Plan:

Basic Matrix Operations
Vector Spaces
Solving Linear Systems
Orthogonality, Projections and Determinants
SVD and Low Rank Approximation
Eigenvalues and Eigenvectors
Convexity and Quadratic Programming
Gradient Descent
Zero Sum Games and Linear Programming

Basic Operations

Let $u = [u_1, u_2, \dots, u_n]^T$, $v = [v_1, \dots, v_n]^T$

- **Dot product** $u \cdot v$

$= u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$

$= u^T v = v^T u$

- **Vector magnitude**

$|u|^2 = u_1^2 + u_2^2 + \cdots + u_n^2$

$= u^T u = u \cdot u$

- **Tranpose** $(AB)^T = B^T A^T$

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

- **Inverse** If A, B are invertible

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

Exercise: express $\|Ax - b\|^2$ in a matrix product form

$$\|Ax-b\|^2 = (Ax-b)^T (Ax-b)$$

$$= ((Ax)^T - b^T)(Ax-b)$$

$$= (x^T A^T - b^T)(Ax-b) = x^T A^T A x - b^T A x - (Ax)^T b + b^T b$$

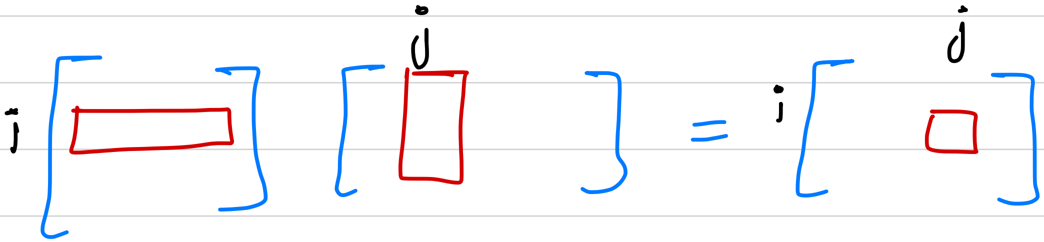
$$= x^T A^T A x - 2 b^T A x + b^T b$$

Matrix Operations

$$\underset{m\times n}{A} = \begin{bmatrix} \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \end{bmatrix}, \quad \underset{n\times p}{B} = \begin{bmatrix} \text{---} \\ \text{---} \\ \text{---} \end{bmatrix}$$

① Entry-wise: $C_{ik} = \sum_{j=1}^{n} A_{i,j} B_{j,k}$ for all $1 \leq i \leq m$, $1 \leq k \leq p$

② Inner Product: C_{ik} = dot product of row i in A, column j in B



③ Column-wise product: $AB = A \begin{bmatrix} | & | & & | \\ B_1 & B_2 & \cdots & B_k \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ AB_1 & AB_2 & \cdots & AB_k \\ | & | & & | \end{bmatrix}$

Exercise: Write $\begin{cases} Ax_1 = b_1 \\ Ax_2 = b_2 \\ Ax_3 = b_3 \end{cases}$ in one matrix matrix product eq.

$$A \begin{bmatrix} | & | & | \\ x_1 & x_2 & x_3 \\ | & | & | \end{bmatrix} = \begin{bmatrix} | & | & | \\ b_1 & b_2 & b_3 \\ | & | & | \end{bmatrix}$$

④ Outer Product

$$\underset{m\times n}{A} = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix}, \quad \underset{n\times p}{B} = \begin{bmatrix} \text{—} & b_1^T & \text{—} \\ \text{—} & b_2^T & \text{—} \\ & \vdots & \\ \text{—} & b_n^T & \text{—} \end{bmatrix}$$

$$\underset{m\times p}{AB} = \underbrace{\underset{m\times 1}{a_1}\ \underset{1\times p}{b_1^T}}_{m\times p} + a_2 b_2^T + \cdots + a_n b_n^T = \sum_{i=1}^{n} a_i b_i^T$$

Example:

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

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

$$= \begin{bmatrix} 1 & -1 \\ 2 & -2 \end{bmatrix} + \begin{bmatrix} 0 & -3 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -4 \\ 1 & -1 \end{bmatrix}$$

④ Outer Product

$$\underset{m\times n}{A} = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix}, \quad \underset{n\times p}{B} = \begin{bmatrix} — b_1^T — \\ — b_2^T — \\ \vdots \\ — b_n^T — \end{bmatrix}$$

$$AB = a_1 b_1^T + a_2 b_2^T + \cdots + a_n b_n^T = \sum_{i=1}^{n} a_i b_i^T$$

The outer product has been ueful in many application!

Example: projection matrix on space spanned by orthanormal vectors $\{v_1, v_2, \ldots, v_n\}$

$$P = v_1 v_1^T + v_2 v_2^T + \cdots + v_n v_n^T \leftarrow \text{We will get to this later today}$$

$$= \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} — v_1^T — \\ — v_2^T — \\ \vdots \\ — v_n^T — \end{bmatrix}$$

Example: SVD forms

Vector Spaces

V is a **vector space** (over reals) if:

① V is closed under scaling: $v \in V,\ \alpha \in \mathbb{R} \Rightarrow \alpha v \in V$

② V is closed under addition: $v, w \in V \Rightarrow v + w \in V$

Examples: $\text{span}\left\{ \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \end{bmatrix} \right\}$, $C(A)$, $N(A)$ and many others $\begin{bmatrix} x \\ -x \end{bmatrix}$

Column Space

$$\underset{m \times n}{A} = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix}$$

$$\Rightarrow C(A) = \text{span}\{a_1, a_2, \ldots, a_n\}$$

$$C(A) = \{Ax \mid x \in \mathbb{R}^n\}$$

$(Ax = b \Rightarrow b \in C(A))$

$$Ax = \begin{bmatrix} | \\ a_1 \\ | \end{bmatrix} x_1 + \begin{bmatrix} | \\ a_2 \\ | \end{bmatrix} x_2 + \cdots + \begin{bmatrix} | \\ a_n \\ | \end{bmatrix} x_n$$

Null Space

$$N(A) = \{x \mid Ax = 0\}$$

Linear Independance: $\{v_1, v_2, \dots, v_n\}$ are linearly independent if any linear that results in a zero must be trivial

$$\lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n = 0 \iff \lambda_1 = \lambda_2 = \dots = \lambda_n = 0$$

Generators: $\{v_1, v_2, \dots, v_n\}$ are generators of V if

$$\operatorname{span}\{v_1, \dots, v_n\} = V$$

Basis: $\{v_1, \dots, v_n\}$ are basis V if $v_1, \dots, v_n$ are L.I. and $\operatorname{span}\{v_1, \dots, v_n\} = V$

$\dim V = \#$ basis

rank(A)
- $=$ max # of L.I. columns $=$ dim (column space)
- $=$ max # of L.I. rows $=$ dim (row space)
- $=$ # of pivots in the rref of A

Rank-Nullity Theorem: $A \in \mathbb{R}^{m \times n}$ $\quad \underbrace{\operatorname{rank} A}_{\#\text{ pivots}} + \underbrace{\dim N(A)}_{\#\text{ free variables}} = n$

How to show check if $v_1, v_2, \cdots, v_n$ are L.I.?

Some ideas:

① Can we easily write one of them as a linear combination of the others? $\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} = 2\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 1\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

② Is the number of vectors more than their dimension? L.D.

③ Check $A = \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{bmatrix}$. If it's invertible $\Longrightarrow$ L.I.
otherwise $\Longrightarrow$ L.D.

because $$A \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} | \\ v_1 \\ | \end{bmatrix} x_1 + \begin{bmatrix} | \\ v_2 \\ | \end{bmatrix} x_2 + \cdots + \begin{bmatrix} | \\ v_n \\ | \end{bmatrix} x_n$$

Inverse of a square matrix

$A \in \mathbb{R}^{n \times n}$ is invertible

$\iff N(A) = \{0\}$

$\iff \text{rank}(A) = n$

$\iff C(A) = \mathbb{R}^n$

$\iff$ A has linearly independent columns

$\iff$ A has n pivots in ref

$\iff \det(A) \neq 0$

Example: inverse of $\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \dfrac{1}{\underbrace{ad-bc}_{\det(A) \neq 0}} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

Solving $\underbrace{\overset{m\times n}{A}\,\underset{n\times 1}{x}}=b_{m\times 1}$

$C(A) = \{Ax \mid x \in \mathbb{R}\}$
$m\times 1$ "b"

- If b is not in the column space of $A \longrightarrow$ no solution
- If b is in $C(A)$:

If $A\underset{n\times 1}{x_p} = b$ and $A\underset{n\times 1}{x_h} = 0$ (i.e. $x_h \in \underset{n\times 1}{N(A)}$)

Then $A(\underline{x_p + x_h}) = \underline{Ax_p} + Ax_h = b + 0 = b$

Therefore $x_p + x_h$ solves $Ax = b$

In general, any solution for $Ax=b$ can be written

as $x_p + \underset{\{0\}}{x_h}$ for some $x_h \in N(A)$ $\quad x_h, x_h'$

$x_p + x_h$
$x_p + x_h'$

Solving $Ax=b$

We can use Gaussian Elemination to solve it

free variables $=\dim(N(A))$

pivots $=\dim(C(A))$

$$\begin{cases} x-2y+3z=4 \\ x+y-3z=7 \\ 3x-4y+5z=8 \end{cases} \Rightarrow \begin{array}{ccc|c} x & y & z & \\ \hline 1 & -2 & 3 & 1 \\ 1 & 1 & -3 & 7 \\ 3 & -4 & 5 & 7 \end{array} \xrightarrow{G.E} \cdots \rightarrow \begin{array}{ccc|c} x & y & z \rightarrow \text{free} & \\ \hline \boxed{1} & -2 & 3 & 1 \\ 0 & \boxed{3} & -6 & 6 \\ 0 & 0 & 0 & 0 \end{array}$$

ref

z can be

→ a free variable

$$\begin{cases} x=z+5 \\ y=2z+2 \end{cases} \Rightarrow \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} z+5 \\ 2z+2 \\ z \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ 0 \end{bmatrix} + z\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \underbrace{\begin{bmatrix} 5 \\ 2 \\ 0 \end{bmatrix}}_{x_p} + \underbrace{z\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}}_{x_n}$$

Orthogonality

- v and u are orthogonal vectors $\Longleftrightarrow$ $\langle v, u \rangle = 0$
  $v^T u = 0$

- V and W are orthogonal subspaces of $\mathbb{R}^n$ if
  $u \cdot w = 0$ for all $u \in U$ and $w \in W$

  $$\dim V + \dim W \leq n$$

- $V^\perp$ = largest orthogonal subspace to V

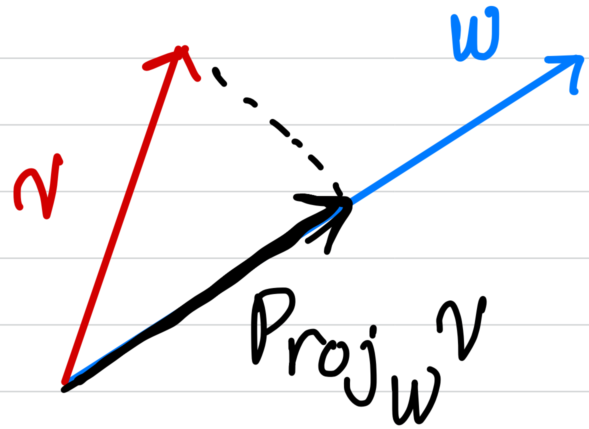
  $$\dim V + \dim V^\perp = n, \qquad (V^\perp)^\perp = V$$

Projection

Projecting a vector on another:

$$\text{Proj}_w v = \underbrace{\left(\frac{v \cdot w}{\|w\|^2}\right)}_{\text{scalar}} w$$

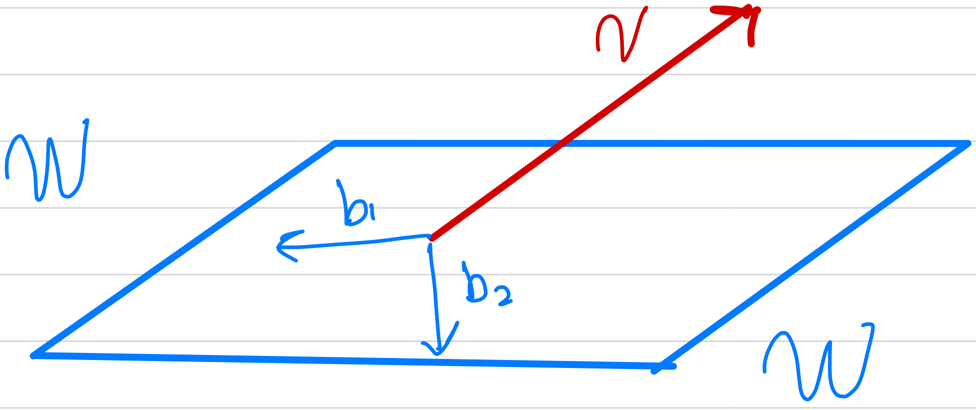
the vector w.



When w is a unit vector $\text{Proj}_w v = (v \cdot w)\, w$

Projecting a vector on a space

Assume that $b_1, b_2, \dots, b_n$ are **orthanomal** bases of the space W



$$\text{proj}_W v = \sum_{i=1}^{n} \text{proj}_{b_i} v = \sum_{i=1}^{n} (v \cdot b_i)\, b_i$$

$$= \sum_{i=1}^{n} b_i (v \cdot b_i) = \sum_{i=1}^{n} b_i (b_i^T v) = \sum_{i=1}^{n} (b_i b_i^T)\, v = P v$$

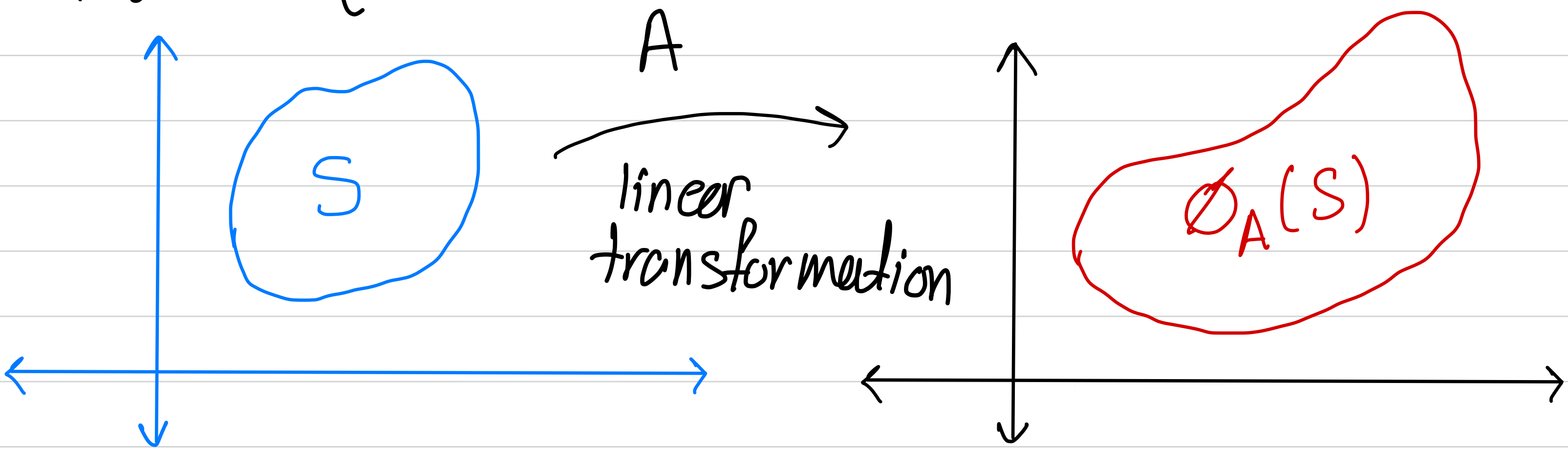
let $P = \sum_{i=1}^{n} b_i b_i^T = \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} — & b_1^T & — \\ — & b_2^T & — \\ & \vdots & \\ — & b_n^T & — \end{bmatrix}$

$$\text{proj}_W v = P v$$

Determinants

The determinant of a square matrix $\boxed{A \in \mathbb{R}^{n \times n}}$ is the scaling factor between $\mathrm{vol}(S)$ and $\mathrm{vol}(\phi_A(S))$

where $\phi(S) = \{Ax : x \in S\}$



- $\det(A^T) = \det(A)$
- $\det(AB) = \det(A)\det(B)$

Ex. If P is a projection matrix, what's $\det(P)$?

$P^2 = P \Rightarrow \det(P^2) = \det(P)$

$\Rightarrow \det(P)^2 = \det(P) \Rightarrow$ $\det(P) = 0$ or 1

Orthogonal matrices

Q is orthogonal $\Longleftrightarrow$ Q is a square matrix and $Q^TQ = I$

if Q is an orthogonal matrix then

- $\|Qv\| = \|v\|$ $\longrightarrow$ Q preserves length of v

proof: $\|Qv\|^2 = (Qv)^T Qv = v^T Q^T Q v$
$= v^T I v = v^T v = \|v\|^2$

- $\langle Qv, Qu \rangle = \langle v, u \rangle$ $\longrightarrow$ Q preserves cos of the angle between u, v

Singular Value Decomposition (SVD)

$$\underset{m\times n}{A} = \underset{m\times m}{U} \overset{m\times n}{\Sigma} \underset{n\times n}{V^T}$$

orthogonal matrix

orthogonal matrix

diagonal matrix

$$A = \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_m \\ | & | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & & \\ & \sigma_2 & & & & \\ & & \ddots & & & \\ & & & \sigma_r & & \\ & & & & 0 & \\ & & & & & \ddots \; 0 \end{bmatrix} \begin{bmatrix} \text{---} & v_1^T & \text{---} \\ \text{---} & v_2^T & \text{---} \\ & \vdots & \\ \text{---} & v_n^T & \text{---} \end{bmatrix}$$

nonzero

left singular vectors

singular values

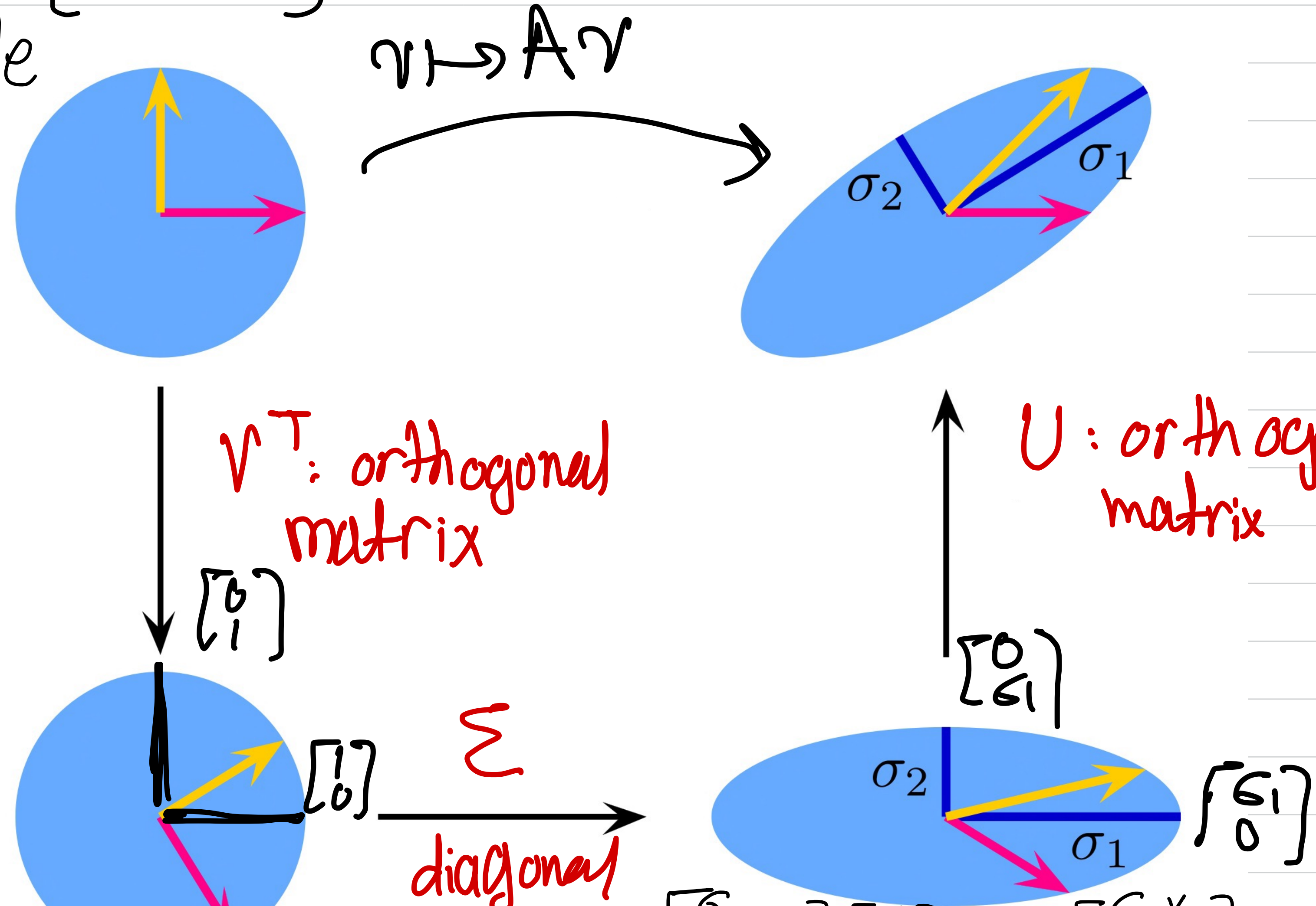
right singular vectors

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T$$

We can breakdown the linear map $Av = U\Sigma V^T v$

$$A = U \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} V^T \qquad v \longmapsto Av$$

unit circle



$$\begin{bmatrix} \sigma_1 & \\ & \sigma_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \sigma_1 x \\ \sigma_2 y \end{bmatrix}$$

The column space and nullspace of $A = U \Sigma V^T$

$$U = \underbrace{\left[\begin{array}{cccccc} | & | & & | & | & & | \\ u_1 & u_2 & \cdots & u_r & u_{r+1} & \cdots & u_m \\ | & | & & | & | & & | \end{array}\right]}_{\mathbb{R}^m}, \qquad V = \left[\begin{array}{ccccccc} | & & | & | & & | \\ v_1 & \cdots & v_r & v_{r+1} & \cdots & v_n \\ | & & | & | & & | \end{array}\right]$$

$u_1, \ldots, u_r$: span column space of A $C(A)$

$u_{r+1}, \ldots, u_m$: span $C^\perp(A)$

$v_1, \ldots, v_r$: basis for $N^\perp(A)$

$\overbrace{v_{r+1}, \ldots, v_n}^{n-r}$: span nullspace of A $N(A)$, orthonormal basis for $N(A)$

Therefore, $\dim C(A) = r$ and $\dim N(A) = n - r$

$\text{Rank}(A) = r$ = # nonzero singular values

The psudo-inverse

$$A = U \overset{\Sigma}{\begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & 0 \end{bmatrix}} V^T \Rightarrow A^+ = V \begin{bmatrix} \sigma_1^{-1} & & & & & \\ & \sigma_2^{-1} & & & & \\ & & \ddots & & & \\ & & & \sigma_r^{-1} & & \\ & & & & 0 & \\ & & & & & 0 \end{bmatrix} U^T$$

---

or $$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T$$

$$\Rightarrow A^+ = \sigma_1^{-1} v_1 u_1^T + \sigma_2^{-1} v_2 u_2^T + \cdots + \sigma_r^{-1} v_r u_r^T$$

AA^+ is the projection matrix onto $C(A)$

A^+A is the projection matrix onto $N(A)^\perp$

Matrix Norm

$$\|A\|_p = \max_{x \text{ such that } |x|=1} |Ax|$$

Properties:

1. $\|A\|_p =$ largest singular value of A
2. $|Ax| \leq \|A\|_p |x|$

<u>Exercise:</u> $\|AB\| \leq \|A\| \, \|B\|$ A, B matrices

$$\|AB\| = \max_{\|x\|=1} \|ABx\| = \|ABx^*\|$$

$$\|AB\| = |A\underbrace{Bx}_{u}| \leq \|A\| \, |u|$$

$u = Bx \Rightarrow$ $\quad |u| = |Bx| \leq \|B\| \, |x| \leq \|B\|$

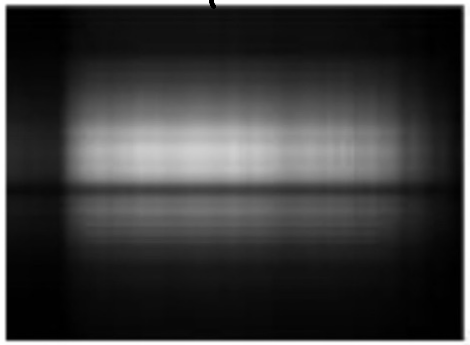
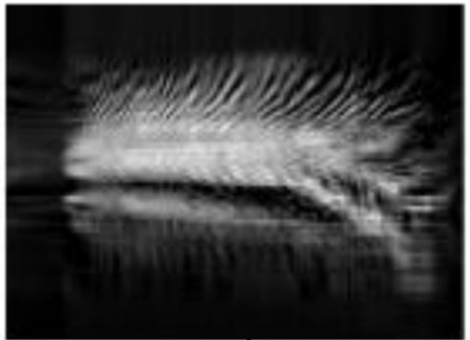
$$\Rightarrow \|AB\| \leq \|A\| \, |u| \leq \|A\| \, \|B\|$$

Low Rank Approximation

$A =$  $= \underbrace{\sigma_1 u_1 v_1^T}_{\text{rank } 1} + \underbrace{\sigma_2 u_2 v_2^T}_{\text{rank } 1} + \cdots + \underbrace{\sigma_r u_r v_r^T}_{\text{rank } 1 \text{ matrix}}$

rank 680

$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$

C_1	C_2	C_5	C_{10}
			
rank 1	rank 2	rank 5	rank 10
$C_1 \; \sigma_1 u_1 v_1^T$	$C_2 = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T$		$C_{10} = \sigma_1 u_1 v_1^T + \cdots + \sigma_{10} u_{10} v_{10}^T$

rank k approximation $C_k = \sigma_1 u_1 v_1^T + \cdots + \sigma_k u_k v_k^T$

$\|A - C_k\| = \sigma_{k+1}$

Eigenvalues and Eigenvectors

$$\underbrace{A\,v}_{\text{eigen vector}} = \overbrace{\lambda\,v}^{\text{eigenvalue}}$$

1. eigenvectors with distinct eigenvalues are linearly independent
2. trace(A) = sum of A's eigenvalues $= \sum_{i=1}^{n} \lambda_i$
3. det(A) = product of A's eigenvalues

Diagonalization

A is diagonalizable if it can be written as

$T D T^{-1}$ where D is diago

① How to find eigenvalues?

- find the characterstic polynomial $p(t) = \det(tI - A)$
- find all roots of $p(t)$

can have repeated roots (e.g. $(t-2)^3(t-1)^1(t-7)^4$)

algebriac multiplicity of 2 is 3, 1 is 1, 7 is 4

② How to find eigenvectors?

$Av = \lambda v$

$v, 2v, 3v, \ldots$

u

Solve for $(A - \lambda I)v = 0$

- find basis of the null space of $A - \lambda I$

geometric multiplicity of $\lambda = \dim N(A - \lambda I)$

Note: if λ has algebriac multiplicity of 1, then you only need to find one eigenvector for λ.

③ How to diagonalize A?

If algebraic multi. = geometric multi. for all eigenvalues, then A is diagonalizable $A = TDT^{-1}$

$$= \underbrace{\begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{bmatrix}}_{T} \underbrace{\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}}_{D} \underbrace{\begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix}^{-1}}_{T^{-1}}$$

then A is diagonalizable

ie

Note: if geometric multiplicity = algebriac multiplicity > 1 for λ then you have to choose independent eigenvectors for each eigenvalue λ

Example $A = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix} = \begin{array}{cc} v_1 & v_2 \end{array}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & \\ & 3 \end{bmatrix} \begin{array}{cc} v_1 & v_2 \end{array}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{-1}$

- Application to Linear Dynamical Systems
  - Consider a linear system $x_{k+1} = Ax_k$ where A is diagonalizable $A = TDT^{-1}$. Want to study its long-term behavior, i.e. what is x_k as $k \to \infty$?
  - Unrolling the recursion, we have

$$x_k = A^k x_0 = TD^kT^{-1}x_0 = T \begin{bmatrix} \lambda_1^k & & & \\ & \lambda_2^k & & \\ & & \ddots & \\ & & & \lambda_n^k \end{bmatrix} T^{-1}x_0$$

  - The only dependency on k is the eigenvalues $\lambda_1, \ldots, \lambda_n$.
  - If $|\lambda_i| > 1$, then $\lambda_i^k \to \infty$.
  - If $|\lambda_i| < 1$, then $\lambda_i^k \to 0$.
  - If $|\lambda_i| < 1$ for all $i = 1, \ldots, n$, then $D^k \to 0$ as $k \to \infty$, and hence $x_k \to 0$ regardless of the initial condition x_0. We say this is a **stable** system.

Positive Semi-definite Matrix

A symmatric matrix $A \in \mathbb{R}^{n \times n}$ is PSD if
square

$$x^T A x \geqslant 0 \text{ for all } x \in \mathbb{R}^n$$

Positive definite (PD) $x^T A x > 0$ for all $x \in \mathbb{R}^n$
$x \neq 0$

How to check that A is psd?

Method 1: write the expression for x^TAx if it can be expressed as sum of squares

$$\begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

$$3a^2 + 2ab + 5b^2 = (a+b)^2 + 2a^2 + 4b^2$$

Method 2: check if all eigenvalues of A are $\geqslant 0$ (non-negative)

if $\lambda_i > 0 \ \forall i \rightarrow$ PD

Method 3: Show that M is a sum of two psds

Method 4: Show that can be expressed as $M = AA^T$ for some $A \in \mathbb{R}^{n \times m}$

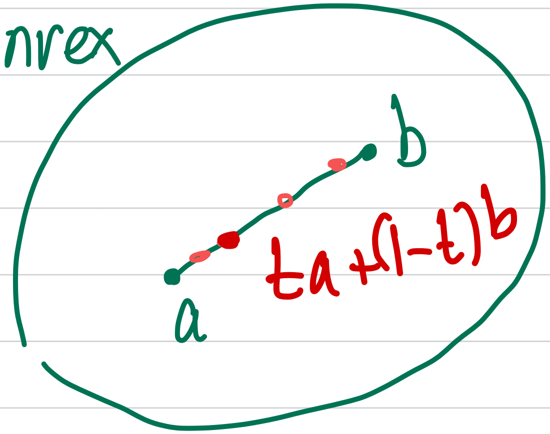
Note: for 2×2 matrices check $\text{trace}(A) \geqslant 0$ & $\det(A) \geqslant 0$

Convexity

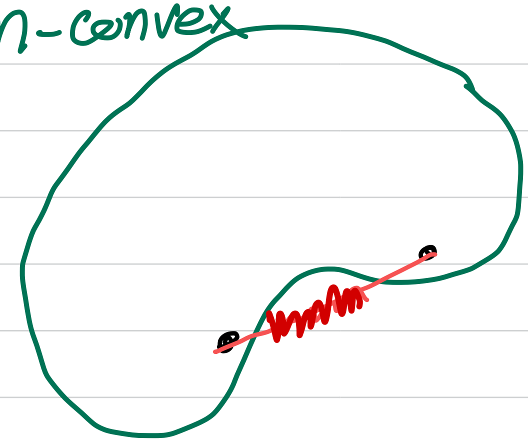
A set C is **convex** if for all a, b in C $\quad ta+(1-t)b \in C$ is in C for any $t \in [0,1]$

In other words, it contains line segment between any two points in the set

convex



non-convex



1. Intersection C_1, C_2 are convex set, then $C_1 \cap C_2$ is convex
2. Maps: If C is convex and ϕ is a $\underbrace{\text{linear}}_{Ax}$ of $\underbrace{\text{affine map}}_{Ax+b}$ then $\phi(C)$ is convex

Example: Show that $S = \{ x \mid Ax \leq b \}$ is convex

Assume that $x, y \in S$, $t \in [0,1]$

$\quad\quad\quad Ax \leq b$ and $Ay \leq b$

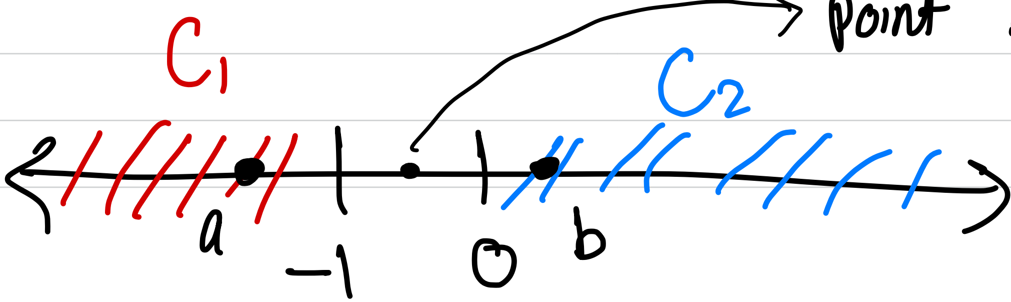
$$c = tx + (1-t)y \in S$$

$$Ac = A(tx + (1-t)y) = t\,Ax + (1-t)\,Ay$$

$$\leq t\,b + (1-t)\,b = b$$

$$\Rightarrow Ac \leq b \Rightarrow c \text{ is in } S$$

- Union is not always convex



point between a, b but not in $C_1 \cup C_2$

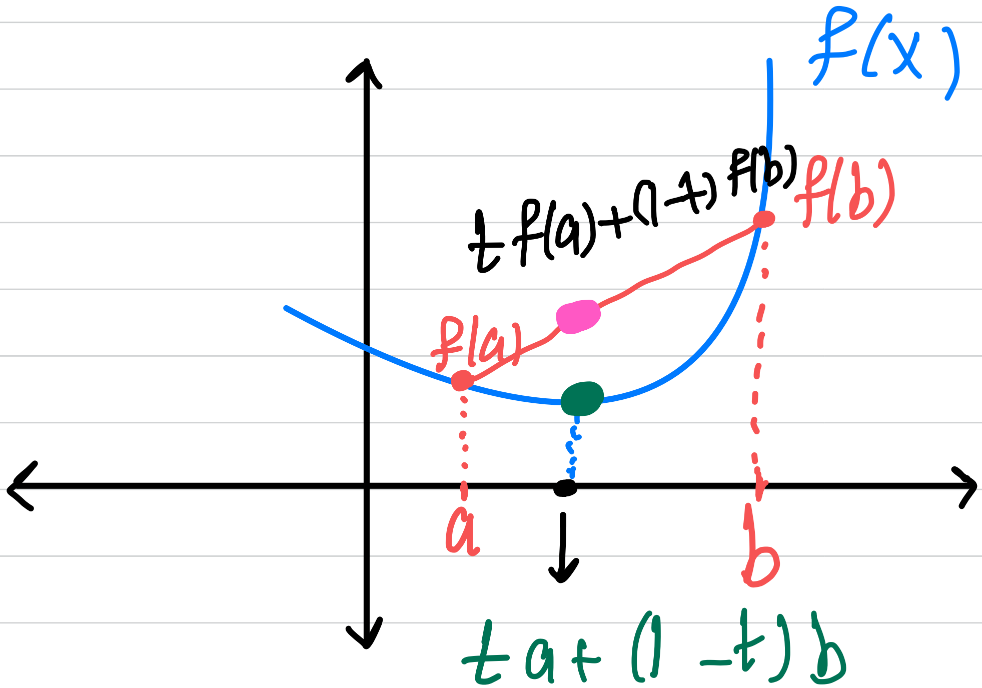
Convex Functions

$f: X \to \mathbb{R}$ is convex if for any $a, b \in X$

$$f(ta + (1-t)b) \leqslant t f(a) + (1-t) f(b)$$

for any $t \in [0,1]$

Geometrically: the graph is below the chord



1. Sum: If $f_1(x)$ and $f_2(x)$ are convex functions, then for $c_1, c_2 \geq 0$, $c_1 f_1(x) + c_2 f_2(x)$ is also a convex function.

2. Pointwise maximum: If $f_1, ..., f_n$ are convex functions, then $f(x) = \max\{f_1(x), ..., f_n(x)\}$ is also convex.

3. Linear change of coordinates: If f is a convex function, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, then $g(x) = f(Ax + b)$ is also convex.

4. Sublevel sets:

$$S_\gamma = \{x \in \mathbb{R}^n : f(x) \leq \gamma\}$$

   If $f(x)$ is convex, then S_γ is a convex set.

5. Epigraph of a function:

$$\mathrm{epi} f = \{(x, y) : x \in \mathbb{R}^n, y \in \mathbb{R} : \quad f(x) \leq y\}.$$

   If f is a convex function, then $\mathrm{epi} f$ is a convex set.

How to show that a function is convex?

Method 1: Use properties (sum & change of coord. are useful)

$g(x, y) = x^2 + (y-1)^2$

sum of convex function

$h(x,y) = (x+y)^2$

$x+y = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$

$g(x) = x^2$

$g\left(\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}\right)$

$= g(x+y)$

$= (x+y)^2$

Method 2: Take second derivative or Hessian

Ex. $f(x) = x^4 \longrightarrow f'': 4 \cdot 3 \, \underline{\underline{x^2}} \geqslant 0 \rightarrow$ convex

Method 3: Use definition

Method 2: Taking 2nd derivative

Single variable

$$f(x) = x^4$$

$$\downarrow$$

$$f''(x) = 12x^2$$

$f''(x) \geq 0$ for all x

Multivariable

$$f(x) = 2x^2 + 4xy + 4y^2$$

$$H(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}$$

If $H(x)$ is PSD then f is convex

↘ always symmatric

Short-cut for quadratic forms:

If $f(v) = v^T A v + b^T v$, then $\quad$ (A is symmetric)

$$f \text{ is convex} \iff A \text{ is PSD}$$

No Need To Find Hessian.

Why? because $H(x) = 2A$

Example: $f(x) = 2x^2 + 4xy + 4y^2 + 5x + 10$

$$= \begin{bmatrix} x & y \end{bmatrix} \underbrace{\begin{bmatrix} 2 & 2 \\ 2 & 4 \end{bmatrix}}_{A} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 5 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + 10$$

$$H(x, y) = 2 \times \begin{bmatrix} 2 & 2 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 4 & 4 \\ 4 & 8 \end{bmatrix}$$

Quadratic functions: $f(x) = x^T A x + b^T x$

$$\nabla f = 2Ax + b$$

$$H(x) = 2A$$

When is f convex? A psd

If f is convex, how to find minimum?

Method 1: Diagonalization Approach: $A = VDV^T$.

Method 2: Solve for $\nabla f = 0$

Method 3: Use gradient desent → approximate solution

Assume that f is a function with Hessian H

ex. $f(x,y) = x^2 + y + \cos(x)$

Let $\lambda_1, \lambda_2, \dots, \lambda_n$ be eigenvalues of H

$$H = \begin{array}{c|c} & j \\ \hline i & \begin{bmatrix} & & \\ & \dfrac{\partial^2 f}{\partial x_i \partial x_j} & \\ & & \end{bmatrix} \end{array}$$

convex function	non-negative eigenvalues	$0 \leq \lambda_i$ for $i = 1, \to n$
smooth & convex	upper bound	$0 \leq \lambda_i \leq L$ for some constant L
Smooth & strongly convex	upper and lower bound	$m \leq \lambda_i \leq L$ $\forall i$ where m is positive

Bounding Eigenvalues of Hessian λ_1, λ_2

__Example 1__: If $\lambda_1 = 3$, $\lambda_2 = 5$ $\Rightarrow$ $\underbrace{3}_{m} \leq \lambda_1, \lambda_2 \leq \underbrace{5}_{L}$

$\Rightarrow$ f is smooth and strongly convex

__Example 2__: If $\lambda_1 = 3 - \sin x$ where $x, y \in \mathbb{R}$
$\lambda_2 = 2 - \sin y$

$$\Rightarrow \quad 2 \leq 3 - \sin x \leq 4$$

$$1 \leq 2 - \sin y \leq 3$$

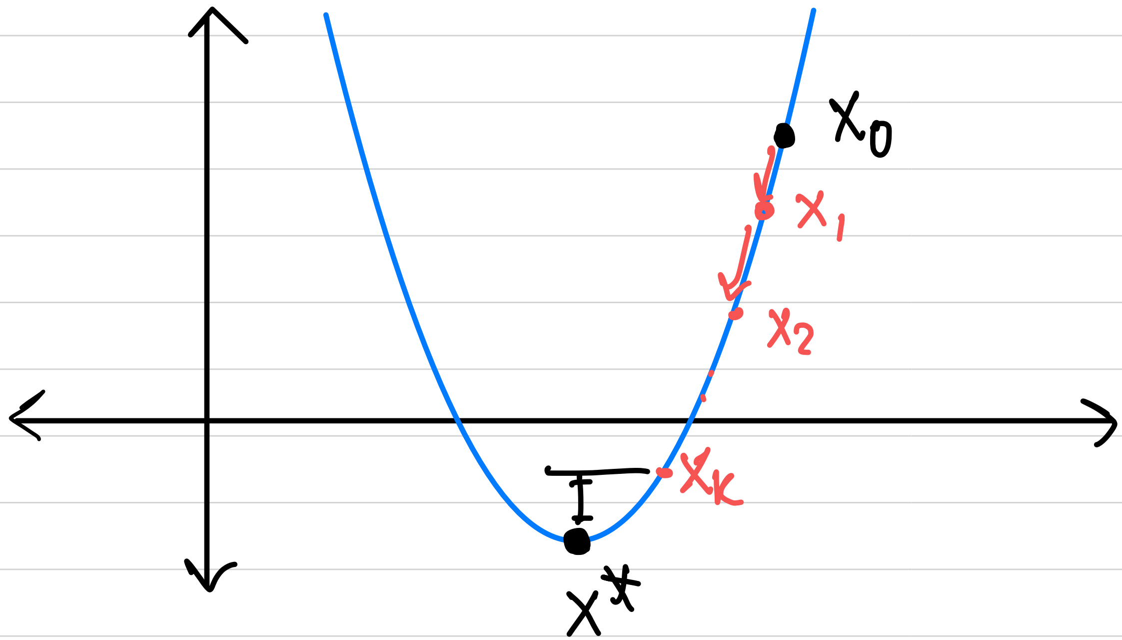
$\Rightarrow$ $\underbrace{1}_{m} \leq \lambda_1, \lambda_2 \leq \underbrace{4}_{L}$ $\Rightarrow$ f is smooth and strongly convex

Algorithm 1: Gradient Descent

Input: initial guess x_0, step size $\gamma > 0$;
while $\nabla f(x_k) \neq 0$ **do**
$\quad x_{k+1} = x_k - \gamma \nabla f(x_k)$
end
return x_k;

Goal: get closer to $f(x^*)$

$$f(x_k) - f(x^*) \leq \varepsilon$$

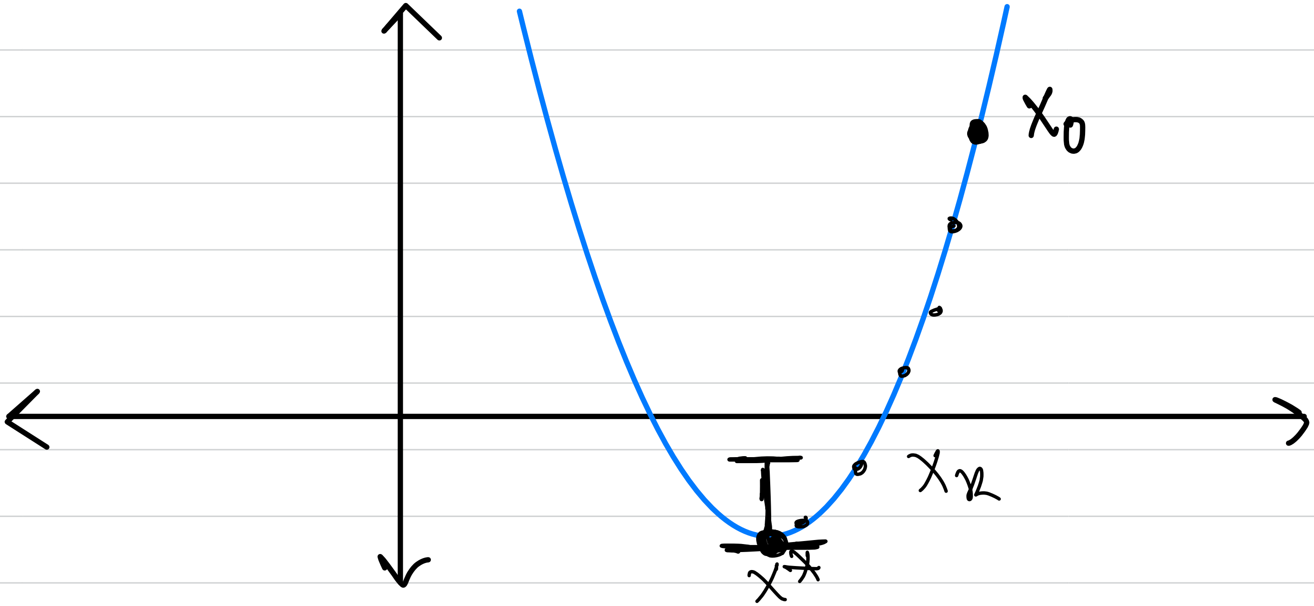


Let f is a convex function with Hessian H

Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$ be eigenvalues of H

	general convex f	quadratic $f = x^TAx + b^Tx$ (assume that $A \neq 0$)
Smooth function step size γ	$\lambda_i \leq L$ $\gamma = \frac{1}{L}$	If $0 < \gamma < \frac{2}{\lambda_1}$ $\rightarrow$ converges
Smooth & Strongly convex step size γ Condition number Q	$m \leq \lambda_i \leq L$ $\gamma = \frac{2}{m+L}$ $Q = \frac{L}{m}$	Best step size $\boxed{\gamma = \frac{2}{\lambda_1 + \lambda_n}}$ condition num $Q = \frac{\lvert\lambda_1\rvert}{\lvert\lambda_n\rvert}$ convergance rate $\lVert I - \gamma A \rVert$ $\frac{Q-1}{Q+1}$

Convergance Rate



Goal : minimize

$f(x_k) - f(x^*)$

k = # steps

smooth & convex : $f(x_k) - f(x^*) \leq \frac{L}{2} \frac{1}{k} \|x_0 - x^*\|^2$

Smooth & strongly convex : $f(x_k) - f(x^*) \leq \frac{L}{2} \left(\frac{Q-1}{Q+1}\right)^{2k} \|x_0 - x^*\|^2$

converges much faster

Logistic Regression

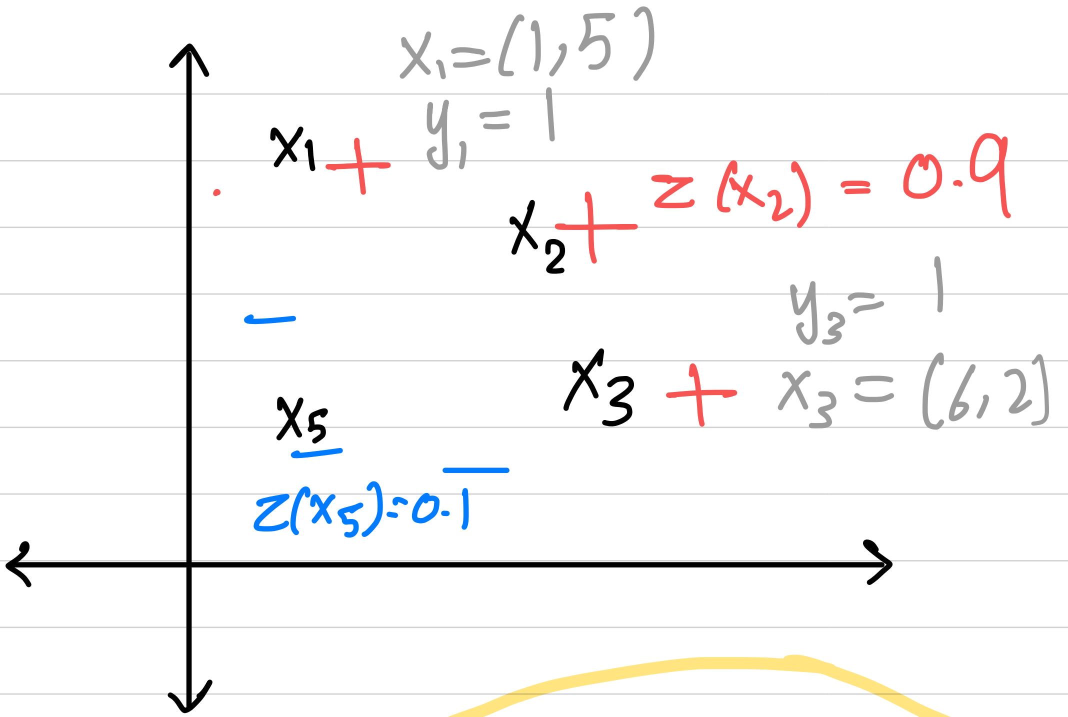
goal: classify points

x: point , y = lablel

$z = w^T x + b \rightarrow$ scalar

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

$$\begin{cases} \text{if } \sigma(z) > 0.5 \rightarrow + \\ \text{if } \sigma(z) \leqslant 0.5 \rightarrow - \end{cases}$$



want to find best w, b

$x_1, x_2, \ldots, x_n$, $y_1, \ldots, y_n$ are given data points

objective function

$$L(x, y; w, b) = -\sum_{\text{samples}} y \log(\sigma(wx+b)) + (1-y)\log(1-\sigma(wx+b))$$

use Gradient Dscent to find w^*, b^*

Linear Programs

Canonical form

Find a vector $x \in \mathbb{R}^n$ that maximizes $c^T x$ ← scalar
subject to $\begin{cases} Ax \leqslant b \\ x \geqslant 0 \end{cases}$

<u>Example:</u>

$$\max \quad x_1 + 2x_2 - x_3$$

$$\text{subject to } \begin{cases} x_1 + x_2 \leqslant 1 \\ x_1 - x_3 \leqslant 5 \end{cases}$$

$$\text{and} \quad x_1, x_2, x_3 \geqslant 0$$

Zero Sum Games:

Payoff Matrix

If Alice plays strategy i
and Bob plays strategy j

$$\Rightarrow \begin{cases} \text{Alice payoff } M_{ij} \\ \text{Bobs payoff } -M_{ij} \end{cases}$$

$$\begin{array}{cc} & \text{Bob's } j^{th} \text{ strategy} \\ \text{Alice's } i^{th} \text{ strategy} & \begin{bmatrix} & & \\ & M_{ij} & \\ & & \end{bmatrix} \end{array}$$

payoff matrix $M_{m\times n}$
for Alice

Expected Payoff

When Alice plays with probablities $p = \begin{bmatrix} p_1 \\ \vdots \\ p_m \end{bmatrix}$
Bob plays with probablities $q = \begin{bmatrix} q_1 \\ \vdots \\ q_n \end{bmatrix}$

Expected payoff for Alice is $p^T M q$.

(because $p^T M q = \sum_{i=1}^{m} \sum_{j=1}^{n} p_i q_j M_{ij}$).

1.3 Duality and Von Neumann's Theorem

Von Neumann's Teorem states that the solution to the above LP is equivalent to the solution for the following LP, which models the game if Bob were forced to reveal his strategy first:

Alice first

$$\begin{aligned} &\max \gamma \\ \text{s.t. } &p^T A \geq \gamma \vec{1} \\ &\vec{0} \leq p \leq \vec{1} \\ &(\vec{1})^T p = 1 \end{aligned}$$

$$\begin{aligned} &\min \lambda \\ \text{s.t. } &\lambda \vec{1} \geq Aq \\ &\vec{0} \leq q \leq \vec{1} \\ &(\vec{1})^T q = 1 \end{aligned}$$

Bob first

Then, the optimal objective function value for these LPs is precisely equal to the game value.

The following is a problem from recitation about Projected Gradient Decsent

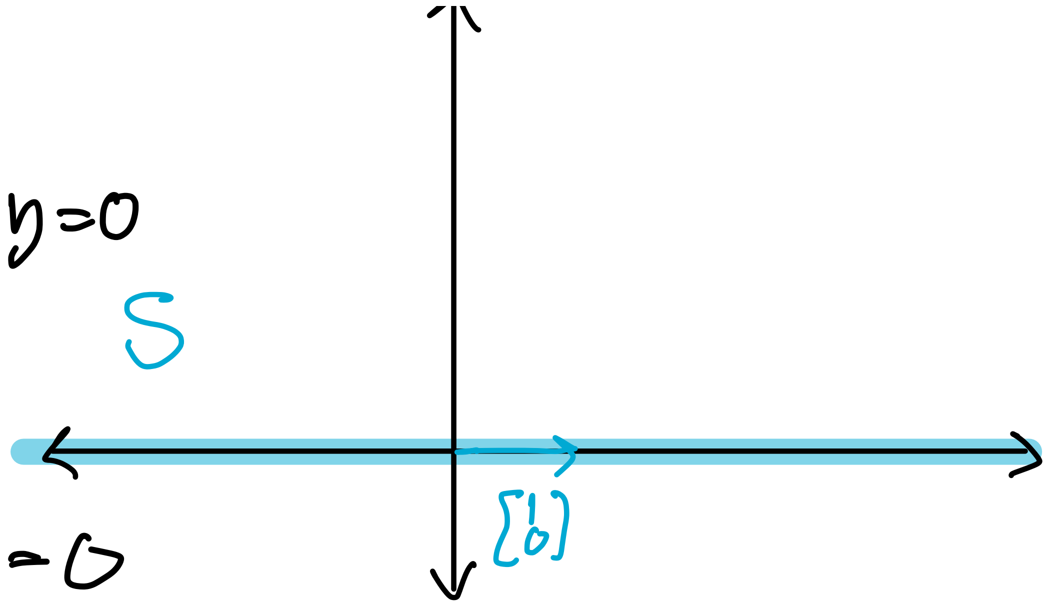
Projected Gradient Descent (Exercise)

6. (Projected Gradient Descent). In this problem we explore gradient descent with convex constraints.

   (a) Let $S = \text{Span}(\{(1, 0)\})$. Write down a matrix A such that S is the nullspace of A.

$$S = \text{Span}\left(\begin{bmatrix}1\\0\end{bmatrix}\right)$$

$\Rightarrow \begin{bmatrix}x\\y\end{bmatrix} \in S$ if and only if $y=0$



$$\begin{bmatrix}0 & 1\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix} \iff y=0$$

$\Rightarrow$ Let $\underline{A = \begin{bmatrix}1 & 0\end{bmatrix}}$

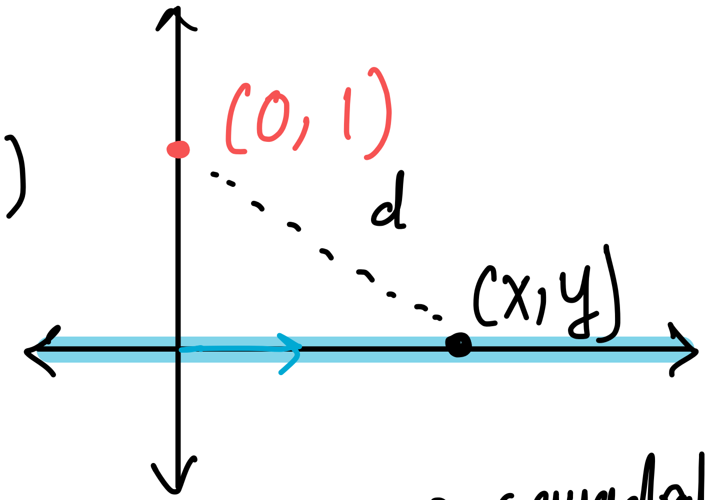
$$S = N(A)$$

Projected Gradient Descent (Exercise)

(b) Write down a quadratic optimization with linear constraints for finding the closest point in S to the point $(0,1)$. That is come up with a function $f(x,y)$ such that

$$\min_{(x,y)\in S} f(x,y)$$

Write down the Hessian of this optimization.



The square of the distance between (x,y) and $(0,1)$ is $x^2+(y-1)^2$

$\Rightarrow$ We want to minimize $f(x,y) = x^2+(y-1)^2$ (quadratic function), such that $x \in S$ or equivelantly $\begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix} = \vec{0}$ (part a)

quadratic optimization: $\min\ x^2+(y-1)^2$ such that $\begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}$

Note that the optimal x^* is $(0,0)$

next page

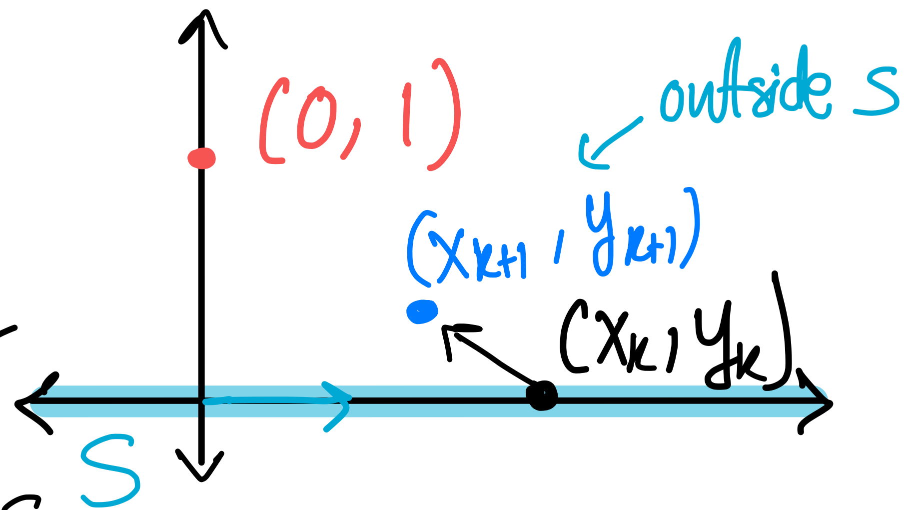
$$f(x,y) = x^2 + (y-1)^2 = x^2 + y^2 - 2y + 1$$

Gradient $\nabla f = \begin{bmatrix} 2x \\ 2y-2 \end{bmatrix}$ | Hessian : $H(x,y) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$

(c) If we move in the direction of the negative gradient $\nabla f(x, y)$ do we stay in S?

No, we don't. Assume that $(x_k, y_k) \in S$

i.e. $y_k = 0$, then



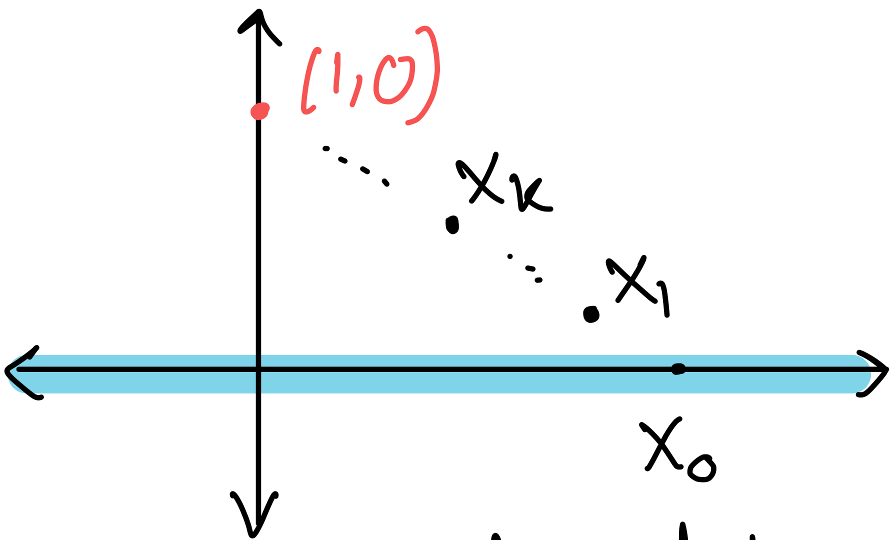
$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - \gamma \overbrace{\begin{bmatrix} 2x_k \\ 2\overset{0}{y_k} - 2 \end{bmatrix}}^{\nabla f} -2$$

$$= \begin{bmatrix} x_k - 2\gamma x_k \\ 0 - 0 + 2\gamma \end{bmatrix}$$ not in S since $y_k = 2\gamma \neq 0$.

(d) The projected gradient descent algorithm takes a step in the direction of the negative gradient $x^*_{t+1} = x_t - \gamma \nabla f(x_t)$ and then sets x_{t+1} to be the projection of x^*_{t+1} onto the subspace S. Write down the projected gradient descent update using A from part (a).

Gradient Descent

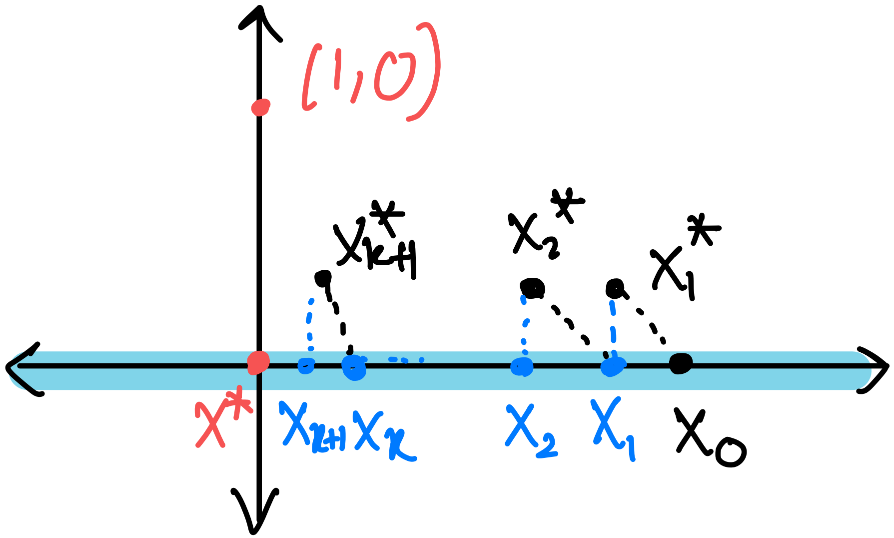
while $\nabla f(x_k) \neq 0$ **do**
| $x_{k+1} = x_k - \gamma \nabla f(x_k)$
end
return x_k;



Not constraint !

Projected Gradient Descent

while $\nabla f(x_k) \neq 0$ **do**
| $x^*_{k+1} = x_k - \gamma \nabla f(x_k)$
$x_{k+1} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} x^*_{k+1}$) projection step
return x_k;



Constraint optimization :)