

LINEAR ALGEBRA

OTHER READING MATERIAL:

THE MATRIX COOKBOOK by Petersen and Pedersen
LINEAR ALGEBRA by Shilov

→ SCALARS →

④ lowercase italics

↳ a single number

↳ specify type : real valued $r \in \mathbb{R}$
natural number $m \in \mathbb{N}$

$$\begin{aligned} i &= 1, \dots, m \\ j &= 1, \dots, n \end{aligned}$$

→ VECTORS →

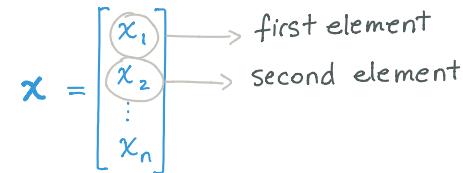
④ lowercase bold

↳ array of numbers arranged in order

↳ identifying points in space

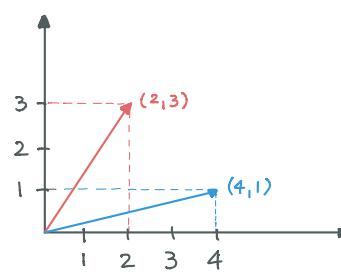
↳ each element is a coordinate along a different axis

↳ defined by magnitude and direction



EXAMPLE: 2D vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$$



$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} \quad \text{set of elements of } \mathbf{x}$$

$$S = \{1, 2, 4\}$$

$$\mathbf{x}_s = \begin{bmatrix} x_1 \\ x_2 \\ x_4 \end{bmatrix}$$

complement of S

$$\mathbf{x}_{-s} = \begin{bmatrix} x_3 \\ x_5 \end{bmatrix}$$

OPERATIONS:

- addition
- subtraction
- scaling
- dot product
- length
- cross product

→ MATRICES →

④ uppercase bold

↳ rectangular 2D array of real numbers

↳ 2 indices : m rows ; n columns

► need to specify the values of its components

Example 3x2 matrix

$$\mathbf{A} = m \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \\ A_{3,1} & A_{3,2} \end{bmatrix} \in \mathbb{R}^{3 \times 2} \quad \mathbf{A}_{:,j} \in \mathbb{R}^3$$

$$\mathbf{A}_{1,:} = \begin{bmatrix} A_{1,1} & A_{1,2} \end{bmatrix} \quad \mathbf{A}_{:,1} = \begin{bmatrix} A_{1,1} \\ A_{2,1} \\ A_{3,1} \end{bmatrix}$$

$\mathbf{A} \in \mathbb{R}^{m \times n}$; $\mathbf{A}_{i,j}$

$\mathbf{A}_{i,:}$ i-th row

$\mathbf{A} = [-a_m^T -]$

$\mathbf{A}_{:,j}$ j-th column

$$\mathbf{A} = \begin{bmatrix} | \\ a_n \\ | \end{bmatrix}$$

$f(\mathbf{A})_{i,j}$ function on elements of \mathbf{A}

OPERATIONS:

- addition
- subtraction
- scaling
- matrix product
- matrix inverse
- trace
- determinant

$$\begin{aligned} \mathbf{A} + \mathbf{B} \\ \mathbf{A} - \mathbf{B} \\ \alpha \mathbf{A} \\ \mathbf{AB} \\ \mathbf{A}^{-1} \\ \text{Tr}(\mathbf{A}) \\ |\mathbf{A}| \end{aligned}$$

→ TENSORS →

↳ array of numbers with more than two axes

$$\mathbf{A}_{i,j,k}$$

OPERATIONS

TRANSPOSE

$$(A^T)_{i,j} = A_{j,i}$$

$$\begin{bmatrix} \text{---} \\ \text{---} \\ \vdots \\ \text{---} \\ \text{---} \end{bmatrix}^T = \begin{bmatrix} \text{---} \\ \text{---} \\ \vdots \\ \text{---} \\ \text{---} \end{bmatrix}^T$$

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \\ A_{3,1} & A_{3,2} \end{bmatrix} = \begin{bmatrix} A_{1,1} & A_{2,1} & A_{3,1} \\ A_{1,2} & A_{2,2} & A_{3,2} \end{bmatrix}$$

- L mirror image of the matrix across the main diagonal \Rightarrow flip rows & columns
- L vector as matrix with one column
 $x = [x_1, x_2, \dots, x_n]^T$
- L scalar : a matrix with one single entry
 $a = a^T$

ADDITION

$$C_{i,j} = A_{i,j} + B_{i,j}$$

- L A and B have to be the same size

$$\begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix} + \begin{bmatrix} B_{1,1} & B_{1,2} \\ B_{2,1} & B_{2,2} \end{bmatrix} = \begin{bmatrix} A_{1,1} + B_{1,1} & A_{1,2} + B_{1,2} \\ A_{2,1} + B_{2,1} & A_{2,2} + B_{2,2} \end{bmatrix}$$

$A + B = C$

SUBTRACTION \rightarrow entry-wise as addition

SCALING

$$B = \alpha A_{i,j} = B_{i,j}$$

$\alpha \in \mathbb{R}$

$$B_{i,j} = \begin{bmatrix} \alpha A_{1,1} & \dots & \alpha A_{1,n} \\ \alpha A_{2,1} & \dots & \alpha A_{2,n} \\ \dots & \ddots & \dots \\ \alpha A_{m,1} & \dots & \alpha A_{m,n} \end{bmatrix}$$

L Scaling A by α

PROPERTIES

$$\begin{aligned} \alpha(\beta A) &= (\alpha\beta)A \\ (\alpha + \beta)A &= \alpha A + \beta A \\ \alpha(A+B) &= \alpha A + \alpha B \\ A(\alpha B) &= \alpha(AB) = (\alpha A)B \end{aligned}$$

PRODUCT

$$\begin{matrix} A & B \\ i \times k & ! \quad k \times j \\ \text{equal} \end{matrix}$$

dimension of C

$$C_{i,j} = A_{i,k} B_{k,j}$$

$$C_{i,j} = \text{row}_i(A) \text{ col}_j(B)$$

$$= \sum_{k=1}^p A_{i,k} B_{k,j}$$

- L dot product $x^T y$
 x, y : same dimensions

$$\begin{bmatrix} A_{1,1} & A_{1,2} & \dots & A_{1,p} \\ A_{2,1} & A_{2,2} & \dots & A_{2,p} \\ A_{3,1} & A_{3,2} & \dots & A_{3,p} \\ \dots & \dots & \ddots & \dots \\ A_{m,1} & A_{m,2} & \dots & A_{m,p} \end{bmatrix} \begin{bmatrix} B_{1,1} & B_{1,2} & \dots & B_{1,n} \\ B_{2,1} & B_{2,2} & \dots & B_{2,n} \\ \dots & \dots & \ddots & \dots \\ B_{p,1} & B_{p,2} & \dots & B_{p,n} \end{bmatrix} = \begin{bmatrix} C_{1,1} & C_{1,2} & \dots & C_{1,n} \\ C_{2,1} & C_{2,2} & \dots & C_{2,n} \\ \dots & \dots & \ddots & \dots \\ C_{m,1} & C_{m,2} & \dots & C_{m,n} \end{bmatrix}$$

$C_{2,2} = A_{2,1} B_{1,2} + A_{2,2} B_{2,1} + \dots + A_{2,p} B_{p,2}$

PROPERTIES

$$\begin{aligned} (A^T)^T &= A & (A+B)^T &= A^T + B^T \\ (AB)^T &= A^T B^T & (\alpha A)^T &= \alpha A^T \\ && \downarrow \text{scalar (real valued)} \end{aligned}$$

PROPERTIES

$$\begin{aligned} A + B &= B + A \\ A + (B+C) &= (A+B)+C \end{aligned}$$

BROADCASTING

matrix-vector addition

$$C = A + b$$

$$C_{ij} = A_{ij} + b_j$$

- L vector b is added to each row of A

✓

E X A M P L E S

$$\begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix} + \begin{bmatrix} b_1 & b_2 \\ b_1 & b_2 \end{bmatrix} = \dots$$

$$\begin{bmatrix} A_{1,1} & A_{1,1} \\ A_{2,1} & A_{2,1} \\ A_{3,1} & A_{3,1} \end{bmatrix} + \begin{bmatrix} b_1 & b_2 \\ b_1 & b_2 \\ b_1 & b_2 \end{bmatrix} = \dots$$

✗

$$\begin{bmatrix} A_{1,1} & A_{1,2} & A_{1,3} \\ A_{2,1} & A_{2,2} & A_{2,3} \end{bmatrix} + \begin{bmatrix} b_1 & b_2 & ? \\ b_1 & b_2 & ? \end{bmatrix} = \dots$$

$(2,3) \neq (1,2)$

PROPERTIES

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

$$A(BC) = (AB)C$$

$$AB \not\equiv BA$$

$$x^T y = y^T x$$

✓ DISTRIBUTIVE

✓ ASSOCIATIVE

✗ COMMUTATIVE

✓ COMMUTATIVE

HADAMARD product

L element-wise

$$A \odot B$$

→ IDENTITY →

matrix

$$I_n \in \mathbb{R}^{n \times n}$$

↳ of order n

$$I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

↳ preserves n -dim vectors $I_n \mathbf{x} = \mathbf{x}; \forall \mathbf{x} \in \mathbb{R}^n$

↳ preserves $m \times n$ matrices $I_m A = A I_n = A; A \in \mathbb{R}^{m \times n}$

INVERSE

$$A^{-1}; A^{-1}A = I_n$$

↳ NON SINGULAR

↳ not all matrices have inverses (e.g. non-square matrices)

↳ inverse exists ▶ full rank matrix

PROPERTIES

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

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

$$(A^{-1})^T = (A^T)^{-1}$$

TRACE

$$\text{Tr}(A) = \sum_{i=1}^n A_{ii}$$

↳ $A \in \mathbb{R}^{n \times n}$

↳ sum of diagonal elements of a matrix

USES:

↳ easier specification of some operations, without the need for summation notation

example: FROBENIUS NORM

$$\|A\|_F = \sqrt{\text{Tr}(AA^T)}$$

PROPERTIES

$$\text{Tr}(A) = \text{Tr}(A^T); A \in \mathbb{R}^{n \times n}$$

$$\text{Tr}(A+B) = \text{Tr}(A) + \text{Tr}(B); A, B \in \mathbb{R}^{n \times n}$$

$$\text{Tr}(\alpha A) = \alpha \text{Tr}(A); A \in \mathbb{R}^{n \times n}, \alpha \in \mathbb{R}$$

$$\text{Tr}(AB) = \text{Tr}(BA); \text{s.t. } AB \text{ is square}$$

$$\text{Tr}(ABC) = \text{Tr}(BCA) = \text{Tr}(CAB)$$

s.t. ABC is square

$\alpha = \text{Tr}(\alpha)$ scalar is its own trace

SYSTEM OF LINEAR EQUATIONS

$$Ax = b; A \in \mathbb{R}^{m \times n} \quad b \in \mathbb{R}^m$$

$x \in \mathbb{R}^n$ unknown

using matrix inverse to solve

$$\begin{aligned} Ax &= b \\ A^{-1}A x &= A^{-1}b \\ I_n x &= A^{-1}b \\ x &= A^{-1}b \end{aligned}$$

- ! if A^{-1} exists, i.e. A is nonsingular
- compute A^{-1} once, so that whenever we change b we find the corresponding x
- represented with only limited precision → should not be used in practice

system of m linear equations in n unknowns

$$A_{1,1}x_1 + A_{1,2}x_2 + \dots + A_{1,n}x_n = b_1$$

$$A_{2,1}x_1 + A_{2,2}x_2 + \dots + A_{2,n}x_n = b_2$$

...

$$A_{m,1}x_1 + A_{m,2}x_2 + \dots + A_{m,n}x_n = b_m$$

each row of A and each element of b provide a constraint

LINEAR DEPENDENCE and Span

set of vectors $\{v^{(1)}, v^{(2)}, \dots, v^{(n)}\}$

LINEAR COMBINATION

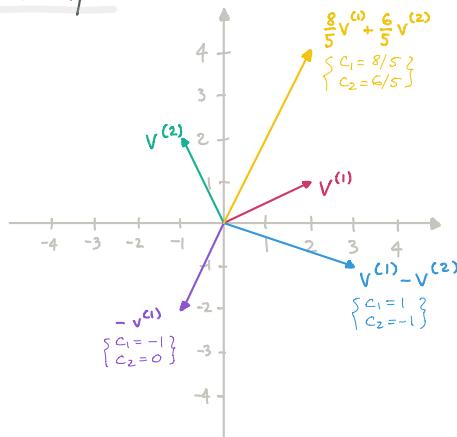
$$\sum_i c_i v^{(i)}$$

- scale vectors $v^{(i)}$ by some scalar c_i
- $c_1, \dots, c_n \in \mathbb{R}$
- $v^{(1)}, \dots, v^{(n)} \in \mathbb{R}^n$

Solutions to $Ax = b$

- | | |
|------------------------|--|
| ✓ $= 1$ | $\implies A^{-1}$ exists; A is square; |
| ✓ $= \infty$ | all columns are linearly independent |
| ✓ $= 0$ | |
| ✗ $1 < \dots < \infty$ | |

Example:



$$\left. \begin{array}{l} v^{(1)} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, v^{(2)} = \begin{bmatrix} -1 \\ 2 \end{bmatrix} \\ c_1 v^{(1)} + c_2 v^{(2)} \\ (1)v^{(1)} + (-1)v^{(2)} = \begin{bmatrix} 3 \\ -1 \end{bmatrix} \\ (-1)v^{(1)} + (0)v^{(2)} = \begin{bmatrix} -2 \\ -1 \end{bmatrix} \\ (8/5)v^{(1)} + (6/5)v^{(2)} = \begin{bmatrix} 2 \\ 4 \end{bmatrix} \end{array} \right\}$$

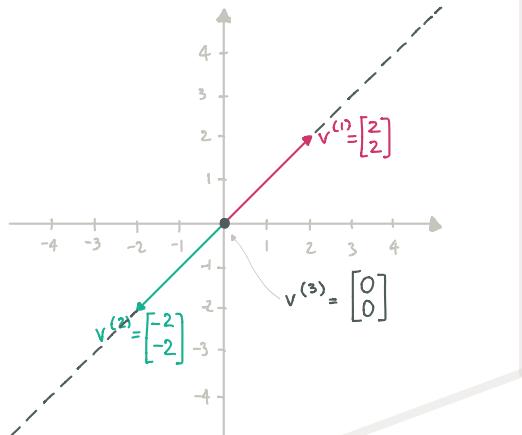
What is the set of all the vectors that we can compute by linear combinations?

↳ all \mathbb{R}^2 can be represented by $v^{(1)}$ and $v^{(2)}$.

→ SPAN

$$\text{span}(v^{(1)}, v^{(2)}) = \mathbb{R}^2$$

Example: $\text{span}(v^{(1)}, v^{(2)}) = \text{line}$
 $\text{span}(v^{(1)}) = \text{line}$
 $\text{span}(v^{(3)}) = 0$



$$\text{span}(\{v^{(1)}, \dots, v^{(n)}\}) = \left\{ u : u = \sum_{i=1}^n c_i v^{(i)}, c_i \in \mathbb{R} \right\}$$

$$Ax = \sum_i x_i A_{:,i}$$

↳ if b is in the span of the columns of A then $Ax = b$ has a solution. → COLUMN SPACE RANGE of A

↳ to have solutions for all values of $b \in \mathbb{R}^m$, then the column space of A has to be all of \mathbb{R}^m .

↳ A must have at least m columns, i.e. $n \geq m$ else, the dimensionality of the column space would be less than m .

↳ only a necessary condition for every b to have a solution

↳ not sufficient condition

LINEAR DEPENDENCE - one of the vectors in a set can be represented by some combination of other vectors in the set
 ↳ decreases the column space

$$v^{(1)} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}; v^{(2)} = \begin{bmatrix} 6 \\ 3 \end{bmatrix} \rightarrow v^{(1)} + 2 \cdot v^{(1)}$$

Set of vectors $\{v^{(1)}, v^{(2)}, \dots, v^{(n)}\}$ are linearly dependent iff

$c_1 v^{(1)} + c_2 v^{(2)} + \dots + c_n v^{(n)} = 0$
 for some c_i 's where at least one is nonzero.

! If each vector adds another dimension to the span then they are linearly dependent

SINGULAR

→ a matrix with lin. dependent columns

LINEAR INDEPENDENCE

- L no vector in the set is a linear combination of other vectors
- L A contains at least one set of exactly m linearly independent columns
→ the span or column space of A is \mathbb{R}^m
- L **COLUMN RANK** of A $A \in \mathbb{R}^{m \times n}$
 - ▷ dimension of the column space of A
 - ▷ size of the largest number of columns of A that form a linearly independent set.
- L column rank = row rank ; $\text{rank}(A)$

necessary & sufficient condition for $Ax = b$ to have solution for every b.

PROPERTIES

$$A \in \mathbb{R}^{m \times n}$$

- $\text{rank}(A) \leq \min(m, n)$
- $\text{rank}(A) = \min(m, n)$ FULL RANK
- $\text{rank}(A) = \text{rank}(A^T)$
- $\text{rank}(BA) \leq \min(\text{rank}(A), \text{rank}(B))$; $B \in \mathbb{R}^{n \times p}$
- $\text{rank}(A+B) \leq \text{rank}(A) + \text{rank}(B)$; $B \in \mathbb{R}^{m \times n}$

→ NORMS →

- L measure of vector size or length
 - ▷ norm is any function $f: \mathbb{R}^n \rightarrow \mathbb{R}$

PROPERTIES

$$x, y \in \mathbb{R}^n; \alpha \in \mathbb{R}$$

- $f(x) \geq 0$
- $f(x) = 0 \Leftrightarrow x = 0$
- $f(x+y) \leq f(x) + f(y)$ TRIANGLE INEQUALITY
- $\forall \alpha \in \mathbb{R} \quad f(\alpha x) = |\alpha| f(x)$ HOMOGENEITY

L^p norm $\|x\|_p = \left(\sum_i |x_i|^p \right)^{1/p}$

L^0 norm

- ▷ number of nonzero elements
- ✗ not OK terminology ; scaling the vector does not change the number of nonzero elements
- ✓ use L^1 instead

L^1 norm $\|x\|_1 = \sum_i |x_i|$

- ▷ grows at the same rate at all locations
- ▷ used when the difference between zero and nonzero elements is very important.

L^∞ norm (max norm)

$$\|x\|_\infty = \max_i |x_i|$$

- ▷ largest magnitude in the vector

FROBENIUS norm

$$x^T x$$

Squared L^2 norm

$$\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$$

- ▷ more convenient mathematically and computationally
- ▷ it increases very slowly near the origin

$$\|A\|_F = \sqrt{\sum_{i,j} A_{ij}^2}$$

DOT PRODUCT of two vectors

$$x^T y = \|x\|_2 \|y\|_2 \cos(\theta)$$

- ▷ where θ is the angle between x and

SPECIAL KIND OF MATRICES and VECTORS

DIAGONAL matrices

- entries only along the main diagonal
- example : Identity matrix
- multiplying by a diagonal matrix is computationally efficient
 $\text{diag}(\mathbf{v})\mathbf{x} = \mathbf{v} \odot \mathbf{x}$
- it doesn't have to be square
 ↳ no inverse

D is diagonal
 $\Leftrightarrow D_{ij} = 0 \forall i \neq j$

$\text{diag}(\mathbf{v})$ ▷ square diagonal matrix with vector \mathbf{v} entries

- $\text{diag}(\mathbf{v})^{-1}$ exists only if every diagonal entry is nonzero.

$$\text{diag}(\mathbf{v})^{-1} = \text{diag}\left[\left(\frac{1}{v_1}, \dots, \frac{1}{v_n}\right)\right]^T$$

SYMMETRIC

$$A^T = A ; A_{ij} = A_{ji}$$

$$\text{ANTI-SYMMETRIC } A = -A^T$$

if $A \in \mathbb{R}^{n \times n}$ then

↳ $A + A^T$ is symmetric

↳ $A - A^T$ is anti-symmetric

ORTHOGONAL VECTORS

- \mathbf{x} and \mathbf{y} are orthogonal to each other if

$$\mathbf{x}^T \mathbf{y} = 0$$

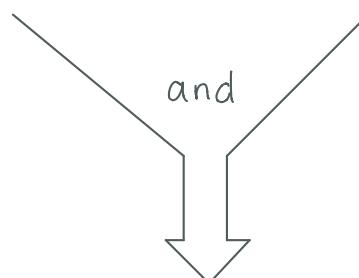
ORTHOGONAL MATRIX

- square
- rows and columns are mutually orthonormal

$$A^T A = A A^T = I$$

$$\downarrow \downarrow \\ A^{-1} = A^T$$

- inverse is cheap to compute
- $\det(A) = \pm 1$ if A is orthogonal



UNIT VECTOR

- vector with unit norm
 $\|\mathbf{x}\|_2 = 1$

ORTHONORMAL

If A is symmetric, ∃ orthogonal matrix P s.t. $P^{-1}AP = D$

- the eigenvalues of A are on the main diagonal of D .

EIGENDECOMPOSITION

Pg. 7

L shows functional properties of matrices

L decompose matrix into

only the scale

of v is altered

EIGENVECTORS \Rightarrow vectors that remain on their span after the transformation

& \Rightarrow axis of rotation

EIGENVALUES \Rightarrow the factor by which vectors are scaled (stretch or squash).

\Rightarrow each eigenvector has an associated eigenvalue

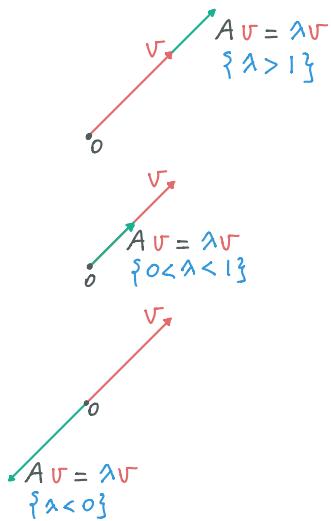
$$A v = \lambda v$$

A : matrix of transformation; A is square

v : eigenvector $\xrightarrow{\text{---}} s v$, $s \in \mathbb{R}$, $s \neq 0 \Rightarrow$ also eigenvector, with the same eigenvalue

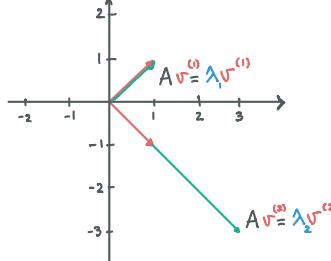
λ : eigenvalue

Scaling space:



$$v^T A = \lambda v^T \quad \text{LEFT EIGENVECTOR}$$

Example: $A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$



Find eigenvalues:

$$\det(A - \lambda I) = \begin{vmatrix} 2-\lambda & -1 \\ -1 & 2-\lambda \end{vmatrix} = (2-\lambda)(2-\lambda) - 1 = 4 - 4\lambda + \lambda^2 - 1 = (\lambda-1)(\lambda-3)$$

$$\Rightarrow \lambda_1 = 1 \text{ and } \lambda_2 = 3$$

Find eigenvectors:

$$\lambda_1 = 1 \quad (A - \lambda_1 I)v = 0$$

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \left. \begin{array}{l} v_1 - v_2 = 0 \\ v_1 - v_2 = 0 \end{array} \right\} \Rightarrow v = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\lambda_2 = 3 \quad (A - \lambda_2 I)v = 0$$

$$\begin{bmatrix} -1 & -1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \left. \begin{array}{l} -v_1 - v_2 = 0 \\ -v_1 - v_2 = 0 \end{array} \right\} \Rightarrow v = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

EIGENDECOMPOSITION

$\{v^{(1)}, v^{(2)}, \dots, v^{(n)}\}$ linearly indep. eigenvectors

$\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ eigenvalues

$$V = [v^{(1)} \ v^{(2)} \ \dots \ v^{(n)}]$$

$$\lambda = [\lambda_1 \ \lambda_2 \ \dots \ \lambda_n]^T$$

$$A = V \text{diag}(\lambda) V^{-1}$$

- ! not every matrix can be decomposed
- ! decomposition can contain complex numbers
- ! every $A \in \mathbb{R}^{n \times n}$ can be decomposed as

$$A = Q \Lambda Q^T$$

where Q is orthogonal composed of eigenvectors of A ; Λ is diagonal with eigenvalues $\Lambda_{ii} \leftrightarrow Q_{:,i}$ in descending order

► A is scaling the space by λ_i in the direction of $v^{(i)}$.

! eigen decomposition of $A \in \mathbb{R}^{m \times m}$ always exists but may not be unique.

symmetric

{iff} any are zero \rightarrow SINGULAR

λ 's $\begin{cases} \text{all positive} & \rightarrow \text{POSITIVE DEFINITE}; \text{ guarantee } x^T A x = 0 \Rightarrow x = 0 \\ \text{all positive or zero} & \rightarrow \text{POSITIVE SEMI DEFINITE}; \text{ guarantee } \forall x, x^T A x \geq 0 \\ \text{all negative} & \rightarrow \text{NEGATIVE DEFINITE} \\ \text{all negative or zero} & \rightarrow \text{NEGATIVE SEMI DEFINITE} \end{cases}$

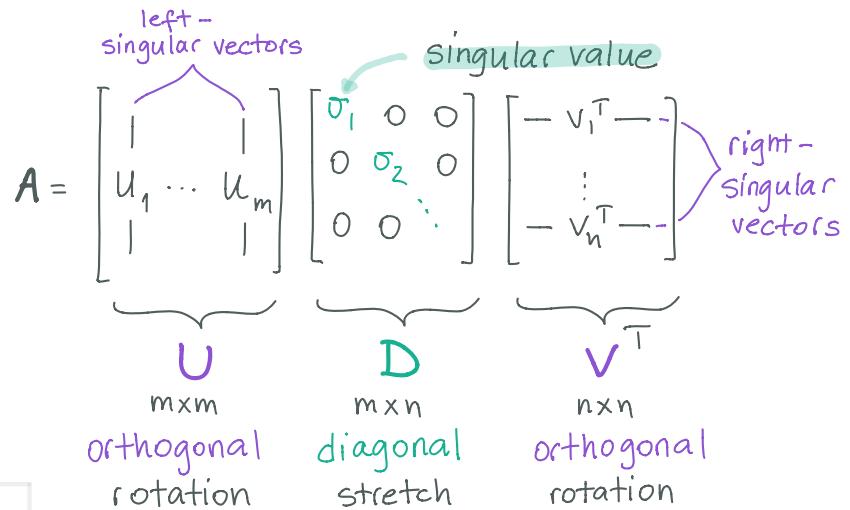
SINGULAR VALUE DECOMPOSITION (SVD)

- L factorize matrix into singular value and singular vectors
- L every real matrix has SVD.
- L used quite a lot in ML applications \Rightarrow e.g. PCA
- L getting around the square matrix requirement but with not as many nice properties as eigen decomposition.
- ! MOST USEFUL: partially generalize matrix inversion for nonsquare matrices

$$A = UDV^T$$

$$\sigma_i = \sqrt{\lambda_i} \quad \text{where } \sigma_i > 0$$

$$\{\lambda_i\} = \text{eigenvalues of } (AA^T) \\ = \text{eigenvalues of } (A^TA)$$



σ_i refers to the strength of A on the i^{th} subspace
 how much distortion can occur under transformation by A .

biggest σ value
 \Downarrow
 most information

$\{u_1, \dots, u_m\}$ \rightarrow left-singular vector
 eigenvectors of AA^T

$\{v_1, \dots, v_m\}$ \rightarrow right-singular vector
 eigenvectors of A^TA

THE MOORE-PENROSE PSEUDOINVERSE

! non square matrices don't have inverse defined.

$\dashrightarrow A = \boxed{}$ } possibility of no solution

$\dashrightarrow A = \boxed{}$ } multiple solutions possible

PSEUDOINVERSE

$$A^+ = V D^+ U^T$$

\dashrightarrow gives x for which Ax is as close as possible to y in terms of $\|Ax - y\|_2$.

\dashrightarrow provides one of the many solutions
 $x = A^+y$ with $\|x\|_2$ among all possible solutions

where U, D, V are SVD of A .

D^+ - pseudoinverse of D

- take reciprocal nonzero elements
- and take transpose

DETERMINANT

$\det : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$

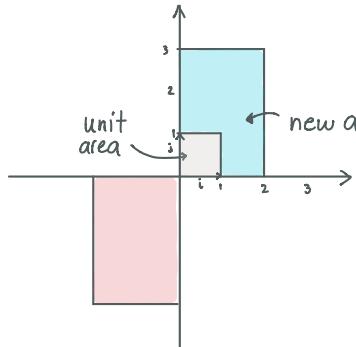
$\mathbb{N} |\mathbf{A}| ; \det(\mathbf{A})$

L exists for every square matrix

L function that maps matrices to real scalars

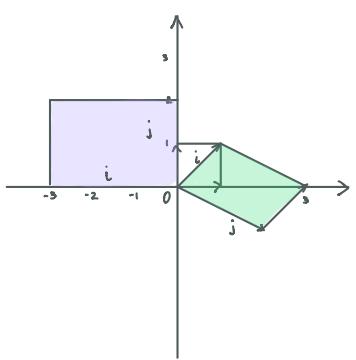
L equal to the products of all the eigenvalues of a matrix

How much area or volume gets scaled?



$$\textcolor{teal}{A} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \Rightarrow \det(A) = 6$$

$$\textcolor{pink}{A} = \begin{bmatrix} 2 & 0 \\ 0 & -3 \end{bmatrix} \Rightarrow \det(A) = -6$$



$$\left. \begin{array}{l} \textcolor{teal}{A} = \begin{bmatrix} 1 & 2 \\ 1 & -1 \end{bmatrix} \Rightarrow \det(A) = -3 \\ \textcolor{pink}{A} = \begin{bmatrix} -3 & 0 \\ 0 & 2 \end{bmatrix} \Rightarrow \det(A) = -6 \end{array} \right\} \det(A) < 0 \quad \begin{array}{l} \text{when orientation} \\ \text{space is inverted} \end{array}$$

$j \rightarrow$
 $i \leftarrow$

$\det(A) = 0 \Rightarrow \text{plane} \Rightarrow \text{line} \Rightarrow \text{point}$
columns of the matrix are lin. dependent.

PROPERTIES

$$\det(A^T) = \det(A) \quad \forall n \times n \text{ matrices}$$

$\det(A) \neq 0 \Leftrightarrow A$ is nonsingular; or

$\det(A) = 0 \Leftrightarrow A$ is singular

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

$$\det \begin{pmatrix} A & B \\ 0 & D \end{pmatrix} = \det(A)\det(D) \quad \text{if } A, D \text{ are square}$$

HOW TO COMPUTE?

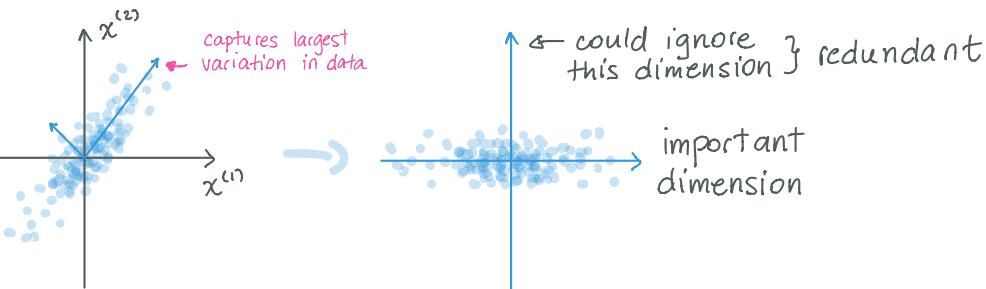
$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc \quad \text{how much is the area squished or stretched.}$$

$$\det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = \det \begin{pmatrix} a & b & c \\ \boxed{d} & \boxed{e} & \boxed{f} \\ \boxed{g} & \boxed{h} & \boxed{i} \end{pmatrix} = a \det \begin{pmatrix} e & f \\ h & i \end{pmatrix} - b \det \begin{pmatrix} d & f \\ g & i \end{pmatrix} + c \det \begin{pmatrix} e & f \\ h & i \end{pmatrix}$$

PRINCIPAL COMPONENTS ANALYSIS (PCA)

Pg. 10

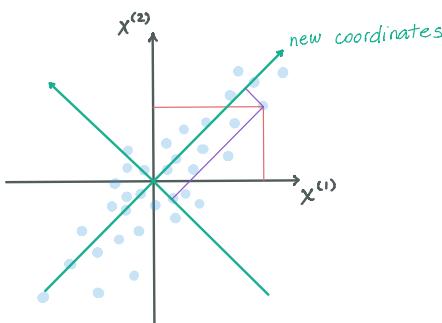
- L unsupervised linear transformation
 - data compression ; dimension reduction
remove unnecessary information



- ! largest variance doesn't lie along the basis of data measurement

- Rotate the data to the new coordinate system
- removes correlation
- orthonormal transformation

- ! Assumptions :
- linearity
→ non-linear features are transformed | centred
 - Principal components are orthogonal.



- points are represented by their projected values on to the original coordinates
- new pair of values since the coordinate system changed.

How to find PCA subspace?

$\mathbf{x}^{(i)} \in \mathbb{R}^n$; need to find l -dimensional subspace ($l < n$) that would represent the data well \Rightarrow projecting original vectors onto l dimensions.

- need a function that would transform \mathbf{x} into \mathbf{c} and another function that would reconstruct an approximation of \mathbf{x}

- since non-informative dimensions are not used (reduced to zero) \mathbf{x} can be only approximated. To capture what is lost we compute the reconstruction error

$$\mathbf{c}^* = \arg \min_{\mathbf{c}} \|\mathbf{x} - g(\mathbf{c})\|_2$$

- ! minimize since we don't want to lose too much information

dimensions ↓
information ~

$$\mathbf{x} \approx g(\mathbf{c})$$

encoding function	$f(\mathbf{x}) = \mathbf{c}$
decoding function	$\mathbf{x} \approx g(f(\mathbf{x}))$
$\mathbf{x}^{(i)} \in \mathbb{R}^n$; $\mathbf{c}^{(i)} \in \mathbb{R}^l$	s.t. $l < n$

$$g(\mathbf{c}) = \mathbf{D}\mathbf{c} ; \mathbf{D} \in \mathbb{R}^{n \times l}$$

matrix of the decoding

- orthogonal
- columns have unit norm

- PCA is defined by the choice of the decoding function.

Note: maximizing variance of the components is the same as minimizing the reconstruction error

STEP 1 How do we find the optimal c^* for every input point x ?

$$c^* = \arg \min_c \|x - g(c)\|_2$$

$$\arg \min_c \|x - g(c)\|_2^2 \quad \text{squared L}^2 \text{ norm and L}^2 \text{ norm minimize the same c}$$

$$\arg \min_c ((x - g(c))^\top (x - g(c)))$$

$$\arg \min_c ((x^\top g(c))^\top (x - g(c)))$$

$$\arg \min_c (x^\top x - x^\top g(c) - g(c)^\top x + g(c)^\top g(c))$$

$$\arg \min_c (x^\top x - 2x^\top g(c) + g(c)^\top g(c))$$

$$\arg \min_c (-2x^\top g(c) + g(c)^\top g(c))$$

$$\arg \min_c (-2x^\top Dc + (Dc)^\top Dc)$$

$$\arg \min_c (-2x^\top Dc + c^\top D^\top Dc)$$

$$\arg \min_c (-2x^\top Dc + c^\top I_n c)$$

$$\arg \min_c (-2x^\top Dc + c^\top c)$$

$$(x+y)^\top = x^\top + y^\top$$

$$x^\top y = y^\top x$$

$x^\top x$ doesn't depend on c

$$g(c) = Dc$$

$$D^\top c^\top = c^\top D^\top$$

orthogonality; unit norm constraint on D

$$\nabla_c (-2x^\top Dc + c^\top c) = 0$$

gradient descent

$$-2x^\top D^\top + 2c = 0$$

$$c = D^\top x$$



$$f(x) = D^\top x$$

ENCODING



$$r(x) = g(f(x)) = g(c) = Dc = DD^\top x$$

DECODING

STEP 2 How to choose matrix D ?

Note:

↳ Use matrix D to decode all points

↳ minimize the Frobenius norm of the matrix errors computed over all dimensions and all points

$$D^* = \arg \min_D \sqrt{\sum_{i,j} (x_j^{(i)} - r(x^{(i)})_j)^2} \quad \text{subject to } D^T D = I$$

let $l=1 \Rightarrow$ 1st principal component↳ D is a single vector d

$$d^* = \arg \min_d \sum_i \|x^{(i)} - dd^T x^{(i)}\|_2^2 \quad \text{subject to } \|d\|_2 = 1$$

$$= \arg \min_d \sum_i \|x^{(i)T} - x^{(i)T} dd^T\|_2^2 \quad \text{subject to } \|d\|_2 = 1 \quad \text{scalar rearrangement}$$

$$= \arg \min_d \sum_i \|X - Xdd^T\|_F^2 \quad \text{subject to } d^T d = 1 \quad X_{i,:} = x^{(i)T}$$

$$= \arg \min_d \text{Tr}((X - Xdd^T)^T(X - Xdd^T)) \quad \|A\|_F = \sqrt{\text{Tr}(AA^T)}$$

$$= \arg \min_d \text{Tr}(X^T X - X^T X dd^T - dd^T X^T X + dd^T X^T X dd^T)$$

$$= \arg \min_d \text{Tr}(X^T X) - \text{Tr}(X^T X dd^T) - \text{Tr}(dd^T X^T X) + \text{Tr}(dd^T X^T X dd^T)$$

$$= \arg \min_d - \text{Tr}(X^T X dd^T) - \text{Tr}(dd^T X^T X) + \text{Tr}(dd^T X^T X dd^T)$$

$$= \arg \min_d - 2\text{Tr}(X^T X dd^T) + \text{Tr}(dd^T X^T X dd^T)$$

$$= \arg \min_d - 2\text{Tr}(X^T X dd^T) + \text{Tr}(X^T X dd^T dd^T)$$

$$= \arg \min_d - \text{Tr}(X^T X dd^T) \quad \text{subject to } d^T d = 1 \quad d^T d = 1$$

$$= \arg \max_d \text{Tr}(X^T X dd^T) \quad \text{subject to } d^T d = 1$$

$$= \arg \max_d \text{Tr}(d X^T X d^T) \quad \text{subject to } d^T d = 1 \quad \text{cycling property}$$

Maximum is obtained by computing eigenvectors of $X^T X$

Covariance matrix

{if data are centered}

↳ deduct the mean

}

[] covariance

↳ variance - should be large

redundancy

↳ variance - should be small