ldots & b_{np} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} = \begin{pmatrix} b_{11}y_1 + \cdots + b_{1p}y_p \\ \vdots \\ b_{n1}y_1 + \cdots + b_{np}y_p \end{pmatrix}
$$

$$
= \begin{pmatrix} \sum_{j=1}^{p} b_{ij} y_j \\ \vdots \\ \sum_{j=1}^{p} b_{nj} y_j \end{pmatrix}
$$

Then

$$
LHS = (x_1, \ldots, x_n) \cdot By
$$

$$
= \sum_{i=1}^{n} \left(x_i \cdot \sum_{j=1}^{p} b_{ij} y_j \right)
$$

$$
= \sum_{i=1}^{n} \sum_{j=1}^{p} b_{ij} x_i y_j
$$

Now consider RHS:

$$
RHS = xB \cdot y
$$

$$
= (x_1 b_{11} + \ldots x_n b_{n1}, \ldots, x_1 b_{1p} + \cdots + x_n b_{np})
$$

$$
= \left(\sum_{i=1}^{n} b_{11} x_i, \ldots, \sum_{i=1}^{n} b_{ip} x_i \right)
$$

$$
= \sum_{j=1}^{p} \left(y_j \cdot \sum_{i=1}^{n} b_{ij} x_i \right)
$$

$$
= \sum_{i=1}^{n} \sum_{j=1}^{p} b_{ij} x_i y_j \qquad\qquad \blacksquare
$$

Proof of Theorem. Observation: $r_i(AB) = r_i(A)B$, indeed:

$$r_i(AB) = (r_1(A) \cdot c_1(B), \dots, r_i(A) \cdot c_p(B))$$

which is the same as

$$r_i(A)B = (\leftarrow r_i(A) \rightarrow)\,(B)$$
$$= (r_1(A) \cdot c_1(B), \dots, r_i(A) \cdot c_p(B))$$

Similarly $c_j(BC) = Bc_j(C)$. (*Exercise*, easy)

To finish the proof, note that the ij-entry of $(AB)C$ is the dot product

$$r_i(AB) \cdot c_j(C) = r_i(A)B \cdot c_j(C)$$
$$= r_i(A) \cdot Bc_i(C) \quad \text{(by Lemma 3.7)}$$
$$= \text{ the } ij - \text{entry of } A(BC). \qquad \blacksquare$$

3.1 Square matrices

Definition. A matrix of size $n \times n$ for some n is called a *square matrix*.

These are nice as we can multiply these matrices by themselves, in fact $A^2 = AA$ is only defined when A is a square matrix.

$$A^2 = (AA)A = A(AA)$$
$$\vdots$$
$$A^n = A^{n-1}A \text{ for all } n \geq 2$$

1 is a very important number as it's a number such that multiplying everything by 1 gets you the same thing. There is a similar thing for matrices:

Definition. The *identity matrix* I_n is the square matrix of size $n \times n$ given by
$$\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & 1 \end{pmatrix}$$
The ij-entry of I_n is 1 if $i = j$, and is 0 otherwise if $i \neq j$.

Proposition 3.6. *For any square matrix A of size $n \times n$, we have $AI_n = I_nA = A$.*

So the identity matrix is an analogue of 1.

Proof.

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

$$= \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

The other way is similar.   ∎

What about the inverse? For many reasons it's important to be able to invert matrices, and at this point we have to be careful as multiplication is not commutative, so there may be two inverses. Let me make a temporary definition first:

> **Definition.** B is the *left inverse* of A if $BA = I_n$.
>
> C is the *right inverse* of A if $AC = I_n$

Proposition 3.7. *Assume that A has both a right inverse C and a left inverse B. Then $B = C$.*

Proof. By Theorem 3.6, we have $(BA)C = B(AC)$. But then by Prop 3.8 we can continue on the left: $C = I_n C = (BA)C = B(AC)$. And simiarly we can continue on the right: $(BA)C = B(AC) = BI_n = B$.   ∎

Warning. Not all matrices are invertible. e.g.

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

Clearly if $A^2 = 0$, then the inverse does not exist; if there is a B such that $BA = I_2$, then $(BA)A = B(AA)$ by associativity, however $AA = A^2 = 0$, hence $B(AA) = 0$. Whereas $(BA)A = I_2 A = A$. Since $A \neq 0$ we have a contradiction. Therefore B does not exist.

A matrix with 2 identical (or proportional) rows is not invertible, e.g.

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

is not invertible. Let me go back to this example later on when we have more theory...

Application to Systems of Linear Equations

Proposition 3.8. *Let A be a square matrix. The linear system $Ax = b$ has exactly one solution when A has an inverse (that A has both left and right inverses).*

Proof. By assumption there is a square matrix B of the same size such that $BA = I_n$. Then $Ax = b$ implies $B(Ax) = Bb$. By associativity, we have

$$B(Ax) = (Ba)x = I_n x = x$$

On the other hand, this equals Bb. Therefore any solution of $Ax = b$ must be equation to Bb.

Claim: Bb is in fact a solution of $Ax = b$.

Indeed, $A(Bb) = (AB)b$. But from Prop 3.8 we know that $b = C$ for which $AC = I_n$. Hence $A(Bb) = (AB)b = I_n b$. So Bb is a solution of $Ax = b$.   ∎

Remark. If $b = 0$, then $x = 0$ is the unique solution of $Ax = 0$.
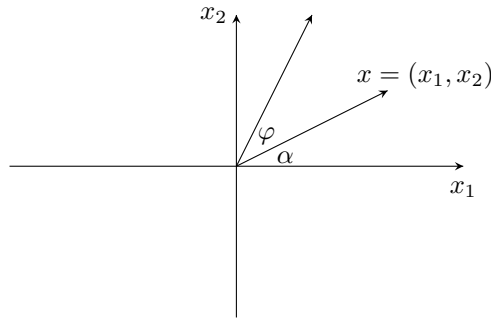

"No, Thursday's out. How about never—is never good for you?"

4 Applications of Matrices

4.1 Geometry

Rotations

Consider a rotation of φ about the origin to the vector x. The vector has components $x_1 = p\cos\alpha$, $x_2 = p\sin\alpha$. Then rotation through through the angle φ sends x to $(p\cos(\alpha + \varphi), p\sin(\alpha + \varphi))$:

$$x = \begin{pmatrix} p\cos\alpha \\ p\sin\alpha \end{pmatrix} \longmapsto \begin{pmatrix} p\cos(\alpha + \varphi) \\ p\sin(\alpha + \varphi) \end{pmatrix}$$

$$= \begin{pmatrix} p\cos\alpha\cos\varphi - p\sin\alpha\sin\varphi \\ p\cos\alpha\sin\varphi + p\sin\alpha\cos\varphi \end{pmatrix}$$

$$= \begin{pmatrix} \cos\psi x_1 + (-\sin\psi)x_2 \\ \sin\psi x_1 + \cos\psi x_2 \end{pmatrix}$$

$$= \begin{pmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Definition. Let
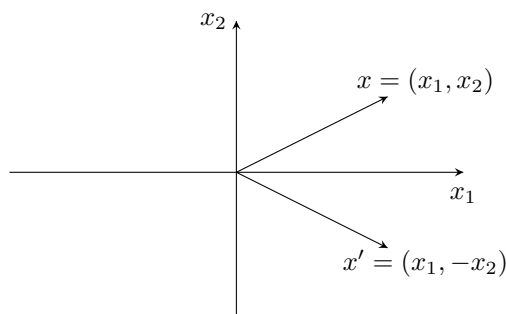$$R_\psi = \begin{pmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{pmatrix}$$
This is the *rotation matrix.*

When we rotate the vector x by ψ we write

$$x \longmapsto R_\psi x \quad \text{(This means ``x goes to $R_\psi x$)}$$
$$x \longmapsto R_\psi x \longmapsto R_\psi R_\psi x$$

Note that $R_\psi R_\psi = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Reflections

The reflection in the x-axis sends (x_1, x_2) to $(x_1, -x_2)$. In terms of matrices the reflection sends

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Definition. (See Sheet 3) The matrix

$$\begin{pmatrix} \cos\psi & \sin\psi \\ \sin\psi & -\cos\psi \end{pmatrix}$$

represents the reflection in the line $x_2 = x$, $\tan\frac{\psi}{2}$.

It's easy to check that $S_\psi^2 = I$. Also $S_\psi S\varphi = R = R_{\psi-\varphi}$. Thus the composition of two reflections is a rotation:
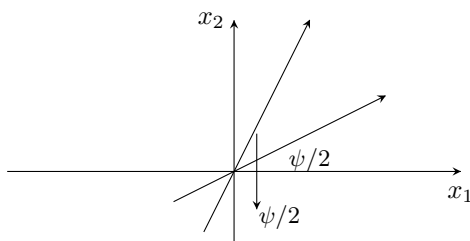
4.2 Markov Chains

Special Halloween Problem: An enchanted village is inhabited by vampires, zombies and goblins. Every full moon:

> 25% of zombies turn into vampires
>
> 10% of zombies turn into goblins
>
> 65% remain as zombies
>
> 20% of vampires turn into zombies
>
> 5% of vampires turn into goblins
>
> 75% stay as vampires

5% of goblins turn into vampires

30% of goblins turn into zombies

65% stay as goblins

Question: *What will happen after n full moons? If $n \to \infty$ are the proportions going to stabilise?*

Let x_1 be the proportion of the vampires, let x_2 and x_3 be the proportions of zombies and goblins respectively.

$$x_1 + x_2 + x_3 = 1 \quad x_i \geq 0$$

After one month the proportions are as follows:

$x_1^{(1)} =$ proportion of vampires after one month.

$$x_1^{(1)} = 0.75x_1 + 0.25x_2 + 0.05x_3$$
$$x_2^{(1)} = 0.20x_1 + 0.065x_2 + 0.30x_3$$
$$x_3^{(1)} = 0.05x_1 + 0.10x_2 + 0.65x_3$$

$$x^{(1)} = \begin{pmatrix} 0.75 & 0.25 & 0.05 \\ 0.20 & 0.65 & 0.30 \\ 0.05 & 0.10 & 0.65 \end{pmatrix} x$$

Let T be this matrix. After n months we have the population represented by

$$x^{(n)} = T^n x$$

Observe: $t_{ij} > 0$. Crucially, the sum of entries in each column of T is 1.

Exercise: For any matrix $T = (t_{ij})$ such that $t_{ij} > 0$ and $\sum_{i=1}^{n} t_{ij} = 1$, if $x = (x_1, \ldots, x_n)^t$ is such that $x_i \geq 0$ and $\sum_{i=1}^{k} x_i = 1$, then Tx has the same properties as x.

This is an example of a *Markov Chain*.

Theorem 4.1

If all entries of some power of T are positive, then there exists a unique vector

$$s = \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix}$$

such that $s_i \geq 0$ and $\sum_{i=1}^{n} s_i = 1$ for which $Ts = s$ (the *stationary state*). For any initial state x, this converges to s. Thus s is an eigenvector of T with eigenvalue 1.

$Ts = s$, or equivalently $(T - I) = 0$ is a system of linear equations. IN our case, solving this system gives

$$s = (\tfrac{37}{86}, \tfrac{34}{86}, \tfrac{15}{86}) = (0.45, 0.4, 0.17)$$

This is the limit distribution of three species.

5 Inverses

> ### Theorem 5.1
>
> Let A be a square matrix. If there exists a square matrix B such that $AB = I$ then this B is unique and satisfies $BA = I$.

Proof. (Also gives a method for finding this B!)

Let X be the square matrix with unknown entires. We want to solve the equation $AX = I$. The entires of X are x_{ij}. We have n^2 unknowns and n^2 equations. We record this as follows: $(A \mid I)$, a $n \times 2n$ matrix.

This is n systems of linear equations in n variables, e.g. for each column of I we have the following system:

$$A \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ \vdots \\ x_{nj} \end{pmatrix} = \text{ the } j\text{th column of } I = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

All these n systems have the same co-efficient matrix A, so we solve them using the same process. Apply reduction to echelon form. Perform elementary row operations on the matrix $(A \mid I)$

Claim: The matrix on the LHS cannot have any rows made entirely of zeros.

Proof of Claim. Remember that D is obtained by row operations from I. We know that two matrices that are obtained from each other by row operations define equivalent linear systems. This means that the linear system $I(y_1, \ldots, y_n) = 0$ has the same solutions as $D(y_1, \ldots, y_n) = 0$. But $(0, \ldots, 0)$ is the only solution to this. Now if D has an all zero row, the system $D(y_1, \ldots, y_n) = 0$ has free variables, hence infinitely many solutions. This contradiction proves that D does not have an all zero row. ∎

Therefore there is a non-zero entry in the bottom row of D. Say this entry is in the jth column. Then the system (1.1) has no solutions (follows from the echelon form method). Therefore, if the matrix on the left has a bottom row made of zeros, then $AX = I$ has no solutions. So A has on right inverse. It remains to consider the case when the matrix on the left has no all-zero rows:

$$(A_{ech} \mid I) = \left(\begin{matrix} 1 & & & & \\ & 1 & & * & \\ & & \ddots & & \\ 0 & & & 1 & \end{matrix} \right. \left| \begin{matrix} \\ \\ D \\ \\ \end{matrix} \right)$$

Perform more elementary row operations to clear the entries above the main diagonal (This is possible because all diagonal entries equal 1). After this step,

we obtain:

$$(I \mid E) = \left(\begin{array}{cccc} 1 & & & \\ & 1 & & \mathbf{0} \\ & & \ddots & \\ \mathbf{0} & & & 1 \end{array} \middle| \; E \right)$$

Since row operations don't change the solution of our linear system, we have $IX = E$. Hence E is a unique solution of the system $AX = I$, i.e. $AE = I$. We've proven that if the right inverse exists it can be obtained by the procedure, and it is unique.

Finally we now prove that $EA = I$:

Consider the equation $EY = I$, where $Y = (y_{ij})$ is a square matrix with unknown entires y_{ij}. Reverse row operations from the first part of the proof to so $(E \mid I) \mapsto (I \mid A)$. $EY = I$ is equivalent to $IY = A$, that is $Y = A$. Therefore $EA = I$. $\blacksquare$

Finding inverses of 2×2 matrices is easy:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \text{ consider } B = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Then

$$AB = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

$$= \begin{pmatrix} ad - bc & 0 \\ 0 & ad - bc \end{pmatrix}$$

$$= (ad - bc)I$$

Case 1. $(ad - bc)$ is non-zero. Then

$$\frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \text{ is the inverse of } A$$

Case 2. $ad - bc = 0$. In this case $AB = 0$. Then the inverse does not exist (If $CA = I$, then $C(AB) = (CA)B = IB = B$. If A is non-zero then B is non-zero, so we get a contradiction for $0 = B$.) Hence A is not invertible.

Let A be a square matrix. If   Lecture 11

$$(A \mid I) \overset{\text{row ops}}{\longrightarrow} \left(\begin{array}{c} * \\ 0 \dots 0 \end{array} \middle| \; * \right)$$

then A has no inverse.

If

$$(A \mid I) \overset{\text{row ops}}{\longrightarrow} (I \mid E)$$

then $AE = I$.

Example 5.2. Let

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

$$(A \mid E) = \left(\begin{array}{ccc|ccc} 1 & 3 & -2 & 1 & 0 & 0 \\ 2 & 5 & -3 & 0 & 1 & 0 \\ -3 & 2 & -4 & 0 & 0 & 1 \end{array} \right)$$

$$\longrightarrow \left(\begin{array}{ccc|ccc} 1 & 3 & -2 & 1 & 0 & 0 \\ 0 & -1 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -19 & 11 & 1 \end{array} \right)$$

$$\longrightarrow \left(\begin{array}{ccc|ccc} 1 & 3 & 0 & -37 & 22 & 2 \\ 0 & -1 & 0 & 17 & -10 & -1 \\ 0 & 0 & 1 & -19 & 11 & 1 \end{array} \right)$$

$$\longrightarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 14 & -8 & -1 \\ 0 & 1 & 0 & -17 & 10 & 1 \\ 0 & 0 & 1 & -19 & 11 & 1 \end{array} \right)$$

So

$$A^{-1} = \begin{pmatrix} 14 & -8 & -1 \\ -17 & 10 & 1 \\ -19 & 11 & 1 \end{pmatrix}$$

Recall that Theorem 5.1 says that if there exists a B such that $AB = I$, then $BA = I$.

Question: *What if there is a matrix C such that $CA = I$: does this imply that $CA = I$?* Yes.

5.1 Transpose

Definition. Let A be a matrix of size $n \times m$. The *transpose* of A is the matrix A^t of size $m \times n$ such that the ij-entry of A^t is the $ji-$entry of A.

E.g. if $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then $A^t = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$.

Some properties:

- $I^t = I$

- $(x_1, \ldots, x_n)^t = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

- $(A^t)^t = A$

- $r_i(A^t) = c_i(A)^t$

- $c_j(A^t) = r_j(A)^t$

Proposition 5.3. $(AB)^t = B^t A^t$

Proof.

$$
\begin{aligned}
\text{The } ij\text{-entry of } (AB)^t &= \text{ the } ji\text{-entry of } AB \\
&= r_j(A) \cdot c_i(B) \\
&= c_i(B) \cdot r_j(A) \\
&= r_i(B^t) \cdot c_j(A^t) \\
&= ij\text{-entry of } B^t A^t \qquad \blacksquare
\end{aligned}
$$

Corollary 5.4. *Let A be a square matrix. If there exists a matrix C such that $CA = I$, then $AC = I$.*

Proof. Suppose we have C such that $CA = I$. Then $(CA)^t = I^t = I$. By Prop 5.2, we have $A^t \cdot C^t = I$ By Theorem 5.1, this implies that $C^t A^t = I$. By Prop 5.2, we obtain $(AC)^t = I$. Apply the transpose to both sides:

$$
AC = ((AC)^t)^t = I^t = I \qquad \blacksquare
$$

Remark. Let A be a square matrix. Then the following properties hold.

(i) If there exists a matrix B such that $AB = I$, then $BA = I$.

(ii) If there exists a matrix C such that $CA = I$, then $AC = I$. Call $B = C = A^{-1}$ the inverse of A.

(iii) If A^{-1} exists, it is unique.

(iv) A^{-1} exists if and only if A can be reduced to I by elementary row operations

(v) A^{-1} exists if and only if $Ax = b$ has a unique solution.

(vi) A^{-1} exists if and only if $Ax = 0$ has a unique solution 0.
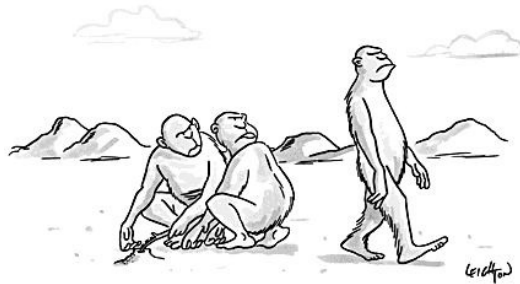
"Well, if it isn't the dawn of civilization."

6 Determinants

Definition (2×2 determinant)**.**

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

Also denoted by $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$

We've seen that if $\det(A) \neq 0$, then A^{-1} exists.

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

If $\det(A) = 0$, then A^{-1} doesn't exist.

Examples 6.1.

$$\det \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} = ad$$

$$R_\varphi = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}$$

$\det R_\varphi = 1$.

$$S_\varphi = \begin{pmatrix} \cos\varphi & \sin\varphi \\ \sin\varphi & -\cos\varphi \end{pmatrix}$$

$\det S_\varphi = -1$.

Exercise: $\det(AB) = \det(A)\det(B)$.

6.1 3×3 determinants

Definition. For a 3×3 matrix, A:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

$$\det(A) = a_{11} \det \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} - a_{12} \det \begin{pmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{pmatrix} + a_{13} \begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}$$

$$= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{21}a_{31}).$$

The same decomposition works for $n \times n$ matrices.

Let A be a square matrix.   Lecture 12

> **Definition.** The ij-minor of A is the square matrix obtained from A by removing the i-th row and the j-th column.

Example 6.2. If
$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 5 & 0 & 8 \end{pmatrix}$$
then
$$A_{11} = \begin{pmatrix} 5 & 6 \\ 0 & 8 \end{pmatrix} \quad A_{13} = \begin{pmatrix} 4 & 5 \\ 5 & 0 \end{pmatrix}$$

We defined $\det(A) = a_{11} \det(A_{11}) - a_{12} \det(A_{12}) + a_{13} \det(A_{13})$.

Proposition 6.3. *Let A be a 3×3 matrix.*

 (i) $\det(A) =$ *expansion in the 2nd row.*

 (ii) $\det(A) =$ *expansion in 3rd row.*

Proof of (i). Note that

$\det(A) = a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{21}a_{31}).$

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

We need to compare this with the following expression:

$$-a_{21}(a_{12}a_{33} - a_{32}a_{13}) + a_{22}(a_{11}a_{33} - a_{31}a_{13}) - a_{23}(a_{11}a_{32} - a_{31}a_{12})$$

Proof of (ii) is similar. ∎

Remark. There are also expansions in the first column, second column etc.

$$\det(A) = a_{11} \det(A_{11}) - a_{21} \det(A_{21}) + a_{31} \det(A_{31})$$

Goal: Show that A^{-1} exists precisely when $\det(A) \neq 0$.

Lemma 6.4. *If a 3×3 matrix A has two identical rows, then $\det(A) = 0$.*

Proof.

$$\det \begin{pmatrix} a & b & c \\ a & b & c \\ x & y & z \end{pmatrix} = \text{ expansion in 3rd row}$$

$$= x \det \begin{pmatrix} b & c \\ b & c \end{pmatrix} - y \det \begin{pmatrix} a & c \\ a & c \end{pmatrix} + z \det \begin{pmatrix} a & b \\ a & b \end{pmatrix}$$

$$= 0$$

Other cases are similar. ∎

Proposition 6.5 (Effect of row operations on determinant)**.**

(i) *Replacing $r_1(A)$ by $r_i(A) + \lambda r_j(A), j \neq i$ does not change $\det(A)$*

(ii) *Swapping $r_i(A)$ and $r_j(A), j \neq i$ multiplies $\det(A)$ by -1*

(iii) *Multiplying $r_i(A)$ by $\lambda \neq 0$ has the effect of multiplying $\det(A)$ by λ.*

Proof. (i) For simplicity assume that $i = 2$ and $j = 1$:

$$A' = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} + \lambda a_{11} & a_{22} + \lambda a_{12} & a_{23} + \lambda a_{13} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

Let's expand this det in the 2nd row:

$$\det(A') = -(a_{21} + \lambda a_{11})\det(A_{21}) + (a_{21} + \lambda_{11})\det(A_{22}) - (a_{23} + \lambda_{13})\det(A_{23})$$

$$= \det(A) + \lambda \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

$$= \det(A) + 0$$

$$= \det(A)$$

(ii) We have:

$$\det \begin{pmatrix} a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

$$= \text{ expand 3rd row}$$

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

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

$$= -\det(A)$$

(iii) Expand in the row that is multiplied by λ. This immediately gives the result.   ∎

Examples 6.6. (i)

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

$$= \text{ expand in the 1st column}$$

$$= \det \begin{pmatrix} -4 & 5 \\ 5 & -1 \end{pmatrix}$$

$$= 4 - 25 = -21$$

(ii) Van der Monde determinant:

$$\det \begin{pmatrix} 1 & x & x^2 \\ 1 & y & y^2 \\ 1 & z & z^2 \end{pmatrix} = \det \begin{pmatrix} 1 & x & x^2 \\ 0 & 1 & x+y \\ 0 & 1 & x+z \end{pmatrix}$$

$$= (y-x) \det \begin{pmatrix} 1 & x & x^2 \\ 0 & 1 & y+x \\ 0 & z-x & z^2-x^2 \end{pmatrix}$$

$$= (y-x)(z-x) \det \begin{pmatrix} 1 & x & x^2 \\ 0 & 1 & x+y \\ 0 & 1 & x+z \end{pmatrix}$$

$$= (y-x)(z-x) \det \begin{pmatrix} 1 & x & x^2 \\ 0 & 1 & x+y \\ 0 & 0 & z-y \end{pmatrix}$$

$$= (y-x)(z-x)(z-y)$$

$$= (x-y)(y-z)(z-x)$$

Corollary 6.7. *If A' is obtained from A by row operations, then $\det(A') \neq 0$ if and only if $\det(A) \neq 0$.*

Proof. A direct consequence of Proposition 6.3.   ∎

Theorem 6.8

Let A be a 3×3 matrix. Then A^{-1} exists if and only if $\det(A) \neq 0$.

Proof. Recall that A can be reduced to echelon form by row operations. Let A' be the matrix in echelon form to which A reduces. Then $\det(A') \neq 0 \iff \det(A) \neq 0$. Hence we are in Case 2 (If A' has an all zero-row then we expand in this row and $\det(A') = 0$.) In Case 2, A' can be reduced to I by further row operations. By Theorem 5.1, A is invertible.

We need to prove that if A^{-1} exists, then $\det(A) \neq 0$. Indeed, if A^{-1} exists, then the echelon form of A has no all-zero rows. Then A can be reduced to I by row operations. Row operations can only multiply det by a non-zero number, and they can be reversed. Therefore, $\det(A) \neq 0$.   ∎

Lecture 13

Remark. For any square matrices A and B of the same size $\det(AB) = \det(A)\det(B)$. If A^{-1} exists, then $AA^{-1} = I$, so $\det(A)\det(A^{-1}) = 1$. Hence $\det(A) \neq 0$ if A^{-1} exists.

Final Comment. If A is a square matrix, then $Ax = 0$ has non-zero solutions if and only if $\det(A) = 0$. (Indeed if $Ax = 0$ has a non-zero solution, then it has at least two distinct solutions, so it has infinitely many solutions. Then A^{-1} does't exist, and $\det(A) = 0$).

7 Eigenvalues and Eigenvectors

> **Definition.** Let A be a $n \times n$ matrix. Then a non-zero vector, v, is called an *eigenvector* of A if $Av = \lambda v$ for some $\lambda \in \mathbb{R}$. In this case λ is called an *eigenvalue* of A corresponding to the eigenvector v.

Remarks. A scalar multiple of an eigenvector is also an eigenvector with the same eigenvalue.

> **Example 7.1.** If $A = \begin{pmatrix} 3 & 2 \\ 2 & 0 \end{pmatrix}$, $v = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$, we have
>
> $$Av = \begin{pmatrix} 3 & 2 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ -2 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix} = -v$$
>
> So v is an eigenvector of A with eigenvalue -1.
>
> But if $w = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, then
>
> $$Aw = \begin{pmatrix} 3 & 2 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \end{pmatrix} \neq \lambda \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$
>
> So w is not an eigenvector of A.

Remark. If A is a scalar multiple of I, $A = c \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, then every non-zero vector is an eigenvector with eigenvalue c.

How do we find eigenvectors and eigenvalues?

v is an eigenvector of A if and only if $Av = \lambda v$ for some $\lambda \in \mathbb{R}$. $Av = \lambda \cdot Iv \implies (\lambda I - A)v = 0$. Here $v \neq 0$. Non-zero solution (such as v) of this system of linear equations exist if and only if $\det(\lambda I - A) = 0$.

Let t be a variable and consider the matrix $tI - A$:

$$tI - A = \begin{pmatrix} t - a_{11} & -a_{12} & \ldots & -a_{1n} \\ -a_{21} & t - a_{12} & \ldots & -a_{2n} \\ \vdots & & & \vdots \\ -a_{n1} & t - a_{n2} & \ldots & t - a_{nn} \end{pmatrix}$$

> **Definition.** The determinant of $tI - A$ is called the *characteristic polynomial* of A. (For us $n = 3$, or $n = 2$)

For example, if $n = 2$, then the characteristic polynomial of A is

$$\det \begin{pmatrix} t - a_{11} & -a_{12} \\ -a_{21} & t - a_{22} \end{pmatrix} = (t - a_{11})(t - a_{22}) - a_{21}a - 12$$

$$= t^2 - (a_{11} + a_{22}) + a_{11}a_{22} - a_{21}a_{12}$$

$$= t^2 - tr(A) + \det(A) = 0$$

Definition. $tr(A)$ is the *trace*, defined as $tr(A) = a_{11} + a_{22}$

Proposition 7.2. *Let A be a 2×2 or 3×3 matrix. Then the eigenvalues of A are the roots of the characteristic polynomial of A, i.e. every eigenvalue λ satisfies $\det(\lambda I - A) = 0$. The eigenvectors of A with eigenvalue λ are non-zero solutions of the system of linear equations $(\lambda I - A)v = 0$.*

Proof. The real numbers λ for which $\det(\lambda I - A) = 0$ are by definition the roots of the characteristic polynomial of A. Hence v is a non-zero solution of $(\lambda I - A)v = 0$. ∎

Example 7.3. Consider

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

(i) Find the characteristic polynomial of A:

$$\det \begin{pmatrix} t - 3 & -2 \\ -2 & t \end{pmatrix} = t^2 - 3t - 4 = 0$$

$$\implies (t + 1)(t - 4) = 0$$

(ii) Solve this equation $(t + 1)(t - 4) = 0$, and find the eigenvalues:

$$\lambda_1 = -1, \; \lambda_2 = 4$$

(iii) Find the eigenvectors with eigenvalue -1:

$$\lambda_1 I - A = \begin{pmatrix} -4 & -2 \\ -2 & -1 \end{pmatrix}$$

Solve

$$\begin{pmatrix} -4 & -2 \\ -2 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0$$

$$\left. \begin{array}{c} -4x_1 - 2x_2 = 0 \\ -2x_1 - x_2 = 0 \end{array} \right\} \implies v_1 = \begin{pmatrix} 1 & -2 \end{pmatrix} \text{ (or any scalar multiple)}$$

Find eigenvectors with eigenvalue 4:

$$\lambda_2 I - A = \begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix}$$

Solve

$$\begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0$$

$$\left. \begin{array}{c} x_1 - 2x_2 = 0 \\ -2x_1 + 4x_2 = 0 \end{array} \right\} \implies v_2 = \begin{pmatrix} 2 & 1 \end{pmatrix} \text{ (or any scalar multiple)}$$

> **Conclusion:** Up to proportionality, A has exactly two eigenvectors with eigenvalue -1 and 4.

Warning. Some 2×2 matrices have only one eigenvector up to scalar multiples!

Example 7.4.
$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

This has characteristic polynomial of

$$\det \begin{pmatrix} t-1 & 1 \\ 0 & t-1 \end{pmatrix} = (t-1)^2$$

So 1 is the unique eigenvalue of A. Now

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

So solving

$$\begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \implies x_2 = 0$$

So $v = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ is the unique eigenvector (up to scalar multiples!)

If A is a 3×3 matrix, then the characteristic polynomial of A is as polynomial of degree 3, with leading coefficient 1. Typically, A will have three different eigenvalues (which may be complex numbers!). There are matrices with 3 non-proportional eigenvectors, but also matrices with fewer eigenvectors, e.g.

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

7.1 Diagonalisation of Matrices

Definition. A matrix is called *diagonal* if all non-diagonal entries are zero:

Lecture 14

$$D = \begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \dots & d_n \end{pmatrix}$$

Properties

- $\operatorname{diag}(d_1, \dots, d_n) + \operatorname{diag}(d_1', \dots, d_n') = \operatorname{diag}(d_1 + d_1', \dots, d_n + d_n')$
- $\operatorname{diag}(d_1, \dots, d_n)\operatorname{diag}(d_1', \dots, d_n') = \operatorname{diag}(d_1 d_1', \dots, d_n d_n')$

- $\operatorname{diag}(d_1, \ldots, d_n)^m = \operatorname{diag}(d_1^m, \ldots, d_n^m)$

- $\det(\operatorname{diag}(d_1, \ldots, d_n)) = d_1 \ldots d_n$

- If $d_1, \ldots, d_n \neq 0$ then $\operatorname{diag}(d_1, \ldots, d_n)^{-1} = \operatorname{diag}(d_1^{-1}, \ldots, d_n^{-1})$

Consider the polynomial

$$p_m x^m + \cdots + p_0 \quad p_i \in \mathbb{R}$$

Then

$$p(A) = p_m A^m + p_{m-1} A^{m-1} + \cdots + p_1 A + p_0 I$$

where A is a square matrix

$D = \operatorname{diag}(d_1, \ldots, d_n)$ and $p(D) = \operatorname{diag}(p(d_1), \ldots, p(d_n))$.

*Can we somehow reduce general matrices to diagonal form?* Yes, in most cases...

Terminology. Let $v_1, \ldots, v_n$ be vectors in $\mathbb{R}^n$.

> **Definition.** $v_1, \ldots, v_n$ are *linearly dependent* if $\not\exists x_1, \ldots, x_n \in \mathbb{R}$ not all equal to 0 such that $x_1 v_1 + \cdots + x_n v_n = 0$.
>
> In the opposite case, $v_1, \ldots, v_n$ are called *linearly independent*.

Examples 7.5.

(i) If one of the vectors is a zero vector, they are linearly dependent, say $v_1 = 0$. Then take $x_1 = 1, x_2 = \cdots = x_n = 0$. $1 \cdot 0 + 0 \cdot v_2 + \ldots 0 \cdot v_n = 0$.

(ii) For $n = 2$, then are v_1 and v_2 linearly dependent? $x_1 v_1 + x_2 v_2 = 0$, say $x_1 \neq 0$, then $v_1 = \frac{x_2}{x_1} v_2$. Conclusion: v_1 and v_2 are linearly dependent iff they are proportional, i.e. v_2 is in this line:
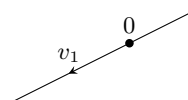
(iii) For $n = 3$, v_1, v_2, v_3 are linearly dependent iff $x_1 v_1 + x_2 v_2 + x_3 v_3 = 0$ and one of the coefficients $\neq 0$, say $x_1 \neq 0$. Then $v_1 = -\frac{x_2}{x_1} v_2 - \frac{x_3}{x_1} v_3$. In $\mathbb{R}^3$, v_1 lies in the plane through $0, v_2, v_3$:
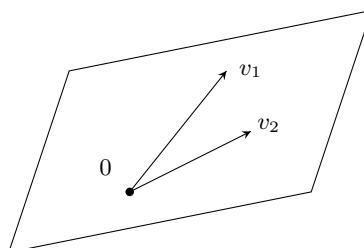
Conclusion: $v_1, v_2, v_3 \in \mathbb{R}^3$ are linearly dependent iff they belong to a plane through 0.

Lemma 7.6. *(i) Let $v_1, v_2 \in \mathbb{R}^2$. Define a 2×2 matrix B such that v_1, v_2 are the columns of B:*

$$B = \begin{pmatrix} | & | \\ v_1 & v_2 \\ | & | \end{pmatrix}$$

Then $\det(B) = 0$ iff v_1 and v_2 are linearly dependent.

(ii) Let $v_1, v_2, v_3 \in \mathbb{R}^3$ and define a 3×3 matrix B so that v_1, v_2, v_3 are the columns of B. Then $\det(B) = 0$ iff v_1, v_2, v_3 are linearly dependent.

Proof of (ii). Note that

$$B \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

$$= x_1 v_1 + x_2 v_2 + x_3 v_3$$

We know that $\det(B) = 0 \iff Bx = 0$ has a non-zero solution, where $x = (x_1, x_2, x_3)^t \iff x_1 v_1 + x_2 v_2 + x_3 v_3 = 0$ where some $x_1 \neq 0 \iff v_1, v_2, v_3$ are linearly dependent. ∎

Theorem 7.7

Let A be a square matrix (3×3 case) with eigenvectors v_1, v_2, v_3 with eigenvalues $\lambda_1, \lambda_2, \lambda_3$ respectively. Define

$$D = \text{diag}(\lambda_1, \lambda_2, \lambda_3) = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$$

P as a matrix with columns v_1, v_2, v_3.

a) Then $AP = PD$

b) If v_1, v_2, v_3 are linearly independent, then P^{-1} exists and $A = PDP^{-1}$

Remark. This is a theorem about the diagonalisation of matrices

Proof. a) What is AP? The i-th column of AP is $c_i(AP) = Ac_i(P)$ (lecture 8)

$$c_i(AP) = Ac_i(P) = Av_i = \lambda_i v_i$$

What is PD?

$$PD = \begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$$

$$= \begin{pmatrix} | & | & | \\ \lambda_1 v_1 & \lambda_2 v_2 & \lambda_3 v_3 \\ | & | & | \end{pmatrix}$$

Hence $AP = PD$.

b) By Lemma 7.2, P^{-1} exists iff v_1, v_2, v_3 are linearly independent.   ∎

> **Definition.** Two square matrices A and B are called *equivalent* if there exists an invertible matrix P such that $A = PBP^{-1}$.

Exercise: Check that this is indeed an equivalence relation.

> **Definition.** A square matrix A is *diagonalisable* if A is equivalent to a diagonal matrix, i.e. $\exists$ an invertible matrix P such that $A = PDP^{-1}$, where D is diagonal.

Proposition 7.8.

(i) If A is a diagonalisable 2×2 matrix, then A has two linearly independent eigenvectors (not multiples of each other)

(ii) If A is a diagonalisable 3×3 matrix, then A has three linearly independent eigenvectors.

Remark. This is in fact an "if and only if" statement (as it follows from Theorem 7.3)

Proof of (ii). $A = PDP^{-1}$ where $D = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3)$. Then $AP = PD$. Next, consider the columns of P as three vectors, call them v_1, v_2, v_3.

$$
PD = \begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}
$$

$$
= \begin{pmatrix} \lambda_1 p_{11} & \lambda_2 p_{12} & \lambda_3 p_{13} \\ \lambda_1 p_{21} & \lambda_2 p_{22} & \lambda_3 p_{23} \\ \lambda_1 p_{31} & \lambda_2 p_{32} & \lambda_3 p_{33} \end{pmatrix}
$$

$$
= \begin{pmatrix} | & | & | \\ \lambda_1 v_1 & \lambda_2 v_2 & \lambda_3 v_3 \\ | & | & | \end{pmatrix}
$$

Now $AP = PD \implies c_i(PD) = \lambda_i v_i \implies c_i(AP) = Ac_i(P) = Av_i$ for $i = 1, 2, 3$. Hence v_i is an eigenvector of A with eigenvalue λ_i.

Claim: v_1, v_2, v_3 are linearly independent. In fact, Lemma 7.2 says P^{-1} exists iff they are linearly independent. Since P^{-1} exists, we're done.   ∎

Corollary 7.9.

(i) *All matrices that are equivalent to*

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

for $\lambda \in \mathbb{R}$, are non-diagonalisable.

(ii) All matrices equivalent to

$$J_{\lambda_1 \mu} = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \mu \end{pmatrix}$$

for any $\lambda_1, \mu \in \mathbb{R}$ are also non-diagonalisable.

Remark. These are in fact all the non-diagonalisable matrices.

Proof. (i) For contradiction, suppose that $A = BJ_\lambda B^{-1}$ (for some invertible matrix B) is diagonalisable. Then $A = PDP^{-1}$, so $BJ_\lambda B^{-1} = PDP^{-1} \implies J_\lambda = B^{-1}PDP^{-1}B$. We know that $(B^{-1}P)^{-1} = P^{-1}(B^{-1})^{-1} = P^{-1}B$. Therefore $J_\lambda = (B^{-1}P)D(B^{-1}P)^{-1}$.

So J_λ is diagonalisable and by Prop 7.4, J_λ must have three linearly independent eigenvectors. Let's find them:

$$tI - J_\lambda = \begin{pmatrix} t - \lambda & -1 & 0 \\ 0 & t - \lambda & -1 \\ 0 & 0 & t - \lambda \end{pmatrix}$$

$$\det(tI - J_\lambda) = (t - \lambda)^3$$

Hence λ is the only eigenvalue of J_λ. The corresponding eigenvectors satisfy $(\lambda I - J_\lambda)v = 0$, i.e.

$$\begin{pmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$\implies x_2 = x_3 = 0, x_1 = a \in \mathbb{R}$. Hence we conclude

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

up to scalar multiples is the only eigenvector. So there are not three linearly independent eigenvectors, a contradiction. Thus $A = BJ_\lambda B^{-1}$ is not a diagonalisable matrix. ∎

Exercise: Prove (ii) similarly.

Examples 7.10. (1) Let

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

In lecture 13, we found this has eigenvalues $\lambda_1 = 4$, $\lambda_2 - 1$. The corresponding eigenvectors are

$$v_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad v_2 = \begin{pmatrix} 1 & -2 \end{pmatrix}$$

So $A = PDP^{-1}$ where

$$D = \begin{pmatrix} 4 & 0 \\ 0 & -1 \end{pmatrix} \quad P = \begin{pmatrix} 2 & 1 \\ -1 & -2 \end{pmatrix}$$

So $A^2 = (PDP^{-1})(PDP^{-1}) = PD^2P^{-1}$. (By induction $A^n = PD^nP^{-1}$ since $PP^{-1} = I$).

$$P^{-1} = \frac{1}{-5}\begin{pmatrix} -2 & -1 \\ -1 & 2 \end{pmatrix} = \frac{1}{5}\begin{pmatrix} 2 & 1 \\ 1 & -2 \end{pmatrix}$$

Therefore, $A^n = PD^nP^{-1}$

$$\implies A^n = \frac{1}{5}\begin{pmatrix} 2 & 1 \\ 1 & -2 \end{pmatrix}\begin{pmatrix} 4^n & 0 \\ 0 & (-1)^n \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 1 & -2 \end{pmatrix}$$

$$= \frac{1}{5}\begin{pmatrix} 2 \cdot 4^n & (-1)^n \\ 4^n & -2(-1)^n \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 1 & -2 \end{pmatrix}$$

$$= \frac{1}{5}\begin{pmatrix} 4^{n=1} + (-1)^n & 2 \cdot 4^n - 2(-1)^n \\ 2 \cdot 4^n - 2(-1)^n & 4^n + 4(-1)^n \end{pmatrix}$$

(2) How do we find a 2×2 matrix B such that $B^3 = A$?

Idea: Take any 2×2 matrix C such that $C^3 = D$ e.g.

$$C = \begin{pmatrix} 4^{\frac{1}{3}} & 0 \\ 0 & (-1) \end{pmatrix}$$

Then $B = PCP^{-1}$ works. Indeed $B^3 = PC^3P^{-1} = PDP^{-1} = A$.

(3) Fibonacci Sequence

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 35, 55, \ldots$$

Formally defined as $f_{n+1} = f_n + f_{n-1} \quad f_0 = 0, \; f_1 = 1$.

Let

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

Notice that

$$\begin{pmatrix} f_{n+1} \\ f_n \end{pmatrix} = A\begin{pmatrix} f_n \\ f_{n-1} \end{pmatrix}$$

Since $f_{n+1} = f_n + f_{n-1}$...

$$\begin{pmatrix} f_{n+1} \\ f_n \end{pmatrix} = A \begin{pmatrix} f_n \\ f_{n-1} \end{pmatrix}$$

$$= A^2 \begin{pmatrix} f_{n-1} \\ f_{n-2} \end{pmatrix}$$

$$= \dots$$

$$= A^n \begin{pmatrix} f_1 \\ f_0 \end{pmatrix}$$

$$= A^n \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

We now know how to find A^n, and this gives a formula for f_n.

(4) Is

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

diagonalisable?

The characteristic polynomial is $(t-1)^2(t-2)$, so $\lambda_1 = 2$, $\lambda_2 = 1$. We find

$$v_1 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$$

is an eigenvector with eigenvalue λ_1.

We can find two non-proportional eigenvectors for $\lambda_1 = 1$. Then three eigenvectors are linearly independent. So A is diagonalisable. [So sometimes double root $\implies$ three eigenvectors and is still diagonalisable.]

"Nobody ever asks 'How's Waldo?'"

8 Conics and Quadrics

Back to 2-dim geometry, we work in $\mathbb{R}^2$.   Lecture 16

Lines are given by linear equations

$$ax_1 + bx_2 + c = 0$$

(polynomial equations in x_1, x_2 of degree 2)

> **Definition.** A conic is a curve in $\mathbb{R}^2$ that can be given by a quadratic equation, i.e. a polynomial equation in x_1, x_2 of degree 2.
>
> $$ax_1^2 + bx_1x_2 + cx_2^2 + dx_1 + cx_2 + f = 0$$

e.g. $x_1^2 + x^2 = r_2$ gives a circle. Also $(ax_1 + bx_2 + c)(a'x_1 + b'x_2 + c') = 0$ is the set of solutions is the union of two lines.
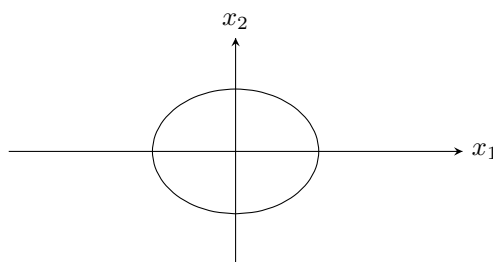
8.1 Standard Forms of Conics

Non-degenerate conics:

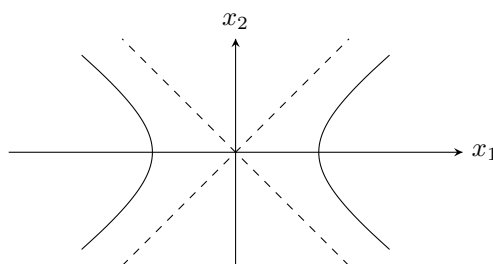For $a \neq 0, b \neq 0 \in \mathbb{R}$ we have

(1) An ellipse

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1$$

(2) A hyperbola

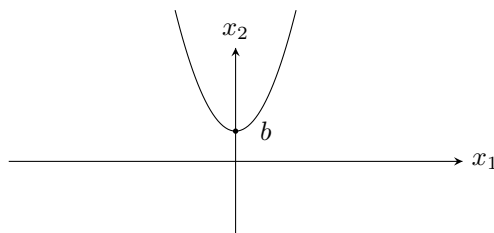$$\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 1$$

(3) An imaginary ellipse

$$-\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 1$$

This defines an empty set (there are no solutions).
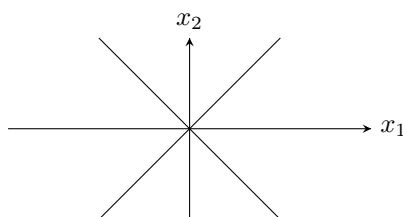
(4) A parabola

$$x_2 = ax_1^2 + b$$

Degenerate conics:

(5) Two lines with a common point

$$\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 0 \implies \left(\frac{x_1}{a} - \frac{x_2}{b}\right)\left(\frac{x_1}{a} + \frac{x_2}{b}\right) = 0$$
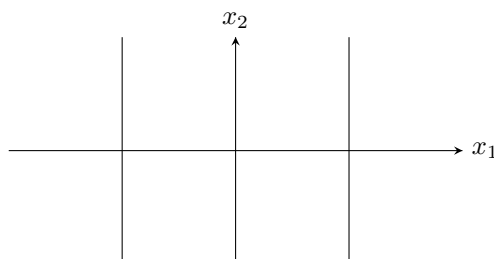

(6) $x_1^2 = 0$, a double line.

(7) A pair of two conjugate complex lines meeting in a real point

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 0 \implies \left(\frac{x_1}{a} - i\frac{x_2}{b}\right)\left(\frac{x_1}{a} + i\frac{x_2}{b}\right) = 0$$

The set of real solutions is one point $(0,0)$.

(8) Two parallel lines

$$\frac{x_1^2}{a^2} = 1 \quad (\frac{x_1}{a} = \pm 1)$$
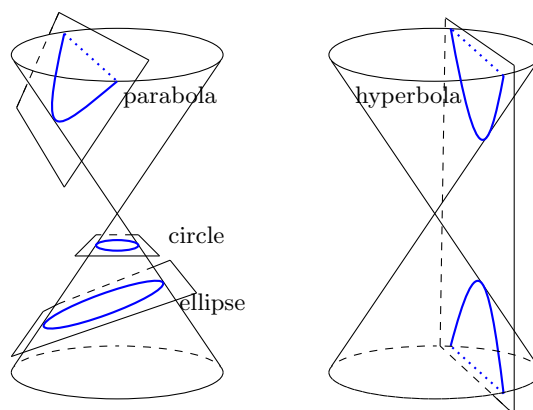
(9) A pair of parallel complex conjugate lines

$$\frac{x_1^2}{a^2} = 1$$

There are no real solutions. $\frac{x_1}{a} = \pm i$.

8.2 Reducing with Trigonometry

Our aim: To use translations and rotations in $\mathbb{R}^2$ to reduce any conic to one of these standard types.

Why conics? These are all slices of the cone $x^2 + y^2 = z^2$:

(i) $z = \text{constant} \implies$ a circle / ellipse.

(ii) $z = y + 1$ into equation $\implies x^2 = 2y + 1$, a parabola.

(iii) $y = 1$ (vertical plane) $\implies$ a hyperbola.

(iv) $z = 0$, $x^2 + y^2$: two conjugate lines with a common real point.

(v) $y = 0$, $x^2 = z^2$: two real lines.

Translations in $\mathbb{R}^2$:
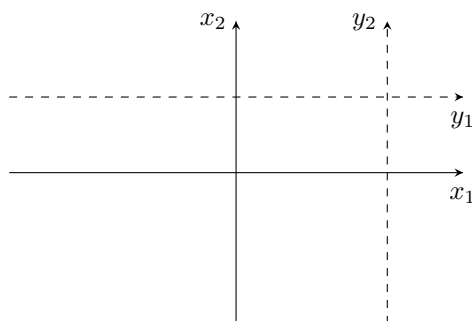
Moving the co-ordinates by (d_1, d_2) has the effect that

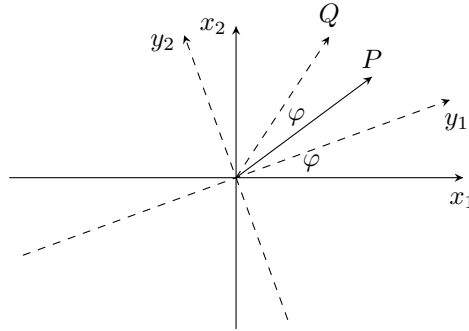$$x_1 = y_1 + d_1$$
$$x_2 = y_2 + d_2$$

Rotations in $\mathbb{R}^2$

Recall that rotating coordinate $P = (p_1, p_2)$ to $Q = (q_1, q_2)$ by φ through O is:

$$\begin{pmatrix} q_1 \\ q_2 \end{pmatrix} = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}$$

What is the effect of rotation on coordinates?

Rotating the (x_1, x_2) coordinate system by φ:

The y_1, y_2 coordinates of Q are simply p_1 and p_2. Therefore, the new coordinates y_1, y_2 and the old coordinates x_1, x_2 are related as follows:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

Let's write quadratic equations in matrix form. Recall that A^t denotes the transpose of A, i.e. the matrix whose $i - j$ entry is the $j - i$ entry of A.

> **Definition.** A square matrix is called *symmetric* if $A^t = A$.

e.g.

$$\begin{pmatrix} a & b \\ b & a \end{pmatrix}, \quad \begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix}$$

Recall that we proved that $(AB)^t = B^t \cdot A^t$ and $(A + B)^t = A^t + B^t$. Consider

$$A = \begin{pmatrix} a & \frac{1}{2}b \\ \frac{1}{2}b & c \end{pmatrix}$$

$$(x_1, x_2) \begin{pmatrix} a & \frac{1}{2}b \\ \frac{1}{2}b & c \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (x_1, x_2) \begin{pmatrix} ax_1 + \frac{1}{2}bx_2 \\ \frac{1}{2}bx_1 + cx_2 \end{pmatrix}$$

$$= x_1(ax_1 + \tfrac{1}{2}bx_2) + x_2(\tfrac{1}{2}bx_1 + cx_2)$$

$$= ax_1^2 + bx_1x_2 + cx_2^2$$

$$(x_1, x_2) \begin{pmatrix} d \\ e \end{pmatrix} = dx_1 + ex_2 = (d, e) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Therefore we can write $ax_1^2 + bx_1x_2 + cx_2^2 + dx_1 + ex_2 + f$ as

$$x^t Ax + (d, e)x + f$$

where $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $x^t = (x_1, x_2)$.

Theorem 8.1

Using a rotation and then translation we can always reduce any quadratic equation to standard form.

Low level proof - Idea: Get rid of the x_1x_2 terms using a rotation. If we can do that, then for some new coefficients a', c', d', e', f' we obtain:

$$a'y_1^2 + c'y_2^2 + d'y_1 + e'y_2 + f'$$

we complete the squares and reduce this equation to

$$a''z_1^2 + c''z_2^2 + f''$$

or (when $c' = 0$)

$$a''z_1^2 + e''z_2 + f''$$

Proof. Rotate through φ, then

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = R_\phi \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

$$\implies x^t Ax = (R\varphi y)^t A(R_\varphi y) = y^t R_\varphi^t AR_\varphi y$$

We want the matrix $R_\varphi^t AR_\varphi y$ to be diagonal (then $y_1 y_2$ terms disappears)

$$R_\varphi^t AR_\varphi = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} a & \frac{1}{2}b \\ \frac{1}{2}b & c \end{pmatrix} R_\varphi$$

$$= \begin{pmatrix} a\cos\varphi + \frac{1}{2}b\sin\varphi & \frac{1}{2}b\cos\varphi + c\sin\varphi \\ -a\sin\varphi + \frac{1}{2}b\cos\varphi & -\frac{1}{2}b\sin\varphi + c\cos\varphi \end{pmatrix} \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}$$

We need to know the non-diagonal terms. The 12 entry is

$$-a\sin\varphi\cos\varphi - \frac{1}{2}b\sin^2\varphi + \frac{1}{2}b\cos^2\varphi + c\sin\varphi\cos\varphi$$

$$= -\frac{1}{2}a\sin 2\varphi + \frac{1}{2}b\cos 2\varphi + \frac{1}{2}c\sin 2\varphi$$

$$= \frac{1}{2}(b\cos 2\varphi - (a - c)\sin 2\varphi)$$

If $a \neq c$, then take φ such that $\tan 2\varphi = \frac{b}{a-c}$ (then this term is 0).

If $a = c$, take $\varphi = \frac{\pi}{4}$ (again, 0).

After this rotation, $x = R_\varphi y$, we accomplished the first step. Then we obtain the equation

$$a' y_1^2 + c' y_2^2 + d' y_1 + e' y_2 + f$$

If $a' \neq 0$, $c' \neq 0$, we can complete the squares and reduce this to

$$a' z_1^2 + c' z_2^2 + f''$$

Then

$$y_1 = z_1 + \alpha_1$$
$$y_2 = z_2 + \alpha_2$$

This is exactly a translation by (α_1, α_2). If $a' \neq 0$, but $c' = 0$, we complete the square for y_1 and do nothing for y_2. Then reduce to $a' z_1^2 + e'' z_2 + f''$. We get a parabola or a degenerate conic.   ∎

Example 8.2. Reduce to standard form

$$5x_1^2 + 4x_1 x_2 + 2x_2^2 + x_1 + x_2 - 1 = 0$$

Step 1:

$$A = \begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix}$$

Use $R_\varphi = \tan 2\varphi = \frac{b}{a-c} = \frac{4}{3}$. We can calculate $\sin \varphi = \frac{2}{\sqrt{5}}$, $\cos \varphi = \frac{1}{\sqrt{2}}$.

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

Substitute into equations and obtain

$$5x_1^2 + 4x_1 x_2 + 2x_2^2 = y_1^2 + 6y_2^2$$

The new equation is

$$y_1^2 + 6y_2^2 - \frac{1}{\sqrt{5}} y_1 + \frac{3}{\sqrt{5}} y_2 - 1 = 0$$
$$= (y_1 - \frac{1}{2\sqrt{5}})^2 + 6(y_2 + \frac{1}{4\sqrt{5}})^2 - \frac{9}{8}$$
$$= z_1^2 + 6z_2^2 - \frac{9}{8}$$

so this is an ellipse.

8.3 Reducing with Eigenvectors

There is a trigonometry-free method!

Note for any rotation matrix, we have $R_\varphi^t = R_\varphi^{-1}$. Indeed

$$R_\varphi = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}$$

$$R_\varphi^t = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}$$

and

$$\begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

> **Definition.** Let P be a square matrix of the size $n \times n$. Then P is called *orthogonal* if $P^t \cdot P = I$. (In otherwords, P^{-1} exists and equals P^t)

Since the rows of P^t are precisely the columns of P, $P^t P = I$ is equivalent to the following property:

$$(c_i(P) \cdot c_j(P)) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

In sheet 6, there is an exercise that says a 2×2 matrix is orthogonal $\iff P$ is a rotation matrix or a reflection matrix; depending on whether $\det(P) = 1$ or $\det(P) = -1$. Note $R_\varphi^t A R_\varphi = R_\varphi^{-1} A R_\varphi$. In the proof of Theorem 8.1, we watned $R_\varphi^t A R_\varphi = R_\varphi^{-1} A R_\varphi$ to be diagonal.

New idea: Use diagonalisation, that

$$P = \begin{pmatrix} | & | \\ v_1 & v_2 \\ | & | \end{pmatrix}$$

where v_1, v_2 are the eigenvectors of A.

Why do v_1, v_2 exist? Why is P a rotation matrix?

Theorem 8.3

Let A be any real symmetric 2×2 matrix

$$A = \begin{pmatrix} a & \frac{1}{2}b \\ \frac{1}{2}b & c \end{pmatrix} \quad \text{such that } A \neq \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix} \text{ for some } \alpha \in \mathbb{R}.$$

Then A has the following properties:

(i) A has two different real eigenvalues λ_1, λ_2 $\lambda_1 \neq \lambda_2$

(ii) If v_1 is an eigenvector, with eigenvalue λ_1 and v_2 is an eigenvector with eigenvalue λ_2. Then $v_1 \cdot v_2 = 0$ (i.e. v_1, v_2 are perpendicular)

(iii) Up to swapping v_1 and v_2 and multiplying v_1, v_2 by a scalar, so

---

> that $||v_1|| = ||v_2|| = 1$, the matrix
>
> $$P = \begin{pmatrix} | & | \\ v_1 & v_2 \\ | & | \end{pmatrix}$$
>
> is a rotation matrix.

Proof. (i)

$$\det \begin{pmatrix} t-a & -\frac{1}{2}b \\ -\frac{1}{2}b & t-c \end{pmatrix} = t^2 - (a+c)t + (ac - \frac{b^2}{4})$$

$$\lambda_1, \lambda_2 = \frac{1}{2}(a+c \pm \sqrt{(a+c)^2 - 4ac - b^2})$$

$$= \frac{1}{2}(a+c \pm \sqrt{(a-c)^2 + b^2})$$

Since $A \neq \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix}$ for some α, we have either $b \neq 0$ or $a - c \neq 0$. Hence $(a-c)^2 + b^2 > 0$. Therefore the characteristic polynomial has two different real roots.

(ii) Let v_1, v_2 be eigenvectors for λ_1, λ_2 respectively. We need to show that $v_1 \cdot v_2 = 0$.

Consider $v_1^t A v_2$:

$$v_1^t A v_2 = v_1^t (A v_2)$$
$$= v_1^t \cdot \lambda_2 v_2$$
$$= \lambda_2 (v_1^t \cdot v_2)$$

But also

$$v_1^t A v_2 = (v_1^t A) v_2$$
$$= (v_1^t A^t) v_2$$
$$= (A v_1)^t v_2$$
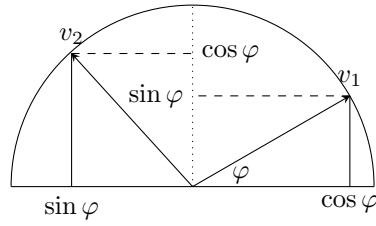$$= (\lambda_1 v_1)^t v_2$$
$$= \lambda_1 v_1^t v_2$$

Therefore $\lambda_2 (v_1^t v_2) = \lambda_1 (v_1^t v_2)$. So $(\lambda_1 - \lambda_2) \cdot (v_1^t v_2) = 0$, but from (i), $\lambda_1 \neq \lambda_2$, so $v_1^t v_2 = 0$.

(iii) Multiply v_1 and v_2 by non-zero numbers so that $||v_1|| = ||v_2|| = 1$. Clearly if

$$v_1 = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \text{ such that } \alpha^2 + \beta^2 = 1$$

then $\alpha = \cos \varphi$, $\beta = \cos \varphi$ for some φ. Since v_2 is also a unit vector and is perpendicular to v_1 if it's either

$$\begin{pmatrix} -\sin \varphi \\ \cos \varphi \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \sin \varphi \\ -\cos \varphi \end{pmatrix}$$

Choose $v_2 = \begin{pmatrix} -\sin\varphi \\ \cos\varphi \end{pmatrix}$. Then

$$\begin{pmatrix} | & | \\ v_1 & v_2 \\ | & | \end{pmatrix} = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} = R_\varphi \qquad \blacksquare$$

Observe that $R_\varphi^t = R_\varphi^{-1}$ (any rotation matrix is an orthogonal matrix). From the diagonalisation theory we know that $PAP^{-1} = \operatorname{diag}(\lambda_1, \lambda_2)$ where

$$P = \begin{pmatrix} | & | \\ v_1 & v_2 \\ | & | \end{pmatrix}$$

Now $P = R_\varphi$, so we have

$$R_\varphi^{-1} A R_\varphi = \operatorname{diag}(\lambda_1, \lambda_2)$$
$$\|$$
$$R_\varphi^t A R_\varphi$$

So if $x = R_\varphi y$, then

$$x^t A x = y^t R_\varphi^t A R_\varphi y$$
$$= y^t R_\varphi^{-1} A R_\varphi y$$

$$\implies x^t A x = y^t \operatorname{diag}(\lambda_1, \lambda_2) y = \lambda_1 y_1^2 + \lambda_2 y_2^2$$

$$(y_1, y_2) \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \lambda_1 y_1^2 + \lambda_2 y_2^2$$

Example 8.4. Same equation as before

$$A = \begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix}$$

$$t^2 - 7t + 6 = (t - 1)(t - 6)$$

so $\lambda_1 = 1$, $\lambda_2 = 6$. Let's find the first eigenvector (2nd follows easily since it's perpendicular).

$$I - A = \begin{pmatrix} -4 & -2 \\ -2 & -1 \end{pmatrix} \implies v_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ -2 \end{pmatrix}$$

($||v_1|| = 1$). Then we don't even have to calculate v_2, and we can take

$$v_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

Check:

$$\frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}$$

has determinant 1. This is R_φ from the beginning of the lecture.

Corollary 8.5. *Any conic can be reduced to the following form:*

$$\lambda_1 y_1^2 + \lambda_2 y_2^2 + d' y_1 + e' y_2 + f'$$

be a rotation. Here λ_1, λ_2 are the eigenvalues of $A = \begin{pmatrix} a & \frac{1}{2}b \\ \frac{1}{2}b & c \end{pmatrix}$.

Obviously $\lambda_1 \lambda_2 = \det(A)$. (since the characteristic polynomial is $t^2 - (a + c)t + \det(A) = (t - \lambda_1)(t - \lambda_2)$).

Therefore if $\det(A) > 0$, then λ_1, λ_2 have the same sign. We complete the squares and obtain

$$\lambda_1 z_1 2 + \lambda_2 z_2^2 = h, \quad \text{where } h \in \mathbb{R}$$

- If $h > 0$, this is an ellipse.

- If $h < 0$, this is an imaginary ellipse.

- If $h = 0$, this is two complex conjugate lines meeting in one real point $(0,0)$

Next, if $\det(A) < 0$, completing the square we obtain

$$\lambda_1 z_1^2 + \lambda_2 z_2^2 = h, \quad h \in \mathbb{R}$$

- If $h > 0$ or $h < 0$, this is a hyperbola.

- If $h = 0$, this a pair of real lines meeting at a point.

Finally if $\det(A) = 0$, then $\lambda_2 = 0$, $\lambda_1 \neq 0$. Complete the square for y_1, so we obtain

$$\lambda_1 z_1^2 + e' y_2 + f'$$

- If $e' \neq 0$, we reduce to $\lambda_1 z_1^2 + e' z_2 = 0$, which is a parabola.

- If $e' = 0$, then $\lambda_1 z_1^2 + f' = 0$. $z_1^2 = -\frac{f'}{\lambda_1}$. If RHS is positive, these are two parallel real lines. If RHS is zero, it's a double line. If RHS is negative, there are no real solutions; this is two parallel complex conjugate lines.

Moral: If we can prove that our conic is non-empty and non-degenerate, then its type is uniquely determined by $\det(A)$.

- $\det(A) > 0 \implies$ ellipse.

- $\det(A) > 0 \implies$ parabola.

- $\det(A) > 0 \implies$ hyperbola.

8.4 Quadric Surfaces

> **Definition.** A *quadric surface* is a surface in $\mathbb{R}^3$ defined by an eqn of degree 2:
>
> $$ax_1^2 + bx_2^2 + cx_3^2 + dx_1x_2 + ex_1x_2 + fx_2x_3 + gx_1 + hx_2 + jx_3 + k = 0$$

Lecture 19
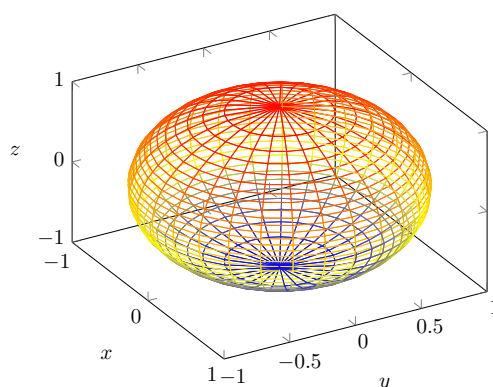
Task: Use an orthogonal transformation, ideally a rotation, to reduce this equation to one of standard form.

Non-denerate forms

Here are non-degenerate forms, with $0 \neq a, b, c \in \mathbb{R}$:

(1) Ellipsoid:

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} + \frac{x_3^2}{c^2} = 1$$

(2) Hyperboloid of one sheet:
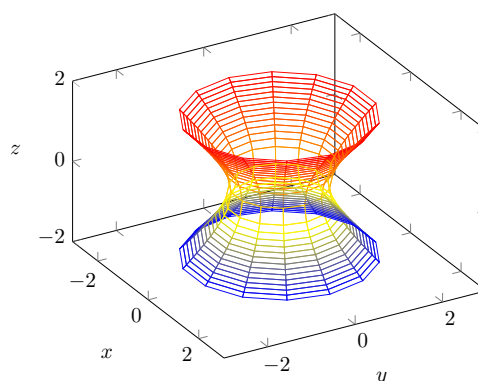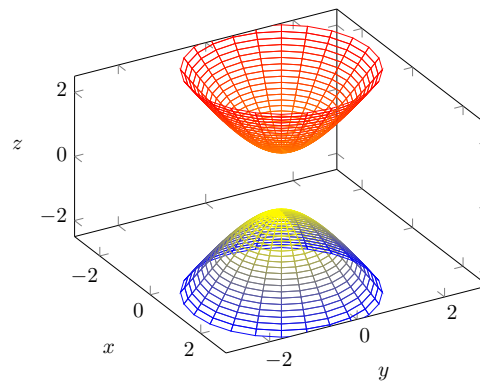
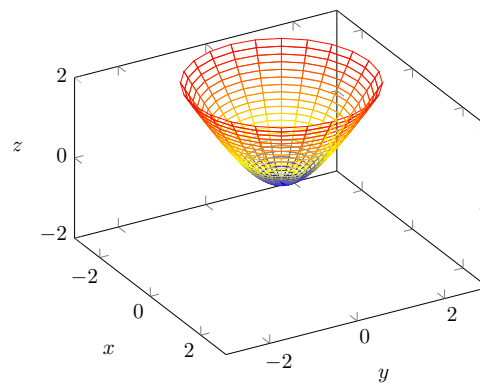$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} - \frac{x_3^2}{c^2} = 1$$

(3) Hyperboloid of two sheets:

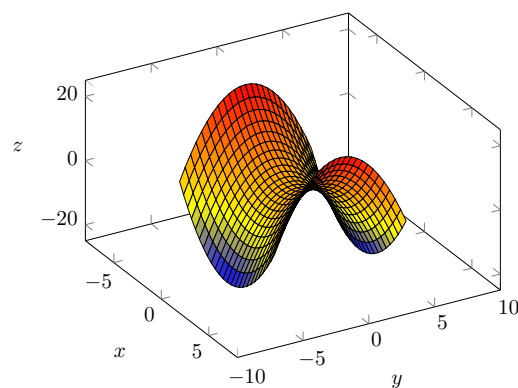$$-\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} + \frac{x_3^2}{c^2} = 1$$

(4) Elliptic Paraboloid:

$$x_3 = \frac{x_1^2}{a^2} + \frac{x_2^2}{b^2}$$
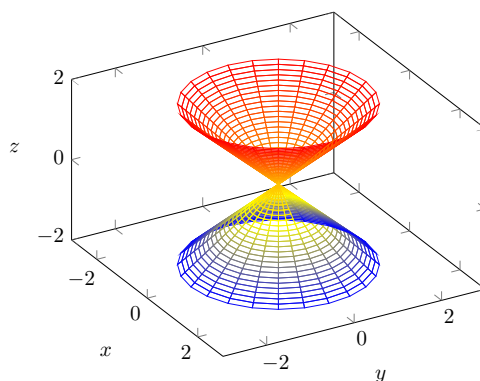

(4) Hyperbolic Paraboloid:

$$x_3 = \frac{x_1^2}{a^2} - \frac{x_2^2}{b^2}$$

Some examples of degenerate quadrics

(5) Cone:

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} - \frac{x_3^2}{c^2} = 0$$

(6) Elliptic cylinder

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1$$

Similarly there are hyperbolic and parabolic cylinders, etc. *You will not be expected to know all the quadrics for the exam.*

8.5 Reducing Quadrics

STEP 1. Write the equation in matrix form.

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \implies x^t A x + (ghj)x + k = 0$$

where

$$A = \begin{pmatrix} a & \frac{d}{2} & \frac{e}{2} \\ \frac{d}{2} & b & \frac{f}{2} \\ \frac{e}{2} & \frac{f}{2} & c \end{pmatrix}$$

STEP 2. Find the eigenvalues of A. In fact any symmetric 3×3 matrix has three real eigenvalues: $\det(tI - A) = (t - \lambda_1)(t - \lambda_2)(t - \lambda_3)$ where $\lambda_i \in \mathbb{R}$. (See sheet 8 for proof).

We've seen before that $\lambda_i \neq \lambda_j$ and v_i is an eigenvector of λ_i, v_j is an eigenvector for λ_j, then v_i is perpendicular to v_j, i.e. $v_i \cdot v_j = 0$. So if $\lambda_1 \neq \lambda_2 \neq \lambda_3$, then we can find v_1, v_2, v_3 of length 1: $||v_1|| = ||v_2|| = ||v_3|| = 1$. These vectors will be perpendicular automatically to each other. The same also works in the general case.

STEP 3.

$$P = \begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix}$$

Then P is an orthogonal matrix, i.e. $P^t P = I \implies P^t = P^{-1}$, since $||v_i|| = 1, i = 1, 2, 3$ and $v_i \cdot v_j = 0$ when $i \neq j$.

STEP 4. Since $P^{-1} = P^t$, $x = Py$ where $y = (y_1, y_2, y_3)^t$. So the equation now looks as follows:
$$x^t A x = y^t P^{-1} A P y$$

But $P^{-1}AP$ is the diagonal matrix

$$D = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$$

Therefore

$$\begin{aligned} x^t A x &= y^t P^{-1} A P y \\ &= y^t \text{diag}(\lambda_1, \lambda_2, \lambda_3) y \\ &= \lambda_1 y_1^2 + \lambda_2 y_2^2 + \lambda_3 y_3^2 \end{aligned}$$

The whole equation now is

$$\lambda_1 y_1^2 + \lambda_2 y_2^2 + \lambda_3 y_3^2 + g' y_1 + h' y_2 + j' y_3 + k = 0$$

If $\lambda_1 \lambda_2 \lambda_3 \neq 0$, complete the squares and get one of two hyperboloids, an ellipsoid, or a degenerate quadric. If $\lambda_1 \lambda_2 \neq 0, \lambda_3 = 0$ we get a paraboloid (one of two types) or a degenerate.

Example 8.6. Reduce to standard form

$$2x_1 x_2 + 2x_1 x_3 - x_2 - 1 = 0$$

Using an orthogonal transformation and a translation.

1. $x^t A x - x_2 - 1 = 0$, where

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

2.

$$\det \begin{pmatrix} t & -1 & -1 \\ -1 & t & 0 \\ -1 & 0 & t \end{pmatrix} = t^3 - 2t$$
$$= t(t - \sqrt{2})(t + \sqrt{2})$$

If this quadric is non-degenerate, it myst be a hyperbolic paraboloid (since $\lambda_1 = 0, \lambda_2 = \sqrt{2}$ and $\lambda_3 = -\sqrt{3} < 0$)

3. Find the eigenvectors $\lambda_1 = 0$, solving $Ax = 0$:

$$\left(\begin{array}{ccc|c} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{array}\right)$$

Solving we find $v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$.

For $\lambda_2 = \sqrt{2}$, solving $Ax = 0$:

$$\left(\begin{array}{ccc|c} \sqrt{2} & -1 & -1 & 0 \\ -1 & \sqrt{2} & 0 & 0 \\ -1 & 0 & \sqrt{2} & 0 \end{array}\right)$$

Solving we find $v_2 = \frac{1}{2} \begin{pmatrix} \sqrt{2} \\ 1 \\ 1 \end{pmatrix}$.

$\lambda_3 = -\sqrt{3} \implies v_3 = \frac{1}{2} \begin{pmatrix} -\sqrt{2} \\ 1 \\ 1 \end{pmatrix}$.

4. Hence

$$P = \begin{pmatrix} 0 & \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ +\frac{1}{\sqrt{2}} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{\sqrt{2}} & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

$$\implies y^t A y + (0, -1, 0) P y - 1$$
$$= \sqrt{2}(y_2 - \alpha)^2 - \sqrt{2}(y_3 - \beta)^2 + \frac{1}{\sqrt{2}}(y_1 - \delta)$$

where $\alpha, \beta, \delta \in \mathbb{R}$. So this is a non-degenerate quadric, hence a hyperbolic paraboloid.
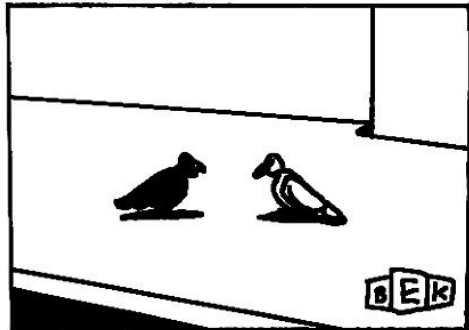
"I can't believe you symbolize peace when you're such a bitch."

9 Geometry in $\mathbb{R}^3$

A vector $x = (x_1, x_2, x_3) \in \mathbb{R}^3$ is such that

- a point P

- a vector $\vec{OP}$

- any vector $\vec{AB}$ of same length and direction as $\vec{OP}$.

Definition. The *length* of $x = (x_1, x_2, x_3)$ is

$$\|x\| = \sqrt{x_1^2 + x_2^2 + x_3^2}$$
$$= |OP|$$

$$(x_1, x_2, x_3) + (y_1, y_2, y_3) = (x_1 + y + 1, x_2 + y_2, x_3 + y_3)$$
$$\lambda(x_1, x_2, x_3) = (\lambda x_1, \lambda x_2, \lambda x_3)$$

Definition (Lines of $\mathbb{R}^3$)**.** $L = \{a + \lambda u : \lambda \in \mathbb{R}\}$ is a line in $\mathbb{R}^3$.

Definition (Dot product)**.** For $x, y \in \mathbb{R}^3$

$$x \cdot y = x_1 y_1 + x_2 y_2 + x_3 y_3$$

Proposition 9.1. *For $x, y \in \mathbb{R}^3$ we have $x \cdot y = \|x\| \cdot \|y\| \cdot \cos \theta$*

9.1 Planes

$n = a$ is a normal vector for the plane.

$x \cdot y = 0 \iff x$ and y are perpendicular. So $x - a$ is perpendicular to $n \iff (x - a) \cdot n = 0$. This is the equation of a plane.

Now $u \cdot (v + w) = u \cdot v + u \cdot w$ so $(x - a) \cdot n = x \cdot n - a \cdot n$. So the equation of the plane is

$$x \cdot n = a \cdot n$$

Write $a \cdot n = c$, a scalar, so you get

$$x \cdot n = c$$

Write $n = (n_1, n_2, n_3)$ then

$$n_1 x_1 + n_2 x_2 + n_3 x_3 = c$$

Example 9.2. We need a, n. $a = (e, e, 0)$ and $n = (0, 0, \pi)$. Then $(x - a) \cdot n = 0$, so $x \cdot n = a \cdot n$. Then

$$a \cdot n = (e, e, 0) \cdot (0, 0, n)$$
$$= 0$$

So the equation is

$$(x_1, x_2, x_3) \cdot (0, 0, n) = 0$$
$$\pi x_3 = 0$$
$$\implies x_3 = 0$$

Proposition 9.3.

(i) The intersection of two planes is

- *Nothing*

- *A line*

- *A plane (if the two planes are the same)*

(ii) If A, B, C are 3 points which don't lie on a common line, then there is a unique plane through A, B, C.

Proof. Problem sheet 7.

Let $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$ and $e_3 = (0, 0, 01)$, so $(x_1, x_2, x_3) = x_1 e_1 + x_2 e_2 + x_3 e_3$.

Definition (Vector product)**.** The vector product = cross product of a, b is

$$a \times b = \begin{vmatrix} e_1 & e_2 & e_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$$
$$= e_1(a_2 b_3 - a_3 b_2) - e_2(a_1 b_3 - a_3 b_1) + e_3(a_1 b_2 - a_2 b_1)$$
$$= (a_2 b_3 - a_3 b_2, a_3 b_1 - a_1 b_3, a_1 b_2 - a_2 b_1)$$

Example 9.4.

$$(1, -1, 0) \times (3, 0, 1) = \begin{vmatrix} e_1 & e_2 & e_3 \\ 1 & -1 & 0 \\ 3 & 0 & 1 \end{vmatrix}$$
$$= (-1, 1, 3)$$

Proposition 9.5. $a \times b$ *is perpendicular to both a, b.*

Proof.

$$a \cdot (a \times b) = \begin{vmatrix} a_1 & a_2 & a_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$$
$$= 0$$

Similarly $b \cdot (a \times b) = 0$.   ∎

Proposition 9.6.   *(i)* $e_1 \times e_2 = e_3$, $e_2 \times e_3 = e_1$, $e_3 \times e_1 = e_2$

  (ii) $a \times a = 0$

  (iii) $b \times a = -a \times b$

  (iv) $(\lambda a) \times b = \lambda(a \times b) = a \times (\lambda b)$ *for any scalar* $\lambda \in \mathbb{R}$

  (v) $a \times (b + c) = a \times b + a \times c$

Proof. Use properties of determinants:
(i)

$$e_1 \times e_2 = \begin{vmatrix} e_1 & e_2 & e_3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix}$$
$$= e_3 = (0, 0, 1)$$

(ii) $a = (a_1, a_2, a_3)$, then

$$a \times a = \begin{vmatrix} e_1 & e_2 & e_3 \\ a_1 & a_2 & a_3 \\ a_1 & a_2 & a_3 \end{vmatrix}$$
$$= 0 = (0, 0, 0)$$

(iii)

                           ∎

Proposition 9.7. *For any vectors* $a, b \in \mathbb{R}^3$ *we have*

$$||a \times b|| = ||a|| \cdot ||b|| \cdot |\sin \theta|$$
$$= \ \textit{the area of a parallelogram } OABC$$

Proof. Squaring both sides:

$$\implies ||a \times b||^2 = (a_2 b_3 - a_3 b_2)^2 + (a_3 b_1 - b_1 a_3)^2 + (a_1 b_2 - a_2 b_1)^2$$

$$\text{Area} = ||a|| \cdot ||h|| = ||a|| \cdot ||b|| \cdot |\sin \theta|$$

$$||a||^2 \cdot ||b||^2 \cdot (1 - \cos^2 \theta) = ||a||^2 ||b||^2 - (||a|| ||b|| \cos \theta)^2$$
$$= (a_1^2 + a_2^2 + a_3^2) \cdot (b_1^2 + b_2^2 + b_3^2) - (a \cdot b)^2$$
$$= (a_1^2 + a_2^2 + a_3^2) \cdot (b_1^2 + b_2^2 + b_3^2) - (a_1 b_1 + a_2 b_2 + a_3 b_3)^2$$
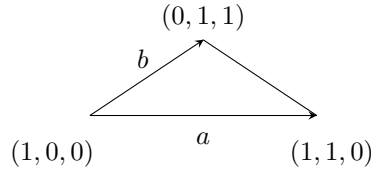$$= a_1^2 b_2^2 + a_2^2 + b_1^2 - 2a_1 a_2 b_1 b_2 + \cdots + (\text{similar terms})$$

Check that the two expressions are identical.   ∎

Remark. $a \times b$ is a vector perpendicular to the plane through O, a and b of length that's equal to the area of the parallelogram build on O, a, b.

Example 9.8. Find the area of a triangle in $\mathbb{R}^3$ with corners $(1,0,0)$, $(1,1,0)$, $(0,1,1)$.

Answer: Area $= \frac{1}{2}$(length of $a \times b$) where $a = (0,1,0)$, $b = (-1,1,1)$

$$
\det \begin{pmatrix} e_1 & e_2 & e_3 \\ 0 & 1 & 0 \\ -1 & 1 & 1 \end{pmatrix} = e_1 - e_2 \cdot 0 + e_3
$$

$$
= e_1 + e_3
$$

$$
= (1,0,1) = a \times b
$$

Hence $||a \times b|| = \sqrt{2} \implies$ area of triangle is $\frac{\sqrt{2}}{2}$.

Volume

Proposition 9.9. *The volume of the parallelepiped built on a, b, c equals $|a \cdot (b \times c)|$*

$$
= \left| \det \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix} \right|
$$

Proof. Let's show that vol$= |a \cdot (b \times c)|$. By previous proposition the area of the base is $(b \times c) = ||b \times c|| \cdot n$ where n is a unit normal to the plane through $0, b$ and c. Therefore

$$
|a \cdot (b \times c)| = ||a|| \cdot ||b \times c|| \cdot |\cos \theta|
$$

where θ is the angle between a and n. But $||a|| \cdot |\cos \theta|$ is the height of the parallelepiped. So we proved that vol$= |a \cdot (b \times c)|$.

Finally

$$
\det \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}
$$

∎

Properties of scalar triple product

Proposition 9.10. *For any $a, b, c \in \mathbb{R}^3$ we have*

(i) $a \cdot (b \times c) = c \cdot (a \times b) = b \cdot (c \times a)$

(ii) a, b, c *are coplanar* $(= 0, a, b, c$ *belong to a plane) if and only if*

$$\det \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix} = 0$$

Proof. (i)

$$\det \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix} = -\det \begin{pmatrix} a_1 & a_2 & a_3 \\ c_1 & c_2 & c_3 \\ b_1 & b_2 & b_3 \end{pmatrix} = \det \begin{pmatrix} c_1 & c_2 & c_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix}$$

Hence $a \cdot (b \times c) = c \cdot (a \times b)$. Similarly we can probe that $c \cdot (a \times b) = b \cdot (c \times a)$.

(ii) a, b, c are coplanar iff volume of parallelepiped built on a, b, c is $0 \iff$ $\det = 0$. ∎

Alternative Proof of (ii). Any plane through 0 can be given by

$$m_1 x_1 + m_2 x_2 + m_3 x_3 = 0$$

(at least one co-efficient $m_i \neq 0$). Then

$$\begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix} \begin{pmatrix} m_1 \\ m_2 \\ m_3 \end{pmatrix} = 0$$

Since a, b, c are contained in this plane. This system of linear equations has a non-zero solution. Therefore $\det = 0$.

This argument can be reversed; conversely if $\det = 0$, then there is a non-zero solution which gives a plane though $0, a, b, c$. ∎

Example 9.11. *Are the points* $(1, 0, 0), (1, 1, 1), (1, 2, 3), (2, 3, 4)$ *in the same plane?*

Let $a = (0, 1, 1), b = (0, 2, 3), c = (1, 3, 4)$.

$$\det \begin{pmatrix} 0 & 1 & 1 \\ 0 & 2 & 3 \\ 1 & 3 & 4 \end{pmatrix} = \det \begin{pmatrix} 1 & 1 \\ 2 & 3 \end{pmatrix} = 1$$

This is not 0, hence the four points do not belong to a common point.

What is $(a \times b) \times c$*?*

We know that this is perpendicular to the plane through $a \times b$ and c. Hence $(a \times b) \times c$ is in the plane through $0, a, b$, so the triple vector product $(a \times b) \times c = (k_1)a + (k_2)b$.

Proposition 9.12.   *For $a, b, c \in \mathbb{R}^3$:*

(i) $(a \times b) \times c) = (a \cdot c)b - (b \cdot c)a$

(ii) $a \times (b \times c) = (a \cdot c)b - (a \cdot b)c$

Remark. Vector product is not associative!

Proof. (i) $a \times b = (a_2 b_3 - a_3 b_2, a_3 b_1 - a_1 b_3, a_1 b_2 - a_2 b_1)$. $(a \times b) \times c =$ the x_1 coordinate of $(a \times b) \times c$ which is $(a_3 b_1 - a_1 b_3)c_3 - (a_1 b_2 - a_2 b_1)c_2$. The x_1 coordinate of RHS is $(a_1 c_1 + a_2 c_2 + a_3 c_3)b_1 - (b_1)1 + b_2 c_3 + b_3 c_3)a_1$. Hence LHS = RHS. Similarly, check that x_2 and x_3 coordinate of LHS and RHS are equivalent.

(ii)

$$
\begin{aligned}
a \times (b \times c) &= -(b \times c) \times a) \\
\implies &= (c \times b) \times a) \\
\implies &= (c \cdot a)b - (b \cdot a)c \\
&= (a \cdot b)b - (a \cdot b)c \qquad\blacksquare
\end{aligned}
$$

Remark. Use this proposition to establish the Jacobi identity:

$$(a \times b) \times c + (c \times a) \times b + (b \times c) \times a = 0$$

This, with the property that $a \times b = -b \times a$ turn $\mathbb{R}^3$ into a *Lie Algebra*.

> **Example 9.13.** $a, b, c, d \in \mathbb{R}^3$. What is $(a \times b) \cdot (c \times d)$?
>
> Answer: Call $u = a \times b$. We proved that
>
> $$
> \begin{aligned}
> u \cdot (c \times d) &= d \cdot (u \times c) \quad \text{(prop 9.7)} \\
> &= d \cdot ((a \times b) \times c) \\
> &= d \cdot ((a \cdot c)b - (b \cdot c)a) \quad \text{(prop 9.8)} \\
> &= (a \cdot c) \cdot (b \cdot d) - (b \cdot c)(a \cdot d)
> \end{aligned}
> $$

Proposition 9.14. *Let $a \in \mathbb{R}^3$, $a \neq 0$. Then $b \times a = 0$ if and only if $b = \lambda a$ for some $\lambda \in \mathbb{R}$.*

Proof. If $b = \lambda a$ then $b \times a = \lambda a \times a = 0$. Conversely assume that $b \times a = 0$. Then certainly $(a \times (b \times a) = a \times 0 = 0$. Then $a \times (b \times a) = (a \cdot a)b - (a \cdot b)a$. We know that this is zero, but $(a \cdot a) = ||a||^2 \neq 0$, since $a \neq 0$, so $b = \frac{(a \cdot b)}{||a||^2}a$. $\blacksquare$

Proposition 9.15. *Let $a, b \in \mathbb{R}^3$. Assume $a \neq 0$. Then the set of $x \in \mathbb{R}^3$ such that $(x - b) \times a = 0$ is the line $\{b + \lambda a \mid \lambda \in \mathbb{R}\}$.*

Proof. $(x - b) \times a = 0 \iff x - b = \lambda a$ for some $\lambda \in \mathbb{R}$ by the previous proposition. $\blacksquare$

9.2 Rotations in $\mathbb{R}^3$

The axis of rotation is the line in $\mathbb{R}^3$ such that every point of this line is fixed by the rotation. We write a rotation by a 3×3 matrix P such that a vector $(x_1, x_2, x_3)^t$ goes to $P(x_1, x_2, x_3)^t$.

Claim:

(i) P is an orthogonal matrix, i.e. $P^t = P^{-1}$

(ii) Any orthogonal matrix has determinant 1 or -1. If $\det P = 1$, then this is a rotation matrix.

(iii) The axis is $\{\lambda v \mid \lambda \in \mathbb{R}\}$, where v is an eigenvector of P with eigenvalue 1, i.e. $Pv = v$

(iv) To find the angle, we take any unit vector w perpendicular to the axis, i.e. $w \cdot v = 0$. Then calculate $w \cdot Pw = ||w|| \cdot ||Pw|| \cos \theta = \cos \theta$.
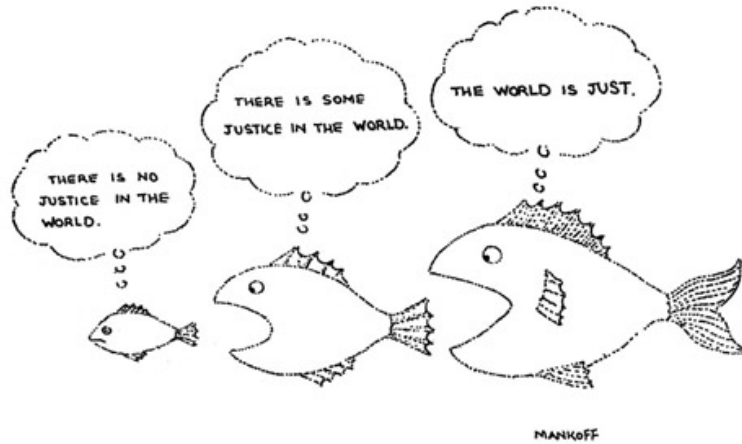
10 Vector spaces

We've seen $\mathbb{R}^2$ and $\mathbb{R}^3$. Similarly, we can work with $\mathbb{R}^n = \{(x_1, \ldots, x_n) \mid x_i \in \mathbb{R}\}$. In the same way we can do linear algebra in $\mathbb{C}^n = \{x_1, \ldots, x_n \mid x_i \in \mathbb{C}\}$. In this case "scalars" mean "complex numbers". The same methods work equally well for binary vectors.

Notation. $\mathbb{F}_2 = \{0, 1\}$ is the binary field such that:

$$0 + 0 = 1 + 1 = 0$$
$$1 + 0 = 0 + 1 = 1$$
$$0 \times 0 = 1 \times 0 = 0 \times 1 = 0$$
$$1 \times 1 = 1$$

So we have three notions of scalars: $\mathbb{R}$, $\mathbb{C}$, $\mathbb{F}_2$.

What is $(\mathbb{F}_2)^n$?

$(\mathbb{F}_2)^n = \{(x_1, \ldots, x_n) \mid x_i \in \mathbb{F}_2\}$

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$$

Here $+$ standard for binary addition, e.g. $(1, 0) + (1, 1) = (0, 1)$.

If $\lambda \in \mathbb{F}_2$, then $\lambda(x_1, \ldots, x_n) = (\lambda x_1, \ldots, \lambda x_n)$.

Now let F be one of $\mathbb{R}, \mathbb{C}$ or $\mathbb{F}_2$. I will refer to F as a *field*.

Rules satisfied by addition of vectors

For any $u, v, w \in V$ [V is in $\mathbb{R}, \mathbb{C}$ or $\mathbb{F}_2$], we have

  <u>A1</u> $u + (v + w) = (u + v) + w$

  <u>A2</u> $u + 0 = 0 + u = u$, $0 = (0, 0, \ldots, 0)$

  <u>A3</u> $u + (-u) = 0$, $u = (u_1, \ldots, u_n) - u = (-u_1, \ldots, -u_n)$

  <u>A4</u> $u + v = v + u$

Rules satisfied by scalar multiplication

For any $\lambda, \mu \in F$ and any $u, v \in V$ we have:

  <u>S1</u> $\lambda(u + v) = \lambda u + \lambda v$

  <u>S2</u> $(\lambda + \mu)u = \lambda u + \mu u$

  <u>S3</u> $\lambda(\mu u) = (\lambda \mu)u$

  <u>S4</u> $1u = u$

> **Definition** (Vector space V over F)**.** Let F be a field (i.e. $\mathbb{R}, \mathbb{C}$ or $\mathbb{F}_2$) and let V be a set, the elements of which are called *vectors*. Suppose we have addition in V i.e. a function $V \times V \to V$ and multiplication by a

scalar, i.e. a function $F \times V \to V$ and an elenet of V called *zero* such that the properties A1-A4 and S1-S4 all hold.

Examples 10.1.

(1) $\mathbb{R}^n, \mathbb{C}^n, (\mathbb{F}_2)^n$

(2) Let S be any set. Consider all functions $S \to |R$. Let's define the structure of a vector space on Functions$(S, \mathbb{R})$.

If $f_1 : S \to \mathbb{R}$ and $f_2 : S \to \mathbb{R}$ are functions, then $f_1 + f_2$ is a function such that for any element $s \in S$, we have $(f_1 + f_2)(s) = f_1(s) + f_2(s)$. The zero element of Functions$(S, \mathbb{R})$ is the zero function, i.e. the function that is equal to zero $\forall s \in S$. If $f \in$ Functions$(S, \mathbb{R})$ then $-f$ is the function such that $(-f)(s) = -f(s), \forall s \in S$.

It's clear that A1-4 hold. Now lets define scalar multiplication: $\lambda \in \mathbb{R}$, $f \in$ Functions$(s, \mathbb{R})$ then λf is the function such that $(\lambda f)(s) = \lambda \cdot f(s), \forall s$.

Check that S1-4 hold, e.g. let's check S4. $\lambda(\mu f)$ is the function such that $\forall s \in S$ we have $(\lambda(\mu f))(s) = \lambda \mu f(s)$, which is the same as the value of the function $(\lambda \mu) \cdot f$.

(3) Similarly the set of functions from S to F is a vector space over F.

(4) Let's be more specific. Let $S = \mathbb{R}$. Consider Functions$(\mathbb{R}, \mathbb{R})$, e.g.

So consider just polynomial functions. This is a subset of the set of all functions $\mathbb{R} \to \mathbb{R}$. Addition of polynomials is again a polynomial, and a scalar of a polynomial is also a polynomial. The set of polynomials is a *subspace* of the vector space of all functions $\mathbb{R} \to \mathbb{R}$.

Polynomials *inherit* addition and scalar multiplication from Functions$(\mathbb{R}, \mathbb{R})$, they are closed under these operations. Hence

$$\mathbb{R}[x] = \{\text{the set of all polynomials with real coeficients}\}$$

is a vector space over $\mathbb{R}$. We can also check A1-4 and S1-4 for polynomials directly.

(5) Now consider polynomials of degree up to d, where d is a fixed positive integer. This set is closed under addition and scalar multiplication so it's also a vector space over $\mathbb{R}$.

e.g. $d = 0$ is the space of polynomials of degree $0 = \mathbb{R}$. $d = 1$ is $ax + b$. So addition is $(ax + b) + (a'x + b) = (a + a')x + (b + b')$ and multiplication is $\lambda(ax + b) = \lambda ax + \lambda b$. So the space of linear polynomials can be identified with $\mathbb{R}^2$. Similarly, the space of quadratic polynomials $= \{ax^2 + bx + c \mid a, b, c \in \mathbb{R}\}$ can be identified with $\mathbb{R}^3$ be associating to $ax^2 + bx + c$ the vector (a, b, c).

Notation. $V \times V = \{(v, v') \mid v, v' \in V\}$. $V \times V \to V$ means a function from pairs of elements of V goes to V. For example addition of vectors is such an operation. $F \times V = \{\lambda, v) \mid \lambda \in F, v \in V\}$. So scalar multiplication is a function $F \times V \to V$.

Let's deduce some consequences of the axioms A1-4, S1-4.

Proposition 10.2. *Let F be a field (i.e. $F = \mathbb{R}, \mathbb{C}$ or $\mathbb{F}_2$). Let V be a vector space over F. For any $v \in V$ and any $\lambda \in F$ we have:*

(i) $0v = \mathbf{0}$

(ii) $\lambda \mathbf{0} = \mathbf{0}$

(iii) If $\lambda \in F$, $\lambda \neq 0$, then $\lambda v = 0$ only if $v = 0$.

(iv) $(-\lambda)v = \lambda(-v) = -(\lambda v)$

(v) $(-1)v = -v$.

Proof.
(i) In F we have $0 = 0 + 0$. Therefore $0v = (0+0)v = 0v + 0v$. Let's add $-Ov$ to both sides $\implies 0v - 0v = 0v + 0v = 0v$. By A1 $\implies 0v - 0v = 0v + (0v - 0v)$ and by A3 $\implies \mathbf{0} = 0v$. $(*)$

(ii) We need to prove $\lambda 0 = 0$. By A2, $0 = 0 + 0$. Then $\lambda 0 = \lambda(0 + 0)$. By S1 $\lambda 0 = \lambda 0 + \lambda 0$. Add $(-\lambda 0)$ to both sides $\implies \lambda 0 - \lambda 0 = \lambda 0 + \lambda 0 + \lambda 0 - \lambda 0$. By $(*) \implies \mathbf{0} = \lambda 0$.

(iii) $\lambda^{-1} \in F \implies \lambda^{-1}(\lambda v) = \lambda^{-1} 0 = 0$. By S3 and S4 $(\lambda^{-1}\lambda)v = 1 \cdot v = v = 0$.

(iv), (v) exercise. ∎

10.1 Subspaces

> **Definition.** Let V be a vector space over F. Let W be a subset of V. Then W is called a *vector space* of V when W itself is a vector space over F with the same addition and scalar multiplication as in V.

Proposition 10.3. *Let W be a subset of a vector space V over F. THen W is a subspace if the following conditions hold:*

(i) W contains 0 (zero vector of V)

(ii) For any $w_1, w_2 \in W$, $w_1 + w_2 \in W$ (W is closed under vector addition)

(iii) For any $w \in W$ and $\lambda \in F$, $\lambda w \in W$ (W is closed under scalar multiplication)

Remark. (i) is a consequence of (ii) and (iii). So to check that $W \subset V$ is a subspace, it is enough to check (ii) and (iii). Indeed suppose W is closed

under addition and multiplication by scalars. Take any $w \in W$, $w \neq 0$. Then $(-1)w = -w$ and $w + (-w) = 0 \in W$ (because W is closed under addition.)

Proof. We must check that A1-A4 and S1-S4 hold in W.

$-w \in W$ because $-w = (-1)w$. W contains 0, for any $w \in W$, $-w \in W$, closed under addition and scalar multiplication. All axioms hold in W because they hold in V by assumption. ∎

Examples of subspaces:

  (i) Trivial subspaces: V is its own subspace. Also $\{0\}$ is a subspace (closed by Prop 10.1 and 10.2).

  (ii) Polynomials with real coefficients $\mathbb{R}[x]$ is a subspace of Functions$(\mathbb{R}, \mathbb{R})$.

  (iii) Polynomials of degree at most d is a subspace of Functions$(\mathbb{R}, \mathbb{R})$ and equally a subspace of $\mathbb{R}[x]$.

Non-examples

  (i) The upper half plane $\{(x_1, x_2) \mid x_1 \geq 0, x_1, x_2 \in \mathbb{R}\}$ is not a subspace of $\mathbb{R}^2$ as it fails to be closed with negative scalar multiplication.

  (ii) The co-ordinate cross, i.e. the union of the x_1-axis and the x_2-axis is not a subspace of $\mathbb{R}^2$. It fails vector addition.

In this lecture V is always a vector space over F, where $F = \mathbb{R}, \mathbb{C}$ or $\mathbb{F}_2$.   Lecture 25

Every vector space V contains "trivial" subspaces $\{0\}$ and V itself. For example, $\mathbb{R}$ contains only trivial subspaces. Indeed, if $W \neq \{0\}$, then $\exists\, w \in W$, $w \neq 0$. Then $\mathbb{R} = \{\lambda w \mid \lambda \in \mathbb{R}\}$.

In $\mathbb{R}^2$ there are non-trivial subspaces, e.g. take $x = (x_1, x_2) \neq 0$. Then λx is a vector subspace. Of course $\{\lambda x \mid \lambda \in \mathbb{R}\}$ is a line through zero. So any line through zero in $\mathbb{R}^2$ is a vector subspace of $\mathbb{R}^2$. These are given by $ax_1 + bx_2 = 0$.

Proposition 10.4. *Let A be a matrix of size $m \times nS$ with entries in I. Then the set of vectors $x \in F^n$ such that $Ax = 0$ is a vector subspace of F^n.*

Proof. By Prop 10.2 we need to check that $Ax = 0$ and $Ay = 0$, then $A(x + y) = 0$. This is clear. Similarly, if $Ax = 0$, then for any $\lambda \in \mathbb{R}$, we have $A\lambda x = \lambda Ax = 0$. ∎

> **Definition.** Let V be a vector space over F, and let $v_1, \ldots, v_n \in V$. The vector $\lambda_1 v_1 + \cdots + \lambda_n v_n$ is called a *linear combination* of $v_1, \ldots, v_n$. Here $\lambda_1, \ldots, \lambda_n \in F$ are arbitrary scalars.

Remark. Clearly $\lambda_1 v_n + \ldots \lambda_n v_n \in V$ for any $\lambda_1, \ldots, \lambda_n \in F$.

> **Example 10.5** ($\mathbb{R}^2$)**.** It is easy to see that if $x, y \in \mathbb{R}^2$ are not multiples of each other, then $\mathbb{R}^2 = \{\lambda x + \mu y \mid \lambda, \mu \in \mathbb{R}\}$
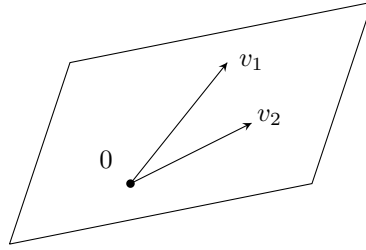
> **Definition.** Let $v_1, \ldots, v_n$ be vectors in V. Then the set of all linear combinations of $v_1, \ldots, v_n$ is called the *linear span* of $v_1, \ldots, v_n$. This is denoted by $\mathrm{Sp}(v_1, \ldots, v_n)$.

Proposition 10.6. *For any $v_1, \ldots, v_n \in V$ the linear span $Sp(v_1, \ldots, v_n)$ is a vector space of V.*

Proof. Again, by Prop 10.2 we need only to check that $\mathrm{Sp}(v_1, \ldots, v_n)$ is closed under $+$ and ($\times$ by scalars).

Indeed

$$\lambda_1 v_1 + \cdots + \lambda_n v_n + (\mu_1 v_1 + \cdots + \mu_n v_n)$$
$$= (\lambda_1 + \mu_1)v_1 + \cdots + (\lambda_n + \mu_n)v_n$$

Also for $\mu \in F$ we have $\mu(\lambda_1 v_1 + \cdots + \lambda_n v_n) = (\mu_1 v_1)v_1 + \cdots + (\mu_n \lambda_n)v_n$.   ∎

Remark. $b = (b_1, \ldots, b_n)^t \in F^n$. *When is b an element of $Sp(v_1, \ldots, v_n)$, where $v_i \in F^n$?*

Consider the $m \times n$ matrix

$$A = \begin{pmatrix} | & & | \\ v_1 & \ldots & v_n \\ | & & | \end{pmatrix}$$

$b \in \mathrm{Sp}(v_1, \ldots, v_n)$ if and only if $b = x_1 v_1 + \ldots x_n v_n$. Equivalently,

$$\begin{pmatrix} b_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} | & & | \\ v_1 & \ldots & v_n \\ | & & | \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = Ax$$

for some $x \in F^n$.

> **Example 10.7.** $V = \mathbb{R}^2$. Consider
>
> $$v_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad v_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 3 \\ 1 \\ -1 \end{pmatrix}$$
>
> *What is $Sp(v_1, v_2, v_3)$?*
>
> Using the remark, we know that $b \in \mathrm{Sp}(v_1, v_2, v_3)$ if and only if $b = Ax$ has a solution.

The augmented matrix of $Ax = b$ is this:

$$\begin{pmatrix} 1 & -1 & 3 & b_1 \\ 2 & 0 & 1 & b_2 \\ 3 & 1 & -1 & b_3 \end{pmatrix}$$

Using Gaussian elimination we get

$$\longrightarrow \begin{pmatrix} 1 & -1 & 3 & b_1 \\ 0 & 2 & -5 & b_2 - 2b_1 \\ 0 & 0 & 0 & b_1 - 2b_2 + b_3 \end{pmatrix}$$

$$\longrightarrow \begin{pmatrix} 1 & -1 & 3 & b_1 \\ 0 & 2 & -5 & b_2 - 2b_1 \\ 0 & 4 & -10 & b_3 - 3b_1 \end{pmatrix}$$

If $b_1 - 2b_2 + b_3 \neq 0$, then on solutions. If $b_1 - 2b_2 + b_3 = 0$, we have solutions (one free variable $\implies$ infinitely many).

Conclusion: $b \in \mathrm{Sp}(v_1, v_2, v_3)$ if and only if $b_1 - 2b_2 + b_3 = 0$. In other words, $\mathrm{Sp}(v_1, v_2, v_3)$ is a plane in $\mathbb{R}^3$ given by the equation $x_1 - 2x_2 + x_3 = 0$.

Example 10.8. Same v_1, v_2 but now let $v_3 = \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix}$.

What is $Sp(v_1, v_2, v_3)$?

We need to solve $Ax = b$, where the augmented matrix is

$$\begin{pmatrix} 1 & -1 & 3 & b_1 \\ 2 & 0 & 1 & b_2 \\ 3 & 1 & 0 & b_3 \end{pmatrix}$$

Once again, follow the same procedure:

$$\longrightarrow \begin{pmatrix} 1 & -1 & 3 & b_1 \\ 0 & 2 & -5 & b_2 - 2b_1 \\ 0 & 4 & -9 & b_3 - 3b_1 \end{pmatrix}$$

$$\longrightarrow \begin{pmatrix} 1 & -1 & 3 & b_1 \\ 0 & 2 & -5 & b_2 - 2b_1 \\ 0 & 0 & 1 & b_1 - 2b_2 + b_3 \end{pmatrix}$$

The system has a solution for any choice of b_1, b_2, b_3 (and there is exactly one solution!) So $\mathrm{Sp}(v_1, v_2, v_3) = \mathbb{R}^3$.

Definition. Let V be a vector space over F, and let $W \subset V$ be a subspace. If $W = \mathrm{Sp}(v_1, \ldots, v_n)$, then $v_1, \ldots, v_n$ is called a *spanning set* for W. Then we say that $v_1, \ldots, v_n$ spans W.

Example 10.9. The vectors $e_1 = (1, 0, \ldots, 0), \ldots, e_n = (0, 0, \ldots, 1)$ span F^n. Indeed $(x_1, \ldots, x_n) = x_1 e_1 + \cdots + x_n e_n$.

Go back to the first example... The plane $x_1 - 2x_2 + x_3 = 0$ is $\mathrm{Sp}(v_1, v_2, v_3)$. In fact, this plane is also $\mathrm{Sp}(v_1, v_2)$. The reason for this is that v_3 is a linear combination of v_1 and v_2. Indeed

$$v_3 = \begin{pmatrix} 3 \\ 1 \\ -1 \end{pmatrix}, \quad v_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad v_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

$$v_3 = \frac{1}{2} v_1 - \frac{5}{2} v_2$$

Therefore $\lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 = (\lambda_1 + \frac{\lambda_3}{2}) v_1 + (\lambda_2 - \frac{5}{2} \lambda_3) v_2$.

Definition. Let V be a vector space over F. Then vectors $v_1, \ldots, v_n$ are *linearly dependent* if there exists $\lambda_1, \ldots, \lambda_n \in F$ not all equal to zero such that $\lambda_1 v_1 + \cdots + \lambda_n v_n = 0 \in V$.

The motivation for this is when $v_1, \ldots, v_n$ are linearly dependent, one of these vectors is a linear combination of all the others.

Indeed some $\lambda_i \neq 0$. To fix ideas assume $\lambda_1 \neq 0$. Then

$$v_1 = -\frac{1}{\lambda_1}(\lambda_2 v_2 + \cdots + \lambda_n v_n)$$

In this case $\mathrm{Sp}(v_1, \ldots, v_n) = \mathrm{Sp}(v_2, \ldots, v_n)$.

Definition. $v_1, \ldots, v_n$j are *linearly independent* if the only way to write the zero vector as a linear combination of $v_1, \ldots, v_n$ is with all coefficients equal to 0, i.e. if $0 = \lambda_1 v_1 + \ldots \lambda_n v_n \implies \lambda_1 = \cdots = \lambda_n = 0$.

Remarks.

(i) One vector $v \in V$ is linearly independent (assuming $v \neq 0$, i.e. v is a non-zero vector)

(ii) Two vectors $v_1, v_2 \in V$ are linearly independent if and only if they are not proportional. $\lambda_1 v_1 + \lambda_2 v_2 = 0$. If $\lambda_1 \neq 0$, then $v_1 = -\frac{\lambda_2}{\lambda_1} v_2$. If $\lambda_2 \neq 0$, then $v_2 = -\frac{\lambda_1}{\lambda_2} v_1$.

(ii) Let $V = F^m$. Suppose $v_1, \ldots, v_n \in V$. Consider the $m \times n$ matrix A such that $v_1, \ldots, v_n$ are the columns of A:

$$A = \begin{pmatrix} | & \cdots & | \\ v_1 & \cdots & v_n \\ | & \cdots & | \end{pmatrix}$$

Then $v_1, \ldots, v_n$ are linearly dependent if and only if the equation $Ax = 0$ has a non-zero solution.

Clearly $x = (x_1, \ldots, x_n)^t$ is a solution, then $x_1 v_1 + \cdots + x_n v + n = 0$. If $Ax = 0 \implies x = 0$, then $v_1, \ldots, v_n$ are linearly independent. If $ax = 0$ has a

non-zero solution then $v_1, \ldots, v_n$ are linearly dependent.

Proposition 10.10. *Any subset of a set of linearly independent set of vectors is linearly independent.*

Proof. Suppose that $v_1, \ldots, v_n$ are linearly independent. Without loss of generality, we can assume that our subset is $v_1, \ldots, v_k$ for some $k \in n$. If $v_1, \ldots, v_k$ are linearly dependent, then we can find $\lambda_1, \ldots, \lambda_n \in F$ (scalars) such that $\lambda_1 v_1 + \ldots \lambda_k v_k = 0$, and not all of $\lambda_1, \ldots, \lambda_k$ are 0.

Consider
$$\lambda_1 v_1 + \cdots + \lambda_k v_k + 0 v_{k+1} + \cdots + 0 v_n = 0$$

This is a linear combination of all $v_1, \ldots, v_n$ with at least one non-zero coefficient. Hence $v_1, \ldots, v_n$ are linearly dependent. This is a contradiction, therefore $v_1, \ldots, v_k$ are linearly independent. ∎

Proposition 10.11. *Let V be a vector space over F. Let $v_1, \ldots, v_n \in V$. Then $v_1, \ldots, v_n$ are linearly dependent if and only if one o them is a linear combination of the others.*

Proof. Suppose $v_1, \ldots, v_n$ are linearly dependent. Then we can find $\lambda_1, \ldots, \lambda_n$ such that $\lambda_1 v_1 + \cdots + \lambda_n v_n = 0$ and one of λ_is, say λ_1 is non-zero. Then

$$v_1 = -\frac{1}{\lambda_1}(\lambda_2 v_2 + \ldots \lambda_n v_n)$$

Conversely if $v_1 = \mu_2 v_2 + \ldots \mu_n v_n$ where μ_i's $\in F$, then

$$(-1)v_1 + \mu_2 v_2 + \cdots + \mu_n v_n = 0$$

$(-1) \neq 0$, so at least one coefficient is non-zero. Hence $v_1, \ldots, v_n$ are linearly dependent. ∎

> **Definition.** Let V be a vector space over F. A set of vectors $v_1, \ldots, v_n \in V$ is called a *basis* of V if
>
> (i) $V = \mathrm{Sp}(v_1, \ldots, v_n)$
>
> (ii) $v_1, \ldots, v_n$ are linearly independent.

In other words, every vector of V is a linear combination of $v_1, \ldots, v_n$ (this is property (i)), and and no v_i is a linear combination of $v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n$.

> **Examples 10.12.**
>
> (i) $V = F^n$. Then $e_1 = (1, 0, \ldots, 0), e_2 = (0, 1, \ldots, 0), \ldots, e_n = (0, 0, \ldots, 1)$ is a basis of F^n.
>
> $(x_1, \ldots, x_n) = x_1 e_1 + \ldots x_n e_n$, so $\mathrm{Sp}(e_1, \ldots, e_n) = F^n$. Clearly $e_1, \ldots, e_n$ are linearly independent: $\lambda_1 e_1 + \cdots + \lambda_n e_n = 0$. This says $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is the zero vector. Hence $\lambda_i = 0$, $i = 1, \ldots, n$. So $e_1, \ldots, e_n$ is a basis.

Warning. There are many other bases! e.g. R^2 has a basis $(1,1)$ and $(1,3)$. Indeed $(1,1)$ and $(1,3)$ are not proportional, hence are linearly independent.

(ii) Let V be a vector space of polynoials in x of degree at most d.

$$V = \{f_d x^d + \cdots + f_1 x_1 + f_0 \mid f_i \in F\}$$

$f_d x^d + \cdots + f_0$ is a linear combination of $1, x, \ldots, x^d$ with coefficients $f_0, \ldots, f_d$. But $1, x, \ldots, x^d$ are linear indepdent vectors of V. So suppose $lambda_0 + \lambda_1 x + \cdots + \lambda_d x^d = 0$ (the zero polynomial.) Conclusion: $1, x, \ldots, x^d$ is a basis of the vector space of polynomials of degree $\leq d$.

Other basis exist: e.g. $1, x+1, x^2+1, \ldots, x^d + 1$ is also a basis.

Remark. Let $V = F^n$. Then $v_1, \ldots, v_n \in V$ form a basis of V if and only if the matrix equation $Ax = 0$ has only one solution $x = 0$. Here

$$A = \begin{pmatrix} | & \cdots & | \\ v_1 & \cdots & v_n \\ | & \cdots & | \end{pmatrix}$$

We've seen that $v_1, \ldots, v_n$ are linearly independent if the only solution is the zero solution. We still need to justify that if $Ax = 0$ has only one solution, then $v_1, \ldots, v_n$ span V.

When does $v_1, \ldots, v_n \in F^n$ form a basis?

Consider

$$A = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix}$$

Answer: $v_1, \ldots, v_n$ are a basis of F^n is and only if the equation $Ax = 0$, where $x = (x_1, \ldots, x_n)^t$ has the unique solution 0. Equivalently, $\det(A) \neq 0$.

x is a solution $\iff x_1 v_1 + \cdots + x_n v_n = 0$. Hence, since $v_1, \ldots, v_n$ are linearly independent the only solution is $x_1 = \cdots = x_n = 0$. $\det(A) \neq 0 \implies A^{-1}$ exists.

To prove that $F^n = \mathrm{Sp}(v_1, \ldots, v_n)$, we must observe that any vector $b = (b_1, \ldots, b_n)^t$ can be written as $b = Ac$ for some $c = (c_1, \ldots, c_n)^t$. This is so because $b = Ac$ says that $b = c_1 v_1 + \cdots + c_n v_n$. Take $c = A^{-1} b$. Then $b = Ac$ as required.

Proposition 10.13. *Let V be a vector space over G with a basis $\{v_1, \ldots, v_n\}$. Then any vector $v \in V$ is uniquely written as $v = \lambda_1 v_1 + \cdots + \lambda_n v_n$.*

Proof. Suppose that

$$v = \lambda_1 v_1 + \cdots + \lambda_n v_n$$

and

$$v = \mu_1 v_1 + \ldots \mu_n v_n$$

Then $(\mu_1 - \lambda_1)v_1 + \cdots + (\mu_n - \lambda_n)v_n = 0$. Since $v_1, \ldots, v_n$ is a basis, these vectors are linearly independent. By the definition of linear independence, we must have $\lambda_1 = \mu_1, \ldots, \lambda_n = \mu_n$. ∎

Theorem 10.14

Let V be a vector space over F. Suppose that $\{v_1, \ldots, v_n\}$ and $\{w_1, \ldots, w_m\}$ are both bases of V. Then $m = n$.

- So any basis of V has the same number of elements.

Proof. $V = \mathrm{Sp}(w_1, \ldots, w_m)$ and $V = \mathrm{Sp}(v_1, \ldots, v_n)$. Thus we can write:

$$v_i = \sum_{j=1}^{n} a_{ij} w_j \quad \text{for some } a_{ij} \in F$$

Let A be the $m \times n$ matrix with entries a_{ij}.

Similarly we write:

$$w_j = \sum_{k=1}^{m} b_{kj} v_k \quad \text{for some } b_{jk} \in F$$

Therefore,

$$v_i - \sum_{j=1}^{n} a_{ij} \cdot \sum_{k=1}^{m} b_{jk} v_k \qquad (*)$$

By Proposition 10.7, there is only one way of writing v_i as a linear combination of $v_1, \ldots, v_n$, namely as $v_i = 0v_1 + \cdots + 1 \cdot v_i + \cdots + 0v_n$.

Hence RHS of $(*)$ is simply v_i. In otherwords, we have

$$\sum_{j=1}^{n} \sum_{k=1}^{m} a_{ij} b_{jk} = 1 \quad \text{if } k = i$$

and

$$\sum_{j=1}^{n} \sum_{k=1}^{m} a_{ij} b_{jk} = 0 \quad \text{if } k \neq i$$

The framed expressions are the entries of AB.

The ii entry of AB is the first sum, that is

$$\sum_{j=1}^{n} \sum_{k=1}^{m} a_{ij} b_{jk} = 1$$

The $1k$-entry of AB, $i \neq k$, is the second sum, that is

$$\sum_{j=1}^{n} \sum_{k=1}^{m} a_{ij} b_{jk} = 0$$

This means $AB = I_m$, the identity matrix of size $m \times m$.

By swapping v_i's and w_j's, we show similarly that $BA = I_n$, the identity matrix of size $n \times n$. Thus $AB = I_n$ and $BA = I_m$.

Recall that the trace of a square matrix is the sum of the diagonal elements.

Fact: $\text{Tr}(AB) = \text{Tr}(BA)$.

This is enough to finish the proof because $\text{Tr}(I_n) = n$ and $\text{Tr}(I_m) = m$.

Proof of Fact. The ii entry of AB is

$$\sum_{j=1}^{n} a_{ij} b_{ji} \implies Tr(AB) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} b_{ji}$$

The jj-entry of BA is

$$\sum_{i=1}^{m} b_{ji} a_{ij}$$

So

$$Tr(BA) = \sum_{j=1}^{n} \sum_{i=1}^{m} b_{ji} \cdot a_{ij}$$
$$= \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ji}$$
$$= Tr(AB)$$

The proves the fact, and hence the theorem. ∎

> **Definition.** Suppose a vector space V over G has a basis consisting of n vectors. Then n is called the *dimension of $V = dimV$*.
>
> If V has a basis (of finitely many elements), then V is called a *finite-dimensional vector space*.

> **Example 10.15.** Let V be the space of all polynomials in x with real coefficients.
>
> **Claim:** V is not finite-dimensional .
>
> *Proof.* Suppose that V is spanned by finitely many polynomials, say $f_1(x), \ldots, f_n(x)$.
>
> Let d be a natural number so that $d > \deg f_i(x)$ with $i = 1, \ldots, n$. Then x^d is not a linear combination of $f_1(x), \ldots, f_n(x)$. ∎

> **Theorem 10.16**
>
> Let V be a vector space. Suppose that $v_1, \ldots, v_n \in V$ span V. Then
>
> (i) $\dim V \leq n$
>
> (ii) There is a subset of $\{v_1, \ldots, v_n\}$, which is a basis of V.

Note that (ii) $\implies$ (i). Indeed $\dim V$ = number of elements in any basis of V, so $\dim V \leq n$, because removing some of the vectors from $\{v_1, \ldots, v_n\}$ we obtain a basis of V. Hence (ii) $\implies$ (i).

Proof of (ii). We use the *casting out process*.

Start with $v_1, \ldots, v_n$. If $v_1, \ldots, v_n$ are linearly independent, then $\{v_1, \ldots, v_n\}$ is a basis. If $v_1, \ldots, v_n$ are linearly independent [i.e. linearly dependent] then we find $\lambda_1, \ldots, \lambda_n \in F$ such that $\lambda_1 v_1 + \lambda_n v_n = 0$ and at least one coefficient $\lambda_i \neq 0$. Say $\lambda_n \neq 0$. Then

$$v_n = -\frac{1}{\lambda_n}(\lambda_1 v_1 + \cdots + \lambda_{n-1} v_{n-1})$$

Therefore V is spanned by $v_1, \ldots, v_{n-1}$, i.e. $V = \mathrm{Sp}(v_1, \ldots, v_n) = \mathrm{Sp}(v_1, \ldots, v_{n-1})$.

Continue like this, we stop when we are met with a linearly independent set of vectors. But we ensure that these vectors span V. Hence we have obtained a basis of V. ∎

We wanted to classify all vector space of $\mathbb{R}^3$. $\dim(\mathbb{R}^3) = 3$, and $\dim(\{0\}) = 0$ Lecture 29 by definition. (We define the dimension of $\{0\}$ to be zero.)

If $V \subset \mathbb{R}^3$ is a vector space, then $\dim(V)$ can be $0, 1, 2,$ or 3. If $\dim(V) = 3$, then V has a basis of 3 vectors, but then it's a basis of $\mathbb{R}^3$. Hence $V = \mathbb{R}^3$.

Felina. Finally we look at an application to error-correcting codes:

10.2 * Error correcting codes *

Now $F = \mathbb{F}_2 = \{0, 1\}$ with binary addition and multiplication. $1 + 1 = 0$.

Consider the n-dimensional vector space $(\mathbb{F}_2)^n = \{(x_1, \ldots, x_n) \mid x_i \in \mathbb{F}_2\}$. We've seen that every vector space of $(\mathbb{F}_2)^k$ is a vector subspace of dimension n over $\mathbb{F}_2$, where $0 \leq m \leq n$.

Let $v_1, \ldots, v_m$ be a basis of V, where $m = \dim(V)$. We know that every vector $v \in W$ is uniquely written as

$$v = \lambda_1 v_1 + \ldots \lambda_m v_m$$

where $\lambda_i \in \mathbb{F}_2$, i.e. $\lambda_i = 0$ or $\lambda_i = 1$.

The number of elements in $(\mathbb{F}_2)^n$ is 2^n .If $V \subset (\mathbb{F}_2)$ has $\dim(V) = m$, then $|V| = 2^m$.

Recall that if $v = (x_1, \ldots, x_n)$ and $w = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$, then the distance between v and w is

$$\sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}$$

This doesn't work well for $\mathbb{F}_2$. Instead consider the *Hamming distance.*

> **Definition.** Let $v, w \in (\mathbb{F}_2)^n$. The *Hamming distance*, $\text{dist}(v, w) =$ the number of $i \in \{1, \ldots, n\}$ such that $v_i \neq w_i$.

For example, $\text{dist}(v, 0) =$ the number of non-zero coordinates of v.

Any reasonable notion of distance should satisfy three axioms:

(i) $\text{dist}(v, v) = 0$

(ii) $\text{dist}(v, w) = \text{dist}(w, v) \quad \forall v, w \in V$

(iii) $\text{dist}(v, w) \leq \text{dist}(u, u) + \text{dist}(u, w)$

    ($\text{dist}(v, w)$ is always a non-negative real number.)

Exercise: Check Axiom (iii).

Error-correcting codes are used to transmit messages in such a way that a distorted message can be decoded correctly.

Messages are first encoded into sequences of 0 and 1 of some length. We get, say 2^d possible messages, encoded into sequences of length 1.

The idea of error correction in the case of 1 error:

The Hamming code

We are going to realise 2^d messages as elements of a vector subspace of $(\mathbb{F}_2)^n$, $V \subset (\mathbb{F}_2)^n$.

The construction: consider the matrix of size $m \times 2$ such that the columns are all vectors with coordinates 0 and 1 except the all-0 vector. E.g. if $m = 3$,

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

Define $V = \{x \in \mathbb{F}^{2^m - 1} \mid Ax = 0\}$.

Crucial property: Suppose $v, w \in V, v \neq w$. We transmit v, w and receive v', w'. We assumed that at most one error occurs during transmission, so either $v' = v$ or $\text{dist}(v, v') = 1$ and similarly for w.

The main point is that the set of $v' \in (\mathbb{F}_2)^n$ such that $\text{dist}(v, v') \leq 1$ does not overlap with the set of $w' \in (\mathbb{F}_2)^n$ such that $\text{dist}(w, w') \leq 1$. Once we've proved that we know that the closest vector to v' is the original vector v.

In fact for any $v, w \in V$, $\text{dist}(v, w) \geq 3$ if $v \neq w$.

Proof.

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

If $\text{dist}(v, w) = 1$, then $v - w$ has exactly one non-zero coordinate. But V has no such vectors! (The matrix has no zero columns).

If $\text{dist}(v, w) = 2$, then $v - w$ has exactly two non-zero co-ordinates, but there are non such vectors in V (The matrix has no two identitcal columns). Hence $\text{dist}(v, w) \geq 3$, if $v \neq w$ in V.   ∎

Proof of the crucial property. If $v' = w'$, $\text{dist}(v, v') \leq 1$ and $\text{dist}(w, w') \leq 1$. Then by the triangle inequality

$$\text{dist}(v, w) \leq \text{dist}(v, v') + \text{dist}(w, w') \leq 2$$

a contradiction.   ∎

The decoding is very easy:

Let $v \in V$, so that $Av = 0$. Let v' be the received vector. We assume that $v' = v$ or $v' = v + e$, where e has exactly one non-zero coordinate.

Recipe for decoding.

Av' is a column vector with 3 co-ordinates. If $Av' = 0$ then $v' = v$ and no errors occurred. Otherwise Av' is a column of A. The position of Av' in A is the coordinate where the error has occurred. Hence we can recover v.

Summary

V is a vector space of dimension

$$\dim V = 2^m - 1 - m$$

Exercise: Show that the system of linear equations that defines V has this number of free variables.

Hence $|V| = 2^{2^m - 1 - m}$, with $V \subset (\mathbb{F}_2)^{2^m - 1}$.

Final Comment: The whole space $(\mathbb{F}_2)^{2^m}$ is the disjoint union of *Hamming spheres* of radius, i.e. $(\mathbb{F}_2)^{2^m - 1}$ is the disjoint union of $\{a' \mid \text{dist}(n - v)\}$ over all $v \in V$. In other words each vector of $(\mathbb{F}_2)^{2^m - 1}$ is at the distance 1 from vector in v, which is unchanged.

- End of Geometry & Linear Algebra -