Mathematical Structures

Axioms are the rules which define a mathematical Structure

Groups

Groups have a set G, a binary operation $*$ and some axioms:

- $*$ is closed on G, i.e. $* : G \times G \to G$
- $*$ is associative for all a, b, c in G

  $$a * (b * c) = (a * b) * c$$

- There exists an identity element $e \in G$ i.e.

  $$\forall a \in G, \quad a * e = e * a = a$$

  for all

  (e.g. for addition, $e = 0$
  multiplication, $e = 1$)

- $*$ has an inverse operator, i.e.

  $$\forall a \in G, \ \exists b \in G. \ a * b = b * a = e$$

  there exists

Examples of groups

- real numbers with +
- Integers with +
- Symmetries
- Even numbers with +

Not groups

- Odd numbers with +
  (1+3=4 → not odd)
  (does not include 0 - no identity)
- Rational numbers with ×
  (no inverse for 0 - can't divide by 0)

Identity is unique

e_1 e_2

Consider $e_1 * e_2 = e_2$ because e_1 is identity

$e_1 * e_2 = e_1$

So $e_1 = e_1 * e_2 = e_2$

$\square$ (QED is for noobs)

Modulo arithmetic - finite group

$\mathbb{Z}_4$

a+b

b \ a	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

Closed ✓

Associative ✓

Identity $= 0$ ✓

Inverse ✓

num	inv
0	0
1	3
2	2
3	1

Distributivity $\to$ $a \cdot (b+c) = (a \cdot b) + (a \cdot c)$

Vector Space

A vector space is a mathematical structure with sets, operations and axioms.

- Has two sets - Scalars F and vectors V.
- V forms a commutative group under addition

  $u + v = v + u$

- F forms a field
- Scalar multiplication - $F \times V \to V$

Notation Conventions

- vector variables in bold or over/underlined.
  e.g. $\underline{u}, \underline{v} \in V$
- write scalars plain, $a, b \in F$
- Vectors in $\mathbb{R}^3$ as row $(1, -2, 0)$ or as column, $\begin{pmatrix} 1 \\ -2 \\ 0 \end{pmatrix}$

Axioms

- Commutative group and field axioms for vectors and scalars
- $a(b\underline{v}) = (ab)\underline{v}$
- $1\underline{v} = \underline{v}$

  one
- $a(\underline{u}+\underline{v}) = a\underline{u} + a\underline{v}$
- $(a+b)\underline{v} = a\underline{v} + b\underline{v}$

Examples

- $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$ over $\mathbb{R}$
  - Vectors in $\mathbb{R}^3$, Scalars in $\mathbb{R}$

$$7 \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 7 \\ 14 \\ 21 \end{pmatrix}$$

- $\mathbb{Z}_2^8$ over $\mathbb{Z}_2$

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

- Polynomials over $\mathbb{R}$

Linear Combinations

Given vectors $\underline{v}_1, \ldots, \underline{v}_n$ and Scalars $a_1, \ldots, a_n$

We can form a linear Combination

$$a_1 \underline{v}_1 + \ldots + a_n \underline{v}_n$$

$$\sum_{i=1}^{n} a_i \underline{v}_i$$

Span: The set of all linear Combinations of a set of vectors

The span of a single vector is a straight line

Spanning Set - A set of vectors S is a spanning set of a vector space (F, V) if $\text{Span}(S) = V$

Linear dependence - A set of vectors $S = \{s_1, \dots, s_n\}$ is linear dependent iff there exists a linear combination

$$\sum_{i=1}^{n} a_i s_i = 0 \text{ and } \exists i \cdot a_i \neq 0$$

i.e. if we can write one vector as a linear combination of the others

$$\text{if } a_1 \underline{v}_1 + a_2 \underline{v}_2 + a_3 \underline{v}_3 = 0$$

$$a_1 \underline{v}_1 = (-a_2) \underline{v}_2 + (-a_3) \underline{v}_3$$

$$\underline{v}_1 = \left(-\frac{a_2}{a_1}\right) \underline{v}_2 + \left(\frac{-a_3}{a_1}\right) \underline{v}_3$$

Linearly independent if not linearly dependent

Basis of a Vector Space

- A subset S of a vector space V is a basis for V if S spans V and S is linearly independent → S is the smallest spanning set of V.
- The standard (canonical) basis for $\mathbb{R}^3$ is $\{(1,0,0), (0,1,0), (0,0,1)\}$

Dimension of a Vector Space

- The cardinality or size of the vector space's basis
- The dimension of $\mathbb{R}^3$ is 3
- We can prove that a dimension is well defined if all basis sets are the same size
- Dimension can be infinite

Basis representation

- A set of vectors is a basis if it spans the whole space and is linearly independent
- for a vector space V the representation of any $\underline{v} \in V$ as a linear combination of a given basis $\{s_1, s_2, \ldots, s_n\}$ is unique
- Proof - by contradiction -

$$v = a_1 \underline{s}_1 + \ldots + a_n \underline{s}_n$$
$$v = b_1 \underline{s}_1 + \ldots + b_n \underline{s}_n$$
$$0 = (a_1 \underline{s}_1 + \ldots + a_n s_n) - (b_1 \underline{s}_1 + \ldots + b_n \underline{s}_n)$$
$$0 = (a_1 - b_1)\underline{s}_1 + \ldots + (a_n - b_n) s_n$$
$$\Rightarrow a_1 - b_1 = 0,\ a_2 - b_2 = 0, \ldots, a_n - b_n = 0$$
$$a_1 = b_1 \ \ldots \ a_n = b_n$$

Not distinct

$\square$

Linear Maps (Linear Transformations)

If we have a function $f : V \to W$ where V and W are vector spaces over a field F then f is a linear map

If for any vectors $\underline{u}, \underline{v} \in V$ and $a \in F$

$$f(\underline{u} + \underline{v}) = f(\underline{u}) + f(\underline{v})$$

$$f(a\underline{v}) = a f(\underline{v})$$

Linear maps respect linear combinations

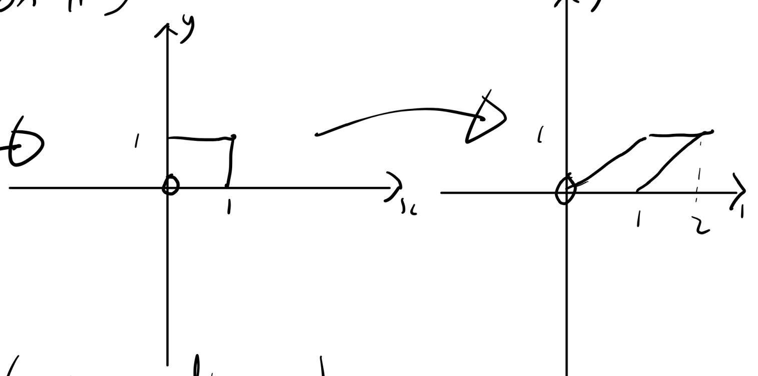
- i.e. $f(a\underline{u} + b\underline{v}) = a f(\underline{u}) + b f(\underline{v})$

If $V = W$, f is an endomorphism

linear maps preserve lines

Examples

- Scaling (on $\mathbb{R}^3$)
- Reflection
- Rotation
- Shearing



- Projection (reducing dimensions)
- Embedding (increasing dimensions)
- Differentiation of polynomials

Every vector $\underline{v}$ can be represented as a linear combination of basis vectors $\underline{s}_i \in S$

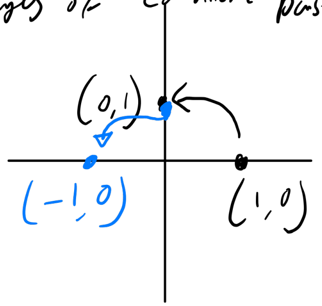
apply linear map f:

$$f(\underline{v}) = f\left(\sum_{i=1}^{n} a_i \underline{s}_i\right) = \sum_{i=1}^{n} a_i f(\underline{s}_i)$$

i.e. linear map characterised by its action on basis vectors.

Rotation 90° anticlockwise

Images of canonical bases $\underline{e}_1\ (1,0)$ and $\underline{e}_2\ (0,1)$



$\therefore\ \underline{e}_1 \mapsto (0,1)$ or $\underline{e}_2$

$\underline{e}_2 \mapsto (-1,0)$ or $-\underline{e}_1$

Turning these into column vectors gives

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

reflection in x axis

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

Calculating from basis vector images:

$$\text{rotate}\left((2,3)\right)$$

$$= \text{rotate}\left(2\underline{e}_1 + 3\underline{e}_2\right)$$

$$= 2 \cdot \text{rotate}(\underline{e}_1) + 3 \cdot \text{rotate}(\underline{e}_2)$$

$$= 2\underline{e}_2 + 3 \cdot -\underline{e}_1$$

$$= (-3, 2)$$

Generalising basis images

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

a_{ij} is the ith component of the image of f of jth basis vector

$$f(\underline{e}_j) = \sum_{i=1}^{m} a_{ij} \underline{e}_i$$

The image of the vector $A\underline{x} = A(x_1, \ldots, x_n)$

is $$f\left(\sum_{j=1}^{n} x_j \underline{e}_j\right) = \sum_{j=1}^{n} x_j f(\underline{e}_j)$$

f is the linear map

$$= \sum_{j=1}^{n} x_j \sum_{i=1}^{m} a_{ij} \underline{e}_i$$

$$= \sum_{j=1}^{n} \sum_{i=1}^{m} a_{ij} x_j \underline{e}_i$$

you can ignore this (I hope)

So the ith component of the image of $\underline{x}$ is

$$\sum_{j=1}^{m} a_{ij} x_j$$

Example

$$\begin{pmatrix} 1 & -2 & 3 \\ -4 & 5 & -6 \end{pmatrix} \begin{pmatrix} 9 \\ 8 \\ 7 \end{pmatrix} = \begin{pmatrix} 1\times 9 - 2\times 8 + 3\times 7 \\ -4\times 9 + 5\times 8 - 6\times 7 \end{pmatrix} = \begin{pmatrix} 14 \\ -38 \end{pmatrix}$$

dimension of domain = 3

dimension of codomain = 2

Identity Matrix - the matrix representation of the identity map

All basis vectors map to themselves via the identity matrix

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

I is a diagonal matrix - if $i \neq j$, $a_{ij} = 0$

Combining linear maps

If we have 2 linear maps f and g represented by matrices A and B, what is the matrix for $(f \circ g)(\underline{v}) = f(g(\underline{v}))$

Need to find images of basis vectors

- B columns contain images of basis vectors under g
- Apply A to columns of B in turn and write as columns

This giving the matrix of the map $f \circ g$ which we write as AB

Matrix Multiplication

Given matrices $A \in R^{m \times n}$ and $B \in R^{n \times k}$

The elements c_{ij} of the product $C = AB \in R^{m \times k}$ are

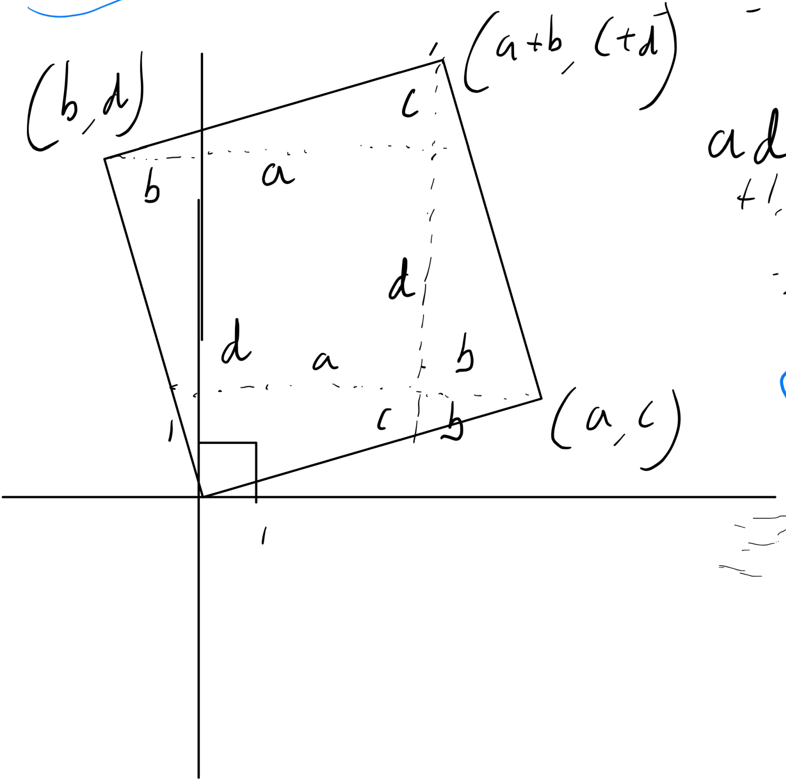
$$c_{ij} = \sum_{l=1}^{n} a_{il} b_{lj}$$

For $1 \leq i \leq m, 1 \leq j \leq k$

This also formalises our definition of multiplying a matrix by a vector $k = 1$

Example:

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



$$ad + \tfrac{1}{2}ac + \tfrac{1}{2}ac + \tfrac{1}{2}bd + \tfrac{1}{2}bd$$

$$\therefore ad + ac + bd$$

please ignore all this :)

The determinant is a measure of how much the area changes by during a linear transformation

Determinant of 2x2 matrix A = $\det(A) = |A|$:

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

$$\begin{vmatrix} 1 & -2 \\ 3 & -4 \end{vmatrix} = 1\cdot(-4) - (-2)\cdot 3 = 2$$

$$\begin{vmatrix} 1 & -2 \\ -2 & 4 \end{vmatrix} = 1\cdot 4 - (-2)\cdot(-2) = 0$$

det is 0 when columns are linearly dependent.

(determinant not defined for non-square matrices).

To find out if vectors are linearly dependent, put em in a matrix and check determinant.

If a matrix is made of columns as vectors $\underline{c}_1, \dots, \underline{c}_n$

$$\det A = \det(\underline{c}_1, \dots, \underline{c}_n)$$

- $\det(a\underline{c}_1, \underline{c}_2, \dots, \underline{c}_n) = a \cdot \det(\underline{c}_1, \underline{c}_2, \dots, \underline{c}_n)$

  Only $\underline{c}_1$ multiplied by a

- $\det(\underline{c}_1 + \underline{c}_1', \dots, \underline{c}_n) = \det(\underline{c}_1, \dots, \underline{c}_n) + \det(\underline{c}_1', \dots, \underline{c}_n)$

  (applies to all columns)
- If $\underline{c}_i = \underline{c}_j$ for some $i \neq j$ then $\det A = 0$ (linearly dependent)
- The determinant of the Identity Matrix is 1.

- Doubling one of the vectors doubles the area/volume
- We can add together the area/vol of cuboids, hypercuboids etc.
- If the image of 2 vectors is the same, the image space has collapsed, e.g. from 3D to 2D

If a matrix B is like A but with two columns swapped, then $\det B = -\det A$

If a matrix A has a column containing all zeroes, then $\det A = 0$

↑ Can write column $\underline{c}_i$ as $0 \times \underline{c}_i$ and take out the 0.

If the scalar multiple of one column is added to another column, the determinant is unchanged.

Determinants can be used to find the area/volume of images or to find out if a set of vectors is linearly dependent.
A matrix has an inverse if $\det M \neq 0$

Determinant of Diagonal Matrix:

$$\begin{vmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & -3 \end{vmatrix} = 2 \cdot \begin{vmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & -3 \end{vmatrix}$$

$$= 2 \cdot 4 \cdot \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -3 \end{vmatrix}$$

multiplication

$$= 2 \cdot 4 \cdot -3 \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}$$

$$= -24$$

Proof of $\det(2\times2)$ using above property:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = a \begin{vmatrix} 1 & b \\ 0 & d \end{vmatrix} + c \begin{vmatrix} 0 & b \\ 1 & d \end{vmatrix}$$

This column is $\begin{pmatrix} 0 \\ d \end{pmatrix} + \begin{pmatrix} b \\ 0 \end{pmatrix}$ so split

$$= a \cdot b \underbrace{\begin{vmatrix} 1 & 1 \\ 0 & 0 \end{vmatrix}}_{=0} + a \cdot d \underbrace{\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}}_{=1} + c \cdot b \cdot \underbrace{\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}}_{=-1} + c \cdot d \cdot \underbrace{\begin{vmatrix} 0 & 0 \\ 1 & 1 \end{vmatrix}}_{=0}$$

$$= a \cdot d - c \cdot b$$

Laplace expansion: $\det C = \sum_{j=1}^{n} (-1)^{i+j} C_{ij} M_{ij}$
of $n \times n$ matrix

$\det(AB) = \det A \det B$
$\det(A^{-1}) = \frac{1}{\det A}$

M_{ij} = det. of Minor, i.e. the det. of matrix C with row i and col. j removed.
This works with any i.

Systems of linear equations - Gaussian elimination

A linear equation in n variables $x_1, \dots x_n$ is an equation of the form $a_1 x_1 + \dots + a_n x_n = b$

where a_i and b are constants

A finite set of linear equations is called a system of linear equations

A solution to such a system is a sequence of $s_1, \dots, s_n$ of numbers such that $x_1 = s_1, \dots, x_n = s_n$ satisfies all equations.

Matrix form of a linear system

A linear system

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

can be written in a matrix form as $A\mathbf{x} = \mathbf{b}$ where

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \text{and } \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

where A is the coefficient matrix of the system.

If A is square and invertible, the solution $\underline{x} = A^{-1}\underline{b}$ can be found.

The augmented matrix is the matrix:

$$(A|\underline{b}) = \left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right)$$

Elementary row operations:

- Multiply an equation (row) through by a non-0 constant
- Interchange two equations (rows)
- Add a constant x one equation to another.

Row echelon form

If we transform the augmented matrix of a linear system in variables x, y, z to the form:

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right)$$

then we know the solution ☺

A matrix is in row echelon form (ref) if it has the following properties:

- If a row is not all 0s then the first non-zero number in it is 1 (the leading 1)
- The rows that are all 0s (if any) are grouped together at the bottom
- If two successive rows are not all 0s then the leading 1 of the higher row occurs further to the left than the leading 1 of the lower row.

A matrix is in reduced ref (rref) if it is in ref, plus

- Each column that contains a leading 1 has 0s everywhere else.

Extracting solutions from ref

Assume that we have transformed the augmented matrix of a linear system to a (r)ref. Examples:

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right) \quad \left(\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right) \quad \left(\begin{array}{ccc|c} 1 & -1 & 0 & 2 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

We have the following possibilities:

- Some row has a leading 1 in the last column.
  Then the system includes equation $0 \cdot x_1 + \ldots + 0 \cdot x_n = 1$.
  Then we know the system has no solutions.
- The number of leading 1s is equal to the number of variables (and there is no leading 1 in the last column).
  Then the system has a unique solution.
- The number of leading 1s is smaller than the number of variables (and there is no leading 1 in the last column).
  Then the system has infinitely many solutions.

General example

Assume a matrix in rref is as follows:

$$\left(\begin{array}{cccc|c} 1 & -1 & 0 & 2 & 2 \\ 0 & 0 & 1 & -1 & 5 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

In equations, this is

$$\begin{array}{rrrrl} x_1 & -x_2 & & +2x_4 & = 2 \\ & & x_3 & -x_4 & = 5 \end{array}$$

- The variables corresponding to the leading 1s (x_1 and x_3 in the example) are the leading variables.
- The other variables are free variables.
- General solution: the leading variables expressed via free variables.
- For the above system: $x_1 = x_2 - 2x_4 + 2, x_3 = x_4 + 5$ (where x_2 and x_4 are arbitrary numbers).

Algorithm - Gaussian elimination (not invented by Gauss)

1. locate pivot column - leftmost column with one or more non-0 values
2. Choose pivot - a non-0 in the pivot column and interchange the first row with the row containing this point to move the pivot row to the top
3. If a is the pivot, multiply the top row by 1/a to get a leading 1
4. Add multiples of the top row as needed so all other values in the pivot column below the pivot are 0.
5. Top row is now 'complete'. Separate the top row from the rest and repeat the process with the matrix made up of the remaining rows.

To find reduced row echelon form: (Gauss-Jordan elimination)

6. Beginning from the bottom-most non-0 row working upwards, add suitable multiples of each row to create 0s above the leading 1s

(see slides for detailed example).

Homogeneous Linear Systems

A linear system $A\underline{x} = \underline{b}$ is homogeneous if $\underline{b}$ is all 0s.
Such a system has a trivial solution: $\underline{x}$ is all 0s. Any other solution is non-trivial.

Theorem: If a homogeneous linear system has n variables and the rref of its augmented matrix has r non-zero rows, the system has $n-r$ free variables.

Corollary: A homogeneous linear system with more variables than equations has infinitely many solutions.

Matrix inverse, rank and Kernel

Reminder of ref:
- All zero rows at the bottom
- Leading 1s in non-zero rows
- Leading 1s move right as you go down rows

rref: Add suitable multiples of each row to create 0s above the leading 1s as well.

Inverse: The inverse, f^{-1}, of a function f undoes the operation of f. Inverse has to be unique for f^{-1} to be a function, otherwise it is a relation. (e.g. x^2 for $x \in \mathbb{Z}$ has no inverse as there are 2 options.)

When $f: X \to Y$ then g is an inverse function of f if $\forall x \in X . g(f(x)) = x$ and $\forall y \in Y . f(g(y)) = y$

e.g. If we want the inverse of $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$

we would want:

$$\begin{pmatrix} w & x \\ y & z \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

this gives us the equations

$w + 3x = 1 \to x = 1 \to w = -2$

$2w + 4x = 0 \to w = -2x$

$y + 3z = 0 \to y = -3z$

$2y + 4z = 1 \to -2z = 1$

$\to z = -\frac{1}{2} \to y = \frac{3}{2}$

$\therefore x = 1,\ w = -2,\ y = \frac{3}{2},\ z = -\frac{1}{2}$

General form of 2x2 inverse:

$$\begin{pmatrix} w & x \\ y & z \end{pmatrix}\underbrace{\begin{pmatrix} a & b \\ c & d \end{pmatrix}}_{\substack{\shortparallel \\ A}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$A^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \quad \text{where } A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

(Inverse does not exist when $\det A = 0$)

In general matrix multiplication is not commutative, however inverse multiplication is.

$$\text{If } AB = I$$

$$A = IA = (AB)A = A(BA)$$

$$\text{So } BA = I = AB$$

$$\therefore AA^{-1} = I = A^{-1}A$$

Inverse by Gauss-Jordan elimination

- Given A we want to find B such that $BA = I$ (or $AB = I$)
- Could write down as a system of equations.
- Can use Gauss-Jordan elimination on augmented matrix $A|I$

Example: find $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}^{-1}$.

$$\begin{pmatrix} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{pmatrix} \xrightarrow[\text{from } R_2]{\text{Sub } R_1} \begin{pmatrix} 1 & 2 & 1 & 0 \\ 0 & -2 & -3 & 1 \end{pmatrix} \xrightarrow[\text{by } -2]{\text{div. } R_2} \begin{pmatrix} 1 & 2 & 1 & 0 \\ 0 & 1 & 3/2 & -1/2 \end{pmatrix}$$

$$\xrightarrow[\text{(reduced ref)}]{\text{Sub } 2R_2 \text{ from } R_1} \begin{pmatrix} 1 & 0 & -2 & 1 \\ 0 & 1 & 3/2 & -1/2 \end{pmatrix}$$

$$\therefore \text{Inverse } \begin{pmatrix} -2 & 1 \\ 3/2 & -1/2 \end{pmatrix}$$

Column Space and Column rank

- All vectors can be written as linear combinations of basis vectors.
- Under a linear map

$$f: X \to Y, \quad f(a\underline{x} + b\underline{y}) = af(\underline{x}) + bf(\underline{y})$$

- $f(X)$ is the span of the columns, called the column space - all possible linear combinations of the columns
- the column rank is the dimension of the column space
- column rank = number of linearly independent columns.

Row rank

- row space is the span of all row vectors
- row rank is the number of linearly independent rows

Row rank = Column rank ☺

∴ rank of a matrix = row rank = column rank

- Each step of G-J elimination leaves column and row rank unchanged
- Rref is reached. Can do more row/column operations to arrive at an identity matrix surrounded by zeroes.
- In this form row rank = column rank (as the identity matrix reached is square)

Implications: - Dimension of $f(X)$ = rank of matrix for f.

- Can calculate rank by G-elimination by counting number of non-zero rows in ref.
- rank = dimension of the image of the linear map represented by a matrix
- A matrix has full rank if the rank is as high as it could be = min(#rows, #columns)

Euclidean Norm

(x_1, x_2) has length $l = \sqrt{x_1^2 + x_2^2}$

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} \quad \text{for } x = (x_1, \dots, x_n)$$

also written as $\|x\|$

A unit vector has length (norm) 1.

A unit vector in the direction of u, $\hat{u} = \dfrac{u}{\|u\|_2}$

Given a vector space X a function $d: X \to R$ is a *norm* if it satisfies

- Triangle inequality. $d(x+y) \leq d(x) + d(y)$ for all $x, y \in X$
- Absolute homogeneity. $d(sx) = |s|\, d(x)$ for all $x \in X$ and scalars s
- Positive definite. If $d(x) = 0$ then $x = 0$

- the length of $u+y$ is less than or equal to the sum of the lengths of u and y

the length of a scalar $\cdot u$ = the scalar $\cdot$ the length of u

If a vector has 0 length, the vector is the zero vector.

Treat $d(u)$ = the length of u → norms are just different ways of defining 'length'. 'Normal' Euclidean length is just one example of a norm

Provable properties of norms

- $d(\underline{0}) = d(0 u) = |0|\, d(u) = 0$

$\therefore d(\underline{0}) = 0$

- $d(\underline{u}) \geq 0$ for all u

0 vector - $\underline{0} = \underline{u} - \underline{u}$

$d(\underline{0}) = d(\underline{u} - \underline{u})$

$d(\underline{u} + -\underline{u}) \leq d(\underline{u}) + d(-\underline{u})$

$= d(\underline{u}) + d(-1 \cdot \underline{u})$

$= d(\underline{u}) + |-1|\, d(\underline{u})$

$= 2d(\underline{u})$

so $d(0) \leq 2d(\underline{u})$

$0 \leq d(\underline{u})$

Manhattan norm: $d_M((u_1, \ldots, u_n)) = \sum_{i=1}^{n} |u_i|$

just add the absolute values of each part of the vector :)

Euclidean and Manhattan norms are examples of p-norms:

p-norms: $\|\underline{u}\|_p = \sqrt[p]{\sum_{i=1}^{n} |u_i|^p}$ → can be written as L_p

Different norms often used in Machine Learning.

when $p=1$ this is the Manhattan Norm $= L_1$

$p=2$ → Euclidean Norm $= L_2$

what if $p \to \infty$?

$\|\underline{u}\|_\infty = \max(|u_1|, \ldots, |u_n|)$ → this is a norm

What if $p \to 0$?

$\lim_{p \to \infty} \|\underline{x}\|_p = \infty$

By assuming $0^0 = 0$:

$L_0 = \underbrace{|x_0|^0 + \ldots + |x_n|^0}$ → This is **Not** a norm

The number of terms in the vector that $\neq 0$ (Since defining $0^0 = 0$)

→ Commonly used, e.g. Hamming distance

Dot Product

$\underline{a} \cdot \underline{b} = \|\underline{a}\|_2 \|\underline{b}\|_2 \cos\theta$

The dot product finds the projection of a vector in the direction of the other

- $\hat{\underline{a}} \cdot \hat{\underline{b}} = \hat{\underline{b}} \cdot \hat{\underline{a}} = \cos\theta$
- $(s\underline{a}) \cdot \underline{b} = s(\underline{a} \cdot \underline{b})$ $\quad s$ = Scalar
- $(\underline{a}_1 + \underline{a}_2) \cdot \underline{b} = \underline{a}_1 \cdot \underline{b} + \underline{a}_2 \cdot \underline{b}$
- $\underline{a} \cdot \underline{b} = 0$ only if they are orthogonal or one is zero
  (orthogonal: at right angles)
- $\underline{a} \cdot \underline{a} = \|\underline{a}\|_2^2$

Calculating Dot product:

If $\underline{a} = (a_1, \ldots, a_n)$ and $\underline{b} = (b_1, \ldots, b_n)$

then $\underline{a} \cdot \underline{b} = \sum_{i=1}^{n} a_i b_i$

Claim: $\underline{a} \cdot \underline{b} = \sum_{i=1}^{n} a_i b_i$

$\underline{a} = \sum_{i=1}^{n} a_i \underline{e}_i$ $\qquad e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix}$ etc

$\underline{b} = \sum_{i=1}^{n} b_i \underline{e}_i$

$(\underline{a} \cdot \underline{b}) = (a_1 e_1 + a_2 e_2 + \ldots) \cdot (b_1 e_1 + \ldots + b_n e_n)$

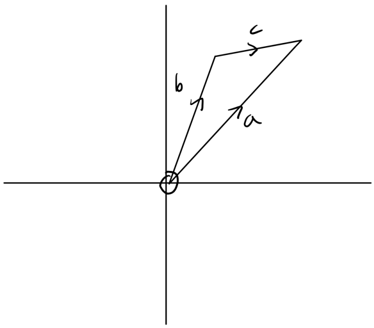
$\underline{e}_i \cdot \underline{e}_j = \|e_i\|_2 \|e_j\|_2 \cos\theta$

$= \cos\theta$

$= \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$ — e_i and e_j are basis vectors so are perpendicular when $i \neq j$ and parallel when $i = j$

$= a_1 b_1 + a_2 b_2 + \ldots + a_n b_n$

Cosine rule: Consider $\underline{a} = \underline{b} + \underline{c}$ and consider $\underline{c} \cdot \underline{c}$



$$\underline{c} \cdot \underline{c} = (\underline{a} - \underline{b})(\underline{a} - \underline{b})$$
$$= \underline{a} \cdot \underline{a} - 2\underline{a} \cdot \underline{b} + \underline{b} \cdot \underline{b}$$
$$\|c\|^2 = \|a\|^2 - 2\|a\|\|b\| \cos\theta + \|b\|^2$$
$$\Downarrow$$
$$c^2 = a^2 + b^2 - 2ab \cos\theta$$

Orthonormality

- Vectors are orthogonal - at right angles - if their dot product is zero.
- Vectors are 'normal' if their l_2 norm is 1.
- A set of vectors is orthonormal if they are all orthogonal and normal
- It is often useful to have an orthonormal basis for a vector space

Kernel of a linear map

If $f: X \to Y$ is a linear map:

Kernel or null space of f is the set of vectors which map to 0, i.e.

$$\{ x \in X : f(x) = 0 \}$$

The kernel forms a vector space - all linear combinations of the kernel are within the kernel.

No inverse of non-square matrices:

Assume $f: X \to Y$ is a linear map

Can an inverse exist if $dim(X) < dim(Y)$?

- $dim(f(X)) \leq dim(X)$, so $dim(f(X)) < dim(Y)$
- Some elements of Y are not in $f(X)$
- No

Can an inverse exist if $dim(X) > dim(Y)$?

- The columns of the matrix cannot be linearly independent
- So there is a non-zero vector that maps to zero (kernel is non-trivial)
- So any inverse cannot be unique because $f(x + k) = f(x)$ if k is in the kernel
- No

Rank-nullity Theorem

Define nullity as the dimension of the kernel

- Given a linear map on vector spaces $T: V \to W$
- If V has finite dimension then:

$$\text{Rank}(T) + \text{Nullity}(T) = \dim(V)$$

($\dim(W)$ not in formula)

Reducing dimensionality important

- Visualisation
- Compression
- Machine learning
- Data representation of text as vectors.

Sparse Matrices : efficiency

Previously we have found that matrix multiplication is $O(n^k)$

From wikipedia : $2.37188 \leq k \leq 3$

Page Rank - an algorithm used by Google to rank web pages in their search engine results, which determines the importance of web results based on the number of sites that link to it.

Basic idea of the algorithm

- Find all the links between web pages
- Construct the directed graph where nodes are pages and edges are links
- Make an initial assignment of PageRank values to each web page (node)
- Transfer the PageRank of each page to the pages it links to (follow edges)
- Weight the Rank according to the number of output links
- Repeat

Solution in the simple case

$$r(p) = \sum_{q \in I(p)} r(q) \,/\, l(q)$$

$$\to \sum \frac{\text{rank of incoming page}}{\text{no. of outgoing links the page has}}$$

for all pages with incoming links to the page p

Where

- r is the rank function
- p and q are pages
- $I(p)$ is the set of pages that link to p
- $l(q)$ is the number of links leaving q

PageRank as a matrix equation

- Give all the pages a number $i = 1 \dots n$
- Find the adjacency matrix A
- Make M the transposed modified adjacency matrix: scale so each column sums to 1
- Write the vector r as $(r_1, \dots, r_n)$

$$Mr = r$$

- More complex version includes *damping*

How to solve the PageRank equation

- Currently more than a billion web pages
- So solve a $10^9 \times 10^9$ matrix
- Good news: successive approximation gets close
- Bad news: just to store the matrix would take 8×10^{18} bytes
- Good news: a lot of them are zeros (sparse) → most pages do not link to every other page on the Internet.
- Bad news: multiplication in numpy is about $O(n^{2.85})$
- Good news: there are quicker ways of multiplying sparse matrices

Santa cares too

- Which reindeer should lead the sleigh?
- What is the weather in Lapland?
- A decent way to forecast the weather is "tomorrow will be the same as today"
- Could simplify to cloudy/snowy/clear

hello :)

Weather as a Markov process

- Look at the history of weather in the place
- Work out transition probabilities between states
- Write transitions as a matrix T
- Write weather today as a (sparse) vector w_0
- $w_1 = Tw_0$
- Then weather in n days time is probably $w_n = T^n w_0$

In long range, forecast is $\underline{w}_\infty$

$T\underline{w}_\infty = \underline{w}_\infty$

In general if $A\underline{v} = \lambda\underline{v}$, then $\underline{v}$ is an eigenvector of A with eigenvalue λ.

↑ A is a matrix.

Basis of eigenvectors

If we can find a basis of eigenvectors then

- Matrix representation for that basis is diagonal
- Change of basis matrix P is invertible
- Write $A = PDP^{-1}$ where D is diagonal

Advantages of diagonalisation

- Diagonal matrices are easy to store (sparse)
- Diagonal matrices are easy to multiply ($O(n)$)
- $A^n = (PDP^{-1})^n = PD^nP^{-1}$
- Much quicker to find powers of matrices
- An example of *matrix factorisation*
- * More in second term *