OPTIMIZATION METHODS and GAME THEORY

→ want to take a decision according to some target

$$\min \{ f(x), \; x \in X \}$$

objective function

constraint set or feasible region $X \subseteq \mathbb{R}^n$ if $X = \mathbb{R}^n$ the decision is unconstrained

vector function

In general $X = \{ x \in \mathbb{R}^n : g(x) \leq 0, \; h(x) = 0 \}$ so

math programming problem

$$\begin{cases} \min f(x) \\ g(x) \leq 0 \\ h(x) = 0 \end{cases}$$

$$\max \{ f(x) : x \in X \} = -\min \{ -f(x) : x \in X \}$$

- convex function $\iff f''(x) \geq 0$ ⇝ sum of convex function are convex

WEIERSTRASS THEOREM

constraints must be continous

let $f : \mathbb{R}^n \to \mathbb{R}$ be continuous and $X \subset \mathbb{R}^n$ closed and bounded then both $\min/\max_{x \in X} f(x)$ admit a global optimal solution

PARETO MINIMUM

→ multi-objective optimization

$\bar{p} \in X$ is called pareto minimum if it's impossible to improve one function without increase the other

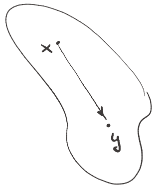
- the two systems must be impossible

$$\begin{cases} F(\bar{p}) - F(p) > 0 \\ E(\bar{p}) - E(p) \geq 0 \\ p \in X \end{cases} \wedge \begin{cases} \ldots \end{cases}$$

CONVEX SET

$x, y \in C$, then C is convex $\iff \forall \alpha \in [0,1]$

$$\alpha x + (1-\alpha) y \in C$$



- se $\forall \alpha \in \mathbb{R}$, $\alpha x + (1-\alpha) y \in C \implies$ set is AFFINE (and convex)
  - points, lines, solution sets of a system of linear equations

$$C = \{ x \in \mathbb{R}^n : A\bar{x} = b \}$$

⇝

$$\alpha x + (1-\alpha) y \in C$$
$$A[\alpha x + (1-\alpha) y] \stackrel{?}{=} b$$
$$\alpha A x + A y - \alpha A y \stackrel{?}{=} b \iff b + b - b = b$$

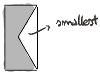
- every sub-space is a convex-set because $\alpha x + \beta y \in C$ with $\beta = 1 - \alpha$
  → the difference is that $\bar{0}$ that is in all subspaces belongs to the affine

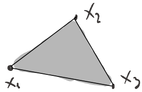
set $\iff$ $A\bar{x} = \bar{0}$ homogenious $\rightarrow$ Affine sets are translated subspaces

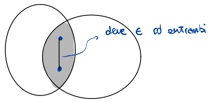
CONVEX COMBINATION of POINTS $\quad y = \sum_{i=1}^{k} \alpha_i x_i \quad$ where $x_i \in [0,1]$ and $\sum \alpha_i = 1$

↳ generalization of 2 points

$\rightarrow$ smallest convex set containing the points







1) if $\{C_i\}_{i \in I}$ is a family of convex set $\Rightarrow$ $\bigcap_{i \in I} C_i$ is convex

2) **convex hull** of a set C is the intersection of all the convex sets containing C

↳ smallest convex set containing C $\rightsquigarrow$ all convex combinations of point in C

POLYHEDRON $\rightarrow$ intersection of a finite number of closed half-spaces $\in \mathbb{R}^n$

$a^T x \leq b$ (convex set) $\quad P = \{ x \in \mathbb{R}^n : Ax \leq b \}$

BALL $\rightarrow$ $B(\bar{x}, r) = \{ z \in \mathbb{R}^n : \|z - \bar{x}\| \leq r \}$

↳ for each norm this is convex

$\|x\|_\infty = \max_i |x_i|$ (chebshev)

$\|x\|_1 = \sum |x_i|$ (manhattan)

$\|x\|_2 = \sqrt{\sum x_i^2}$ (euclidian)

$\|x\|_p = \sqrt[p]{\sum x_i^p}$

$\|x\|_A = \sqrt{x^T A x}$

↳ positive definite and symm

$x^T A x > 0 \quad \forall x$

NORM: on a vector space is a function $p: X \rightarrow \mathbb{R}$ such that:

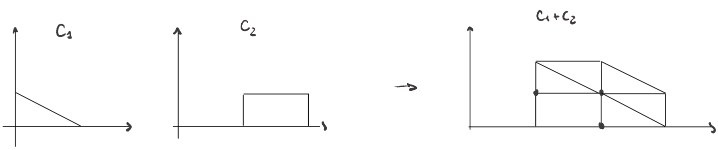
1) $p(x+y) \leq p(x) + p(y) \quad \forall x, y \in X$
2) $p(\alpha x) = |\alpha| p(x) \quad \forall x \in X, \ \forall \alpha \in \mathbb{R}$
3) $p(x) = 0 \iff x = 0$

$\rightarrow$ follows that $p(x) \geq 0 \quad \forall x \in X$

$p(x) = p(2x - x) \leq p(2x) + p(-x) = 3p(x) \quad \Rightarrow \quad p(x) \geq 0$

OPERATION that maintains CONVEXITY $\quad C_1 + C_2 = \{ x_1 + x_2 : x \in C_1, y \in C_2 \}$

also αC and $C_1 - C_2$



CLOSURE of a set C is $C \cup \{\text{limit points}\}$ $\equiv$ intersection of all closed set $\supseteq C$

INTERIOR is the union of all open set $\supseteq C$ $\equiv$ $\{x : \exists \varepsilon > 0 : B(x,\varepsilon) \subseteq C\}$

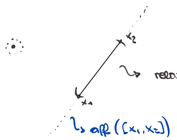
Th: If C is convex $\Rightarrow$ $\text{int}(C) \wedge \text{cl}(C) \wedge \text{ri}(C)$ are convex

$\text{aff}(C)$ = smallest affine set containing C

RELATIVE INTERIOR $\longrightarrow$ non-empty (if C is non-empty) and convex

convex

$$\text{ri}(C) = \{x \in C : \exists \varepsilon > 0 : \text{aff}(C) \cap B(x,\varepsilon) \subseteq C\}$$



↳ relative interior $\text{ri}([x_1,x_2]) = (x_1,x_2)$

SEPARATION of CONVEX SET $\longrightarrow$ if they are linearly seperable

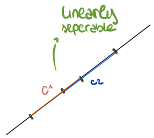
$$a^T x \geq \beta \quad \forall x \in A \quad \wedge \quad a^T x \leq \beta \quad \forall x \in B$$

8

- proper if strict inequality used
- they could have some point in common

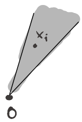
Th: A, B non empty convex sets in $\mathbb{R}^n$, then A, B are ~~properly~~ linearly seperable $\iff$ $\text{ri}(A) \cap \text{ri}(B) = \emptyset$

- Two disjoint and convex set are always properly linearly seperable



AFFINE FUNCTION $f(x) = Ax + b$ $\longrightarrow$

1) if C is convex $\Rightarrow$ $f(C) = \{f(x),\ x \in C\}$ is convex
2) if $C \subseteq \mathbb{R}^m$ is convex $\Rightarrow$ $f^{-1}(C) = \{x \in \mathbb{R}^n : f(x) \in C\}$ is convex

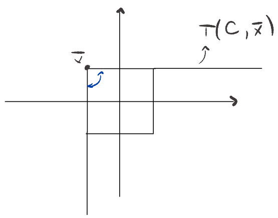


CONE A set C is a cone if $\forall x \in C,\ \forall \lambda \geq 0 \Rightarrow \lambda x \in C$

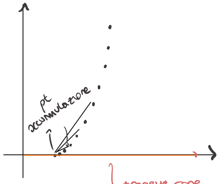
↳ must always contains zero (and the line from 0 to x_i)

- could also not be convex $\longrightarrow$ union of two lines are a cone but not convex

- **TANGENT CONE:** $T_C(\bar{x}) = \left\{ d \in \mathbb{R}^n : \exists \{z_k\} \subset C,\ \exists \{t_k\} > 0,\ z_k \to \bar{x},\ t_k \to 0,\ \lim_{k \to +\infty} \frac{z_k - \bar{x}}{t_k} = d \right\}$



tangent cone to $\bar{x}$



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

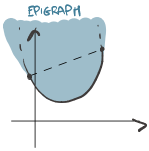
· **RECESSION CONE** given a polyhedron $P = \{\bar{x} : Ax \leq b\}$

the recession cone of P is defined as:

$\alpha Ad \leq b - Ax$
$\forall x \Rightarrow Ad \leq \bar{0}$

$$rec(P) = \{ d : x + \alpha d \in P \quad \forall x \in P, \ \alpha \geq 0 \} = \{ d : Ad \leq 0 \}$$

CONVEX FUNCTIONS let $C \subseteq \mathbb{R}^n$ be convex, then a function $f : \mathbb{R}^n \to \mathbb{R}$ is convex on C if:



$$f(\alpha y + (1-\alpha)x) \leq \alpha f(y) + (1-\alpha) f(x) \qquad \forall x,y \in C, \ \forall \alpha \in [0,1]$$

· STRICTLY CONVEX $\quad f(\alpha y + (1-\alpha)x) < \alpha f(y) + (1-\alpha) f(x) \quad \forall x,y \in C, \ x \neq y, \ \forall \alpha \in (0,1)$

↳ in any segment I don't want a linear behaviour

· STRONGLY CONVEX $\quad f(\alpha y + (1-\alpha)x) \leq \alpha f(y) + (1-\alpha) f(x) - \frac{\tau}{2}\alpha(1-\alpha)\|y-x\|_2^2$

< 0 for $\alpha \neq 0,1$

$\exists \tau > 0 :$
$\forall x,y \in C$
$\forall \alpha \in [0,1]$

↳ $\exists$ a quadratic function that under-estimate the function itself

$\to epi(f) \cap (C \times \mathbb{R})$ is convex $\Leftrightarrow$ f is convex

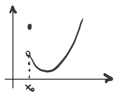
(limito il dominio)

CONCAVE FUNCTION given a convex set $C \subseteq \mathbb{R}^n$, a function $f : \mathbb{R}^n \to \mathbb{R}$ is concave if $-f$ is convex ⇝ hypograph is convex

· a linear function is both convex and concave $\quad f(x) = c^T x + b$

strongly convex $\Rightarrow$ strictly convex $\Rightarrow$ convex

Th: Let $f: \mathbb{R}^n \to \mathbb{R}$ be convex on C convex $\Rightarrow$ f is continuous on $ri(C)$
relative interior

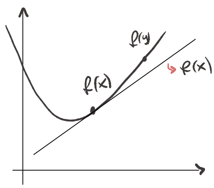


Th: f is strongly convex $\Leftrightarrow$ $\exists \tau > 0$: $f(x) - \frac{\tau}{2}\|x\|_2^2$ is convex

FIRST ORDER CONDITIONS

tutti spazio intorno

f (continuosly differentiable on C) is convex on C (open and convex) $\Leftrightarrow$ $f(y) \geq f(x) + (y-x)^T \nabla f(x)$



$\hookrightarrow f(x) + (y-x)^T \nabla f(x)$ $\to$ global under-estimator

strictly: $f(y) > f(x) + (y-x)^T \nabla f(x)$ $\quad \forall x,y \quad x \neq y$

strongly: $f(y) \geq f(x) + (y-x)^T \nabla f(x) + \frac{\tau}{2}\|y-x\|_2^2$

SECOND ORDER CONDITION

f (twice continuosly differentiable) and C (open and convex)

hessian

f is convex $\Leftrightarrow$ $\forall x \in C$, $\nabla^2 f(x)$ is positive semidefinite

$v^T \nabla^2 f(x) v \geq 0 \quad \forall v \in \mathbb{R}^n, \ \forall x \in C$

every eigenvalues are ≥ 0

could be semi-definite

- if $\nabla^2 f(x)$ is positive definite $\forall x \in C$ $\Rightarrow$ f is strictly convex on C
- f is strongly convex on C $\Leftrightarrow$ $\exists \tau > 0$: $\nabla^2 f(x) - \tau I$ is positive semidefinite $\forall x \in C$

$v^T \nabla^2 f(x) v \geq \tau \|v\|_2^2$ or $\lambda_i \geq \tau \quad \forall \lambda_i \in$ eigenvalues

CONVEXITY OF QUADRATIC FUNCTIONS

$n \times n$ symmetric matrix

$f(x) = \frac{1}{2} x^T Q x + c^T x$

is also the hessian

$\nabla f(x) = \frac{1}{2}\left(Qx + (x^T Q)^T\right) + c = Qx + c$ — depends on how Q is defined

Th:

1) If f is convex $\to$ Kf is convex
2) If f_1 and f_2 are convex $\to$ $f_1 + f_2$ is convex
3) If f is convex $\Rightarrow$ $f(Ax+b)$ is convex
4) If $f_1, \dots, f_m$ are convex $\Rightarrow$ $f(x) = \max_i \{f_i\}$ is convex
5) If $f_1, \dots, f_m$ are convex $\Rightarrow$ $f(x) = \sup_i \{f_i\}$ is convex

Ex

1) $f(x) = e^{px}$ $\quad p \in \mathbb{R} \setminus \{0\}$ when is strictly convex, but not strongly on $\mathbb{R}$

$f''(x) = p^2 e^{px}$ $\qquad p^2 e^{px} > 0 \Rightarrow p \neq 0$

$p^2 e^{px} \geq \tau \quad \forall x \in \mathbb{R}$ $\Rightarrow$ impossibile perché una volta fissato τ
$\lim_{x \to -\infty} p^2 e^{px} \to 0$

2) $f(x) = x^p \qquad x \in \mathbb{R}_+ \setminus \{0\}$

$f''(x) = p(p-1)\, x^{p-2}$ $\qquad p(p-1)x^{p-2} > 0 \iff p(p-1) > 0$ $\quad p<0 \vee p>1$

• Need to check for $p=1 \vee p=0$ but are two lines

$p(p-1)\,x^{p-2} \geq \tau \quad \forall x \in \mathbb{R}_+ \setminus \{0\} \to$ $p = 2 \Rightarrow 2 \geq \tau$ ok
otherwise problem $\leftrightarrow 0 / +\infty$

3) $f(x) = \|x\|$ $\qquad \|\alpha x + (1-\alpha) y\| \leq \|\alpha x\| + \|(1-\alpha) y\| = \alpha \|x\| + (1-\alpha)\|y\|$

$\to$ not strongly because $f(\lambda x) = \lambda \|x\|$ that is linear

COMPOSITION $\qquad f: \mathbb{R}^n \to \mathbb{R} \qquad g: \mathbb{R} \to \mathbb{R}$

$g \circ f = -x^2$

1) If f is convex (x^2) and g is convex ($-x$) and non-decreasing $\to$ $g \circ f$ is convex
2) If f is convex and g is concave and non-increasing $\to$ $g \circ f$ is concave
3) If f is concave and g is convex and non-increasing $\to$ $g \circ f$ is convex
4) If f is concave and g is concave and non-decreasing $\to$ $g \circ f$ is concave

• SUB-LEVEL SET $\quad S_K(f) = \{ x \in \mathbb{R}^n : f(x) \leq K \}$ $\qquad K \in \mathbb{R}$

If f is convex $\to$ $S_K(f)$ is convex $\forall K$

dim: $\quad x_1, x_2 \in S_K \Rightarrow f(\alpha x_1 + (1-\alpha) x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2) \leq \alpha K + (1-\alpha) K = K$

then $\quad \alpha x_1 + (1-\alpha) x_2 \in S_K$

QUASI-CONVEX FUNCTION : given C convex, f is said quasi-convex on C if $S_k(f) \cap C = \{x \in C : f(x) \leq k\}$ are convex $\forall k \in \mathbb{R}$

- $f(x) = \sqrt{|x|}$ is quasi-convex
- $\log(x)$ is quasi-convex $\wedge$ quasi-concave

OPTIMIZATION PROBLEM

objective

$$f_* = \min \{ f(x) : x \in X \}$$

constraints

$$X = \{ x \in \mathbb{R}^n : g(x) \le 0 \ \wedge \ h(x) = 0 \}$$

feasible region

P is referenced also by "mathematical programming problem"

$v(P) = \inf \{ f(x) : x \in X \}$ is called optimal value

- $v(P) \in \mathbb{R} \Rightarrow$ bounded from below
- $v(P) = -\infty \Rightarrow$ unbounded
- $v(P) = +\infty \Rightarrow$ infeasible $X = \emptyset$

GLOBAL OPTIMAL SOLUTION $x^* \in X \quad f(x^*) \le f(x) \quad \forall x \in X$

$X_* = \operatorname{argmin} \{ f(x) : x \in X \} \rightsquigarrow$ set of global minima

LOCAL OPTIMAL SOLUTION $x^* \in X : \quad f(x^*) \le f(x) \quad \forall x \in X \cap B(x^*, r)$ with $r > 0$

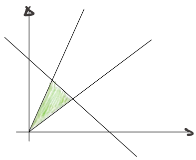
WEIERSTRASS if f is continuous and X is closed and bounded $\Rightarrow$ at least one global optimum exists

proof: $v(P) = \inf_{x \in X} f(x)$ bounded but may not converge $\{x^k\} \subseteq X : f(x^k) \to v(P)$. Since x^k is bounded $\Rightarrow$ Bolzano-Weierstrass (every bounded sequence admits a convergent subsequence) says that $\exists$ a subsequence $\{x^{k_p}\} \to x^*$. Since X is closed $\Rightarrow x^* \in X$. $f(x^{k_p}) \to f(x^*)$ since f continuous $\Rightarrow f(x^*) = v(P)$

Corollary: If f is continuous, X closed and $\exists k \in \mathbb{R}$: $S_k(f) = \{ x \in X : f(x) \le k \}$ (closed) is non empty and bounded $\Rightarrow \exists$ a global minimum

$$\begin{cases} \min \ e^{x_1 + x_2} \\ x \in X \doteq \{ x_1 - x_2 \le 0 , \ -2x_1 + x_2 \le 0 \} \end{cases}$$

closed but unbounded



but $S_2 = \{ x \in X : f(x) \le 2 \} = \begin{cases} x_1 + x_2 \le \ln 2 \\ x_1 - x_2 \le 0 \\ -2x_1 + x_2 \le 0 \end{cases}$ is closed and bounded

Corollary: if f is continous and coercive $\left(\lim_{\substack{\|x\| \to +\infty \\ x \in X}} f(x) = +\infty \right)$

and $X \neq \emptyset$ is closed $\Longrightarrow$ a global minimum exist

↳ has an end

- because I can take $\bar{x} \in X$, $K = f(\bar{x})$ $\Rightarrow$ S_K will be non-empty and bounded (because moving away from the origin the function will pass K)

Th: f convex on convex set X $\Rightarrow$ any local optimum of f is a global optimum

proof: x^* a local optimum $f(x^*) \le f(z)$ $\forall z \in X \cap B(x^*, r)$

by contraddiction: $f(y) < f(x^*)$ then take $\alpha \in [0,1]$:

$$\alpha x^* + (1-\alpha)y \in B(x^*, r) \Rightarrow$$

$$f(\alpha x^* + (1-\alpha)y) \le \alpha f(x^*) + (1-\alpha) f(y) < \alpha f(x^*) + (1-\alpha) f(x^*) = f(x^*)$$

- **Th:** f is strictly convex on convex set X and admits a global optimum x^* $\rightarrow$ x^* is the unique solution

proof: by contraddiction $\hat{x} \in X$, $\hat{x} \neq x^*$ but $f(\hat{x}) - f(x^*)$ $\rightarrow$

$$f(\alpha x^* + (1-\alpha)\hat{x}) < \alpha f(x^*) + (1-\alpha) f(\hat{x}) = f(x^*) \qquad \forall \alpha \in (0,1]$$

- **Th** f a strongly convex on $\mathbb{R}^n$ and X is closed $\rightarrow$ $\exists$ a global minimum

proof: f convex is continous $f(x) = \psi(x) + \tau \|x\|_2^2$ with $\psi(x)$ convex

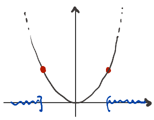
schwartz inequality

$$f(x) = \psi(x) + \tau\|x\|_2^2 \ge a^T \bar{x} + \tau \|x\|_2^2 \ge -\|a\|_2 \|x\|_2 + \tau \|x\|_2^2 \rightarrow +\infty$$

Corollary: If f strongly convex on $\mathbb{R}^n$ and X is closed and convex $\Rightarrow$ $\exists!$ global minimum

positive definite

any quadratic programming problem $\min \frac{1}{2} x^T Q x + c^T x$ $\quad x \in X$
and X closed and convex $\Rightarrow$ $\exists!$ global minimum (strictly and strongly convex)



-strictly convex function in $\mathbb{R}^n$ but X not convex $\Rightarrow$ more global minimum

Th: Assume X (es. $\mathbb{R}^n$) an open set and f be differentiable at $x^* \in X$. If x^* is the local optimum $\Rightarrow$ $\nabla f(x^*) = 0$

proof: by contradiction $\nabla f(x^*) = d$, then $f(x^* - dt) = \phi(t)$
$\varphi'(0) = -d^T \nabla f(x^*) = -\|\nabla f(x^*)\|_2^2 < 0 \Rightarrow f(x^* - td) < f(x^*)$ for small enough that is impossible

Th: X open set (→ otherwise boundary problems (non differentiable)) and $x^* \in X$ be a local optimum $\Rightarrow$
1) $\nabla f(x^*) = 0$
2) $\nabla^2 f(x^*)$ is positive semidefinite

NECESSARY CONDITION

Th: X open set and $x^* \in X$ and
1) $\nabla f(x^*) = 0$
2) $\nabla^2 f(x^*)$ is positive definite
$\Rightarrow$ x^* local minima

SUFFICIENT CONDITION

Th! f be differentiable convex on the open convex set X then $x^* \in X$ is a global minimum $\Leftrightarrow$ $\nabla f(x^*) = 0$

- elimino i punti di sella essendo convex

proof: $f(x) - f(y) \geq (x-y)^T \nabla f(y)$ with $y = x^*$ $\Rightarrow$ $f(x^*) \leq f(x)$ $\forall x \in X$
↳ linearization at y for its convexness

⇝ OTHER WAY

- se f è strictly convex $\Rightarrow$ il minimo è unico nel teorema

Consider the quadratic problem
$$\begin{cases} \min f(x) = \frac{1}{2}x^T Q x + c^T x \\ x \in \mathbb{R}^n \end{cases}$$
→ unconstrained, convex

$\Rightarrow \exists$ a global optimum x^* for P $\Leftrightarrow$
1) $Qx^* + c = 0$ ⇒ $\nabla f = 0$
2) Q is positive semidefinite ⇒ convex

- If it is positive-definite $\Rightarrow$ f is strictly convex problem $\Rightarrow \exists! x^*$

$Qx^* + c = 0 \rightarrow x^* = -Q^{-1}c$
↳ non singular ($\det Q \neq 0$), so invertible

- Strict $\Leftrightarrow$ Strongly if is a quadratic problem because λ_i doesn't depend on x

- If Q is positive-semidefinite (Q may be not invertible)

  Rouché-Capelli: the system has a solution if $\text{rank}(Q) = \text{rank}(Q|-c)$
  → could be infinity solutions

  - no solution if $Q = [0]$ but the linear problem is un-constrained

Consider the problem
$$\begin{cases} \min f(x) \\ g(x) \leq 0 \\ h(x) = 0 \end{cases}$$
is said convex if:

1) $f(x)$ is convex
2) $g(x)$ are convex ⇝ level set is convex
3) $h(x)$ are affine (both convex and concave)

⇒ X is convex

$$\begin{cases} \min \; x_1^2 + x_2^2 \\ \dfrac{x_1}{1+x_2^2} \leq 0 \\ (x_1+x_2)^2 = 0 \end{cases} \xRightarrow{\text{equivalent}} \begin{cases} \min \; x_1^2 + x_2^2 \\ x_1 \leq 0 \\ x_1 + x_2 = 0 \end{cases}$$
that is convex

- If P is convex then any stationary point is a global minimum

ALGORITHM for solving unconstrained problem

$f(x^1) > \dots > f(x^k) > \dots$

$x^{k+1} = x^k + t_k d^k$ $\quad t_k > 0 \quad d^k \in \mathbb{R}^n$

descent direction: is a vector such that $d^T \nabla f(x^k) < 0$

directional derivative

$d_k = -\nabla f(x^k)$ that is the steepest descent direction

GRADIENT METHOD

1) choose x^0, $k=0$
2) if $\nabla f(x^k) = 0$ STOP otherwise 3
3) $d^k = -\nabla f(x^k)$ $\Rightarrow$ compute an optimal solution for:

$t_k = \underset{t>0}{\operatorname{argmin}} \; f(x^k + t d^k)$ $\rightarrow$ one dimensional

t_k is the step size

$\begin{cases} x^{k+1} = x^k + t_k d^k \\ k = k+1 \end{cases}$ go to step 2

- I will come to an end if there is a stationary point

let f be continously differentiable

1) $(d^k)^T d^{k+1} = 0 \quad \forall k$ $\rightsquigarrow$ ORTHOGONAL

I want to minimize $d^T \nabla f(x^k)$

$\min f(x^k + t d^k) \Rightarrow \frac{d}{dt} f(x^k + t d^k) = 0 \Leftrightarrow$

$(d^k)^T \nabla f(x^k + t_k d^k) = 0 \Rightarrow (d^k)^T \nabla f(x^{k+1}) = 0$

converging sequence

2) If $\{x_k\} \rightarrow x^* \Rightarrow \nabla^T f(x^*) \cdot \nabla f(x^*) = 0$

$\Leftrightarrow \|\nabla f(x^*)\|^2 = 0 \Leftrightarrow \nabla f(x^*) = 0$

- f is coercive (level set is bounded) $\Rightarrow$ $\forall x^0$, $\{x^k\}$ is bounded (also $\{f(x^k)\}$) and any of its cluster points (accumulation point) is a stationary point

→ If f is convex and coercive → the stationary point is a global minimum

→ If f is strongly convex $\Rightarrow$ $\{x_k\}$ → unique minimum

QUADRATIC CASE (Q positive definite)

taylor expansion (exact because is quadratic)

$$f(x^k + td^k) = f(x^k) + (td^k)^T \nabla f(x^k) + \frac{1}{2}(td^k)^T Q (td^k)$$

$$= f(x^k) + (d^k)^T g^k t + \frac{1}{2}(d^k)^T Q d^k t^2$$

$\nabla f(x^k) = Qx^k + c$

$$\frac{d f(x^k + td^k)}{dt} = (d^k)^T Q d^k t + (d^k)^T g^k = 0 \qquad t_k = -\frac{(d^k)^T g^k}{(d^k)^T Q d^k}$$

the only variable

$$\rightarrow \quad t_k = \frac{\|g^k\|^2}{(g^k)^T Q g^k} \qquad \Rightarrow \quad x^{k+1} = x^k + \frac{\|g^k\|^2}{(g^k)^T Q g^k} g^k$$

since positive definite $\neq 0$

ERROR BOUND

f quadratic with Q positive definite, x^* global minimum $\Rightarrow$ $\{x^k\}$ satisfy:

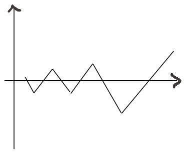
$$\|x^{k+1} - x^*\|_Q \leq \left(\frac{\frac{\lambda_n}{\lambda_1} - 1}{\frac{\lambda_n}{\lambda_1} + 1}\right) \|x^k - x^*\|_Q$$

max and minimum eigenvalue $\quad 0 \leq \lambda_1 \leq \dots \leq \lambda_n$

$\frac{\lambda_n}{\lambda_1}$ influence the rate of convergence

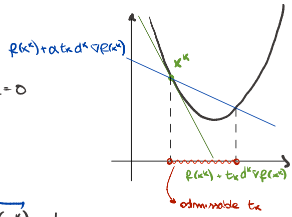
example $\quad f(x) = x_1^2 + 10x_2^2 \qquad x^0 = (10, 1) \qquad g = \begin{pmatrix} 2 \\ 20 \end{pmatrix}$

$$\|x^{k+1} - x^*\|_Q = \frac{9}{11} \|x^k - x^*\|_Q$$



When the function f is not quadratic $\Rightarrow$ the exact line search may be computationally expensive

GRADIENT METHOD with Armijo inexact line search



1) set $\alpha, \gamma \in (0,1)$ and $\bar{t} > 0$. choose $x^0 \in \mathbb{R}^n$, $k=0$
2) If $\nabla f(x^k) = 0$ STOP otherwise 3)
3) $d^k = -\nabla f(x^k)$ $\quad t_k = \bar{t}$

only variable

$$\text{while } f(x^k + t_k d^k) > f(x^k) + \alpha t_k \overbrace{(d^k)^T \nabla f(x^k)}^{\text{negative}} \text{ do :}$$
$$t_k = \gamma t_k$$

this converge

$$x^{k+1} \leftarrow x^k + t_k d^k \qquad k \leftarrow k+1 \qquad \text{then go to 2)}$$

TH: If f is coercive $\Rightarrow$ $\forall x^0$ (starting point) the $\{x^k\}$ is bounded and any accumulation point is a stationary point for f.

CONJUGATE GRADIENT METHOD

- quadratic case $\quad f(x) = \frac{1}{2}x^TQx + c^Tx$, where Q is positive definite

$$d^k = \begin{cases} -g^0 & k=0 \\ -g^k + \beta_k d^{k-1} & k \geq 1 \end{cases}$$

$\rightarrow$ require that d^k and d^{k-1} are conjugate in respect to Q

$$(d^k)^T Q\, d^{k-1} = 0$$

$$(-g^k + \beta_k d^{k-1})^T Q\, d^{k-1} = 0 \Rightarrow (-g^k)^T Q\, d^{k-1} + \beta_k (d^{k-1})^T Q\, d^{k-1} = 0$$

$$\beta_k = \frac{(g^k)^T Q\, d^{k-1}}{(d^{k-1})^T Q\, d^{k-1}}$$

- then I can perform an exact line search $\quad t_k = -\dfrac{(g^k)^T d^k}{(d^k)^T Q d^k}$

descent direction

$$(g^k)^T d^k = (g^k)^T(-g^k + \beta_k d^{k-1}) = -\|g^k\|^2 + \cancel{\beta_k (g^k)^T d^{k-1}} < 0$$

$\rightarrow 0$ because its exact line search

- for quadratic functions it requires at most n iterations

dimensions of the space

1) Alternative formula: $t_k = \frac{\|g^k\|^2}{(d^k)^T Q d^k}$ (general case)

2) Alternative formula: $\beta_k = \frac{\|g^k\|^2}{\|g^{k-1}\|^2}$ (general case)

3) If we don't find the minimum after k iteration $\{g^0, \dots, g^k\}$ are $\perp$

4) $\{d^0, \dots, d^k\}$ are conjugate wrt Q and x^k is the minimum of f on:
$x_0 + \langle d^0, \dots, d^k \rangle$

If Q has r distinct eigenvalues $\Longrightarrow$ the method find the minimum in r iterations

$$\|x^k - x^*\|_Q \leq 2 \left(\frac{\sqrt{\frac{\lambda_n}{\lambda_1}} - 1}{\sqrt{\frac{\lambda_n}{\lambda_1}} + 1} \right)^k \|x^0 - x^*\|_Q \qquad \forall k > 0$$

$k = n \Rightarrow \|x^k - x^*\| \leq 0$

$$\|x^k - x^*\|_Q \leq \frac{\lambda_{n-k+1} - \lambda_1}{\lambda_{n-k+1} + \lambda_1} \|x^0 - x^*\|_Q \qquad \forall k > 0$$

- can extend the method to other function

NEWTON METHOD want to find a stationary point $\nabla f(x) = 0$

$$\nabla f(x) \simeq \nabla f(x^k) + \nabla^2 f(x^k)(x - x^k)$$

$\rightsquigarrow$ compute the minimum of the second order approximation that will be x^{k+1}

$$\nabla^2 f(x^k)(\overbrace{x - x^k}^{d}) = -\nabla f(x^k) \quad \Longrightarrow \quad x^{k+1} - x^k = d$$

$$\{x^k\} \to x^* \quad \Rightarrow \quad \nabla f(\overset{x^*}{\cancel{x^k}}) + \nabla^2 f(\overset{x^*}{\cancel{x^k}})(\overset{x^*}{\cancel{x^{k-1}}} - \overset{x^*}{\cancel{x^k}}) = 0$$

altrimenti non converge (perché può avere zone flat con stesso valore)

Convergence Theorem: If x^* is a local minimum of f and $\nabla^2 f(x^*)$ is positive definite $\Longrightarrow \exists \delta > 0 : \forall x^0 \in B(x^*, \delta)$ the sequence $\{x^k\}$ converge to x^* and:

$$\|x^{k+1} - x^*\| \leq C \|x^k - x^*\|^2 \quad \rightsquigarrow \text{ quadratic convergence} \quad \text{for some } C > 0,\ k > 0$$

- function must be doubly differentiable
- local convergence, if we start far from x^* then it is not convergent
  $\{1, -1, 1, -1 \dots\}$

Th: If f is strongly convex $\Rightarrow$ we have global convergence because d^k is a descent direction

positive definite

$$\nabla f(x^k)^T d^k = -\nabla f(x^k)^T \left[\nabla^2 f(x^k)\right]^{-1} \nabla f(x^k) < 0$$

- Could be strictly convex with Q positive definite and coercive

- I can perform an inexact line search with d given by the Newton Method

Th: If f strongly convex $\Rightarrow$ $\forall x^0 \in \mathbb{R}^n$ the sequence $\{x^k\} \to$ global minimum. If $\alpha \in (0, \frac{1}{2})$ and $\tilde{t} = 1$ $\Rightarrow$ the convergence is quadratic

basic: $x^{k+1} \leftarrow x^k + d^k$

line search: $x^{k+1} \leftarrow x^k + t^k d^k$

CONSTRAINED optimization problem

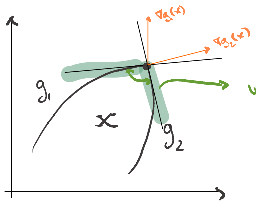
- if is unconstrained $\Rightarrow$ local minimum has $\nabla f(x^*) = 0$

  I want to find a generalization for constrained problemes

$$\begin{cases} \min f(x) \\ x \in X = \{x \in \mathbb{R}^n : g \leq 0 \ \wedge h = 0\} \end{cases}$$

f, g, h continuously differentiable

Tangent cone $T_X(x^*) = \left\{ d \in \mathbb{R}^n : \exists \{z_k\} \subset X, \ \exists \{t_k\} > 0, \ z_k \to x^*, \ t_k \to 0, \ \lim_{k \to +\infty} \frac{z_k - x^*}{t_k} = d \right\}$



with small steps I remain inside the feasible set (first order approximization of X)

optimality condition

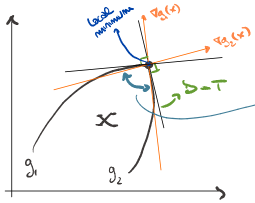
Th: If x^* is a minimum of P $\Rightarrow$ $\nabla f(x^*) d \geq 0 \quad \forall d \in T_X(x^*)$

$A(x^*) \doteq \{ j : g_j(x^*) = 0 \}$ constraints that are active on that point $\in X$

$$D(x^*) = \left\{ d \in \mathbb{R}^n : \begin{array}{ll} d^T \nabla g_j(x^*) \leq 0 & \forall j \in A(x^*) \\ d^T \nabla h_k(x^*) = 0 & \forall k \end{array} \right\}$$

first order feasible direction cone at $x^* \in X$.

ACQ - Abadie constraint qualification holds if for $x^* \in X$: $T_X(x^*) = D(x^*)$



$\nabla f(x^*)$ must be in this area otherwise $\nabla f(x^*) \cdot d < 0$

$$\Rightarrow \quad -\nabla f(x^*) = \lambda_1 \nabla g_1(x^*) + \lambda_2 \nabla g_2(x^*) \qquad \lambda_1, \lambda_2 \geq 0$$

KKT – Karush Kuhn Tucker : If $x^* \in X$ is a local minimum $\wedge$ ACQ holds at x^*

$\Rightarrow \exists \lambda^* \in \mathbb{R}^m \wedge \exists \mu^* \in \mathbb{R}^p$: (x^*, λ^*, μ^*) is a solution of :

$$\begin{cases} \nabla f + \sum^{m} \lambda_i \nabla g_i(x^*) + \sum^{p} \mu_j \nabla h_j(x^*) = 0 \\ \lambda_i \, g_i(x^*) = 0 \quad \forall i \\ \lambda^* \geq 0 \\ g(x^*) \leq 0 \\ h(x^*) = 0 \end{cases}$$

$\lambda_i \, g_i(x^*) = 0 \quad \forall i$ $\longrightarrow$ only active constraints are important

$\lambda^* \geq 0$ $\longrightarrow$ garantees the position of the $\nabla f(x^*)$

Defining the lagrangian function $\quad L(x, \lambda, \mu) \doteq f(x) + \sum^{m} \lambda_i g_i(x) + \sum^{p} \mu_j h_j(x)$

$\Rightarrow$ I can use the constraint $\nabla_x L(x, \lambda, \mu) = 0$

Sufficient conditions for ACQ

1) g_j and h_k are affine $\forall j, k$ $\Rightarrow$ ACQ holds $\forall x \in X$

2) SLATER CONDITION : g_j are convex $\forall j$ and h_k affine $\forall k$ and $\exists \bar{x} \in X$: $g_j(\bar{x}) < 0 \;\wedge\; h(\bar{x}) = 0$ $\rightarrow$ ACQ holds $\forall x \in X$

3) If $x^* \in X$ and $\{ \nabla g_j(x^*), \nabla h_k(x^*) \;\; \forall j, k \}$ are linearly independent
$\Rightarrow$ ACQ holds at x^*

- for maximum points I could use $\min \{ -f(x) \}$ or set $\lambda \leq 0$

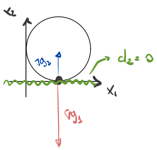
Es 1 – ACQ assumpions

$$\begin{cases} \min \; x_1 + x_2 \\ (x_1 - 1)^2 + (x_2 - 1)^2 - 1 \leq 0 \\ x_2 \leq 0 \end{cases}$$

$x^* = (1, 0)$ is the global optimum

$T_X(x^*) = \{0\}$

$D(x^*) = \left\{ d : \begin{array}{l} d^T \begin{pmatrix} 2(x_1 - 1) \\ 2(x_2 - 1) \end{pmatrix} \leq 0 \\ d^T \begin{pmatrix} 0 \\ 1 \end{pmatrix} \leq 0 \end{array} \right\} = \left\{ d \in \mathbb{R}^2 : d^T \begin{pmatrix} 0 \\ -2 \end{pmatrix} \leq 0 \right\} = \{ d \in \mathbb{R}^2 : d_2 = 0 \}$



$-\nabla f(x^*) = \lambda_1 \begin{pmatrix} 0 \\ -2 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

$-\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \lambda_1 \begin{pmatrix} 0 \\ -2 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ $\rightsquigarrow$ impossibile

ES 2 - necessary but not sufficient

$$\begin{cases} \min\ x_1+x_2 \\ -x_1^2-x_2^2+2 \leq 0 \end{cases}$$

ACQ holds at $(1,1)$ because $\nabla g\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}-2\\-2\end{pmatrix} \neq \bar{0}$

$$-\begin{pmatrix}1\\1\end{pmatrix} = \lambda_1 \begin{pmatrix}-2\\-2\end{pmatrix} \Rightarrow \lambda_1 = \frac{1}{2}$$

. I found a solution to the system but x^* is not a local optimum

Th: If P is convex (f, g are convex and h affine) and (x^*, λ^*, μ^*) solves KKT system $\Rightarrow$ x^* is a global solution

sufficient condition

Dim $L(x,\lambda,\mu)$ is convex wrt x if P is convex (sum of convex functions)
$\Rightarrow$ $\nabla_x L(x^*,\lambda^*,\mu^*)=0$ means that x^* is a global solution for $L(x,\lambda^*,\mu^*)$

$$L(x^*,\lambda^*,\mu^*) \leq L(x,\lambda^*,\mu^*) \quad \forall x \in \mathbb{R}^n \Rightarrow$$

$$f(x^*) + \cancel{\lambda^{*T} g(x^*)} + \cancel{\mu^{*T} h(x^*)} \leq f(x) + \lambda^{*T} g(x) + \mu^{*T} h(x)$$

solves KKT λ^*, μ^*

$$\text{If } x \in X \text{ (feasible set)} \Rightarrow f(x^*) \leq f(x) + \underbrace{\lambda^{*T} g(x)}_{\leq 0} + \underbrace{\mu^{*T} h(x)}_{=0} \leq f(x)$$

DUALITY

Given $\lambda \geq 0$ and $\mu \in \mathbb{R}^p$ the problem:

$$\begin{cases} \inf\ L(x,\lambda,\mu) \\ x \in \mathbb{R}^n \end{cases}$$ is called lagrangian relaxation of P

$$\varphi(\lambda,\mu) = \inf_{x\in\mathbb{R}^n} L(x,\lambda,\mu)$$ is the LAGRANGIAN DUAL FUNCTION

Th: $\forall \lambda \geq 0,\ \mu \in \mathbb{R}^p \Rightarrow \varphi(\lambda,\mu) \leq v(P)$ → optimal solution of P

$$\varphi(\lambda,\mu) \leq f(x) + \lambda^T g(x) + \mu^T h(x) \quad \forall x \in \mathbb{R}^n$$

in $x \in X$ $\quad \varphi(\lambda,\mu) \leq f(x)$

$$\varphi(\lambda,\mu) \leq \inf f(x) = v(P)$$

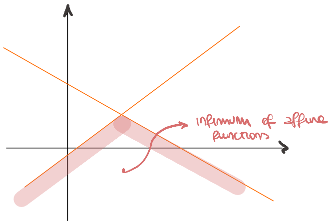
↳ becomes minimum if the solution exists

• for every λ and μ, $\varphi(\lambda,\mu)$ is a lower bound

PROPERITIES of $\varphi(\lambda,\mu)$

1) Is concave because is the inf of a family of affine functions wrt λ,μ

φ is inf of set of concave functions → AFFINE
φ is $-\sup$ of set of convex functions
$\Rightarrow$ $\varphi = -\text{convex} = \text{concave}$



2) May be $-\infty$ at some point

3) May not be differentiable

$$\begin{cases} \max \ \varphi(\lambda,\mu) \\ \lambda \geq 0 \\ \mu \in \mathbb{R}^p \end{cases}$$ is called Lagrangian dual problem

D is always equivalent to a convex problem even if P is non convex. Maximization of concave function in concave set

$$v(D) \leq v(P)$$ → WEAK DUALITY PROPERTY

LINEAR PROGRAMMING

$$\begin{cases} \min \ c^T x \\ Ax \geq b \end{cases} \qquad L(x,\lambda) = c^T x + \lambda^T(b - Ax) = \lambda^T b + (c^T - \lambda^T A)x$$

$$\varphi(\lambda) = \inf_x L(x,\lambda) = \begin{cases} -\infty & \text{if } c^T - \lambda^T A \neq 0 \\ \lambda^T b & \text{if } c^T - \lambda^T A = 0 \end{cases}$$

Dual problem: $$\begin{cases} \max \ \varphi(\lambda) \\ \lambda \geq 0 \end{cases} \Rightarrow \begin{cases} \max \ \lambda^T b \\ \lambda \geq 0 \\ c^T = \lambda^T A \end{cases}$$

ES

$$\begin{cases} \min \ \frac{1}{2}x^Tx \\ Ax = b \end{cases} \qquad L = \frac{1}{2}x^Tx + \mu^T(b - Ax)$$

H is I $\Rightarrow$ strongly convex

$$\varphi(\mu) = \inf_{x \in \mathbb{R}^n} L(x,\mu) = \min_{x \in \mathbb{R}^n} L(x,\mu)$$

$$\nabla_x L = 0 \iff x = A^T\mu$$

$$\Rightarrow \varphi(\mu) = \frac{1}{2}\mu^T A A^T \mu + \mu^T(b - AA^T\mu)$$
$$= -\frac{1}{2}\mu^T A A^T \mu + \mu^T b$$

ALWAYS POSITIVE SEMI-DEFINITE $\Rightarrow$ CONCAVE

$$\begin{cases} \max \ -\frac{1}{2}\mu^T A A^T \mu + \mu^T b \\ \mu \in \mathbb{R}^p \end{cases}$$

$y = A^T\mu \Rightarrow \mu^T A A^T \mu = y^T y = \|y\|^2$

STRONG DUALITY holds if $v(D) = v(P)$ and D admits an optimal solution

Ex

$$\begin{cases} \min \ -x^2 \\ x - 1 \le 0 \\ -x \le 0 \end{cases}$$

$v(P) = -1$

$L(x, \mu, \lambda) = -x^2 + \lambda_1(x-1) - \lambda_2 x$

$\varphi(\lambda) = -\infty \quad \forall (\lambda_1, \lambda_2) \in \mathbb{R}^2 \Rightarrow v(D) = -\infty$

f, g, a continuously differentiable

Th: P is convex and $\exists x^*$ global solution $\wedge$ ACQ holds at x^* $\Rightarrow$

1) $v(D) = v(P)$

2) (λ^*, μ^*) is optimal for D $\iff$ (λ^*, μ^*) is a KKT multipliers vector associated with x^*

Dim of 1) $L(x,\lambda,\mu)$ is convex $\qquad (\lambda^*,\mu^*)$ are KKT multipliers associated with x^* $\quad \exists$ because ACQ holds at x^*

$$\nabla_x L(x^*,\lambda^*,\mu^*) = 0 \qquad \lambda^* \geq 0 \qquad (\lambda^*)^T g(x^*) = 0$$

$$v(D) \geq \varphi(\lambda^*,\mu^*) = \inf_x L(x,\lambda^*,\mu^*) \overset{L \text{ convex}}{=} L(x^*,\lambda^*,\mu^*)$$

$$L(x^*,\lambda^*,\mu^*) = f(x^*) + \cancel{\lambda^{*T} g(x^*)} + \cancel{\mu^{*T} h(x^*)} = f(x^*) = v(P)$$

but for weak duality $v(P) \geq v(D) \Rightarrow v(P) = v(D)$

May holds for non-convex P

Es.
$$\begin{cases} \min & -x_1^2 - x_2^2 \\ & x_1^2 + x_2^2 - 1 \leq 0 \end{cases} \qquad L(x,\lambda) = (\lambda-1)x_1^2 + (\lambda-1)x_2^2 - \lambda$$

$$\varphi(\lambda) = \inf_x L(x,\lambda) = \begin{cases} -\infty & \text{if } \lambda < 1 \\ -\lambda & \text{if } \lambda \geq 1 \end{cases}$$

$$D: \begin{cases} \max & -\lambda \\ & \lambda \geq 1 \end{cases} \qquad \lambda = 1 \text{ is the solution}$$
$v(D) = -1$

KKT
$$\begin{cases} \nabla_x L = 0 \\ \lambda_i g_i = 0 \\ \lambda \geq 0, \; g(x) \leq 0 \end{cases} \Rightarrow \begin{cases} (\lambda-1)x_1 = 0 \\ (\lambda-1)x_2 = 0 \\ \lambda(x_1^2 + x_2^2 - 1) = 0 \\ \lambda \geq 0, \quad x_1^2 + x_2^2 - 1 \leq 0 \end{cases}$$

$$\begin{cases} \lambda = 1 \\ x_1^2 + x_2^2 - 1 = 0 \end{cases} \Rightarrow \text{optimal for D and P}$$
(not happen always)

$$\begin{cases} \lambda = 0 \\ x_1, x_2 = 0 \end{cases} \rightsquigarrow \text{not optimal}$$
(if it was convex all are optimal)

Th: (x^*, λ^*, μ^*) is a saddle point of L if:

$$L(x^*, \lambda, \mu) \leq L(x^*, \lambda^*, \mu^*) \leq L(x, \lambda^*, \mu^*) \quad \forall x \in \mathbb{R}^n \ \forall \lambda^{?}, \mu$$

$\iff$ x^* is optimum of P $\quad (\lambda^*, \mu^*)$ is optimum of D
$\wedge$ $v(P) = v(D)$

$L(x^*, \lambda^*, \mu^*)$ is
- a local maximum in respect to λ, μ fixed x^*
- a local minimum in respect to x fixed λ^*, μ^*

CLASSIFICATION PROBLEM

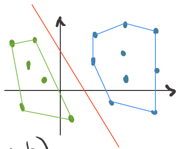
$\rightarrow$ want to classificate a vector $x \in \mathbb{R}^n$ by encoding a particular label

SVM $\rightarrow$ supervised classification method for a vector of data, according to previously classified data (with known labels)

$A, B \subset \mathbb{R}^n$ two finite sets with known labels (1 for A, -1 for B)
- $\mathbb{R}^n$ is the input space
- $A \cup B$ is the training set

In binary classification we assume A and B are strictly linearly separable
$\Rightarrow H = \{x \in \mathbb{R}^n : w^T x + b = 0\}$ such that

$$w^T x^i + b > 0 \quad \forall x^i \in A$$
$$w^T x^j + b < 0 \quad \forall x^j \in B$$



Then we can use the decision function $f(x) = \text{sgn}(w^T x + b)$

Th: strict linearly separation of A and B $\iff$ $\text{conv}(A) \cap \text{conv}(B) = \emptyset$

convex hull

compact set

Def: H is the separation hyperplane $\Rightarrow$ the margin of separation of H is def. as the minimum distance between H and $A \cup B$

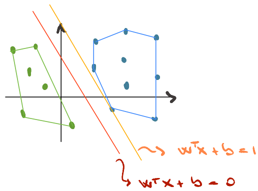
$$\rho(H) = \min_{x \in A \cup B} \frac{|w^T x + b|}{\|w\|}$$

$\rightarrow$ I want to find the maximum margin hyper-plane

Th: max separation is given by the following convex quadratic P:

$$\begin{cases} \min \ \frac{1}{2}\|w\|^2 \\ w^T x^i + b \geq 1 & \forall x^i \in A \\ w^T x^j + b \leq -1 & \forall x^j \in B \end{cases}$$

constants



The distance between the two hyperplane $w^Tx+b=1$ and $w^Tx+b=-1$ $\longrightarrow$ $d = \frac{2}{\|w\|}$

consider a point $\hat{x}$ in $w^T\hat{x}+b=1$ $\Rightarrow$ $d = \frac{|\overbrace{w^T\hat{x}+b}^{=1}+1|}{\|w\|} = \frac{2}{\|w\|}$

Assume $A = \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{p\times n}$, $w \in \mathbb{R}^n$, $b \in \mathbb{R}$

$$\begin{array}{c} \quad n+1 \\ \left(\begin{array}{c|c} \begin{matrix} -A \\ (m\times n) \end{matrix} & \begin{matrix} -e_m \\ (m\times 1) \end{matrix} \\ \hline \begin{matrix} B \\ (p\times n) \end{matrix} & \begin{matrix} e_p \\ (p\times 1) \end{matrix} \end{array}\right) \end{array} \begin{matrix} m \\ + \\ p \end{matrix}$$

$$\begin{cases} \min\limits_{w,b} \ \frac{1}{2}(w,b)\, C \begin{pmatrix} w \\ b \end{pmatrix} \\ D\begin{pmatrix} w \\ b \end{pmatrix} \leq d \end{cases}$$

with $n=2$ $\quad C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

$(1,1,\dots,1) \in \mathbb{R}^m$ but column vector

$$-w^Tx^j - b \leq -1 \qquad D = \begin{pmatrix} -A & -e_m \\ B & e_p \end{pmatrix} \qquad d = \begin{pmatrix} -e_m \\ -e_p \end{pmatrix}$$
$$w^Tx^i + b \leq -1$$

The dual: $\quad \ell = |A \cup B|$

for any point $x^i \in A \cup B$ define a label $y_i = \begin{cases} 1 & x^i \in A \\ -1 & x^i \in B \end{cases}$

LINEAR SVM

$$\begin{cases} \min\limits_{w,b} \ \frac{1}{2}\|w\|^2 \\ 1 - y^i(w^Tx^i + b) \leq 0 & \forall i = 1,\dots,\ell \end{cases}$$

Since P is convex $\Rightarrow$ strong duality since ACQ holds

$\lambda \in \mathbb{R}^\ell$ $\qquad L(\underbrace{w,b}_{\text{original var.}},\lambda) = \frac{1}{2}\|w\|^2 + \sum_{i=1}^{\ell} \lambda_i [1 - y^i(w^Tx^i + b)]$

$$= \frac{1}{2}\|w\|^2 - \sum_{i=1}^{\ell} \lambda_i y^i w^T x^i - b\sum_{i=1}^{\ell}\lambda_i y^i + \sum_{i=1}^{\ell}\lambda_i$$

If $\sum_{i=1}^{\ell} \lambda_i y^i \neq 0 \implies \min_{w,b} L = -\infty$

If $\sum_{i=1}^{\ell} \lambda_i y^i = 0 \implies L$ strongly convex wrt w

$$\nabla_w L(w,b,\lambda) = w - \sum_{i=1}^{\ell} \lambda_i y^i x^i = 0$$

minimum found at the unique stationary point

$$\implies L(w,b,\lambda) = \frac{1}{2}w^T w - w^T w + \sum \lambda_i = -\frac{1}{2}w^T w + \sum\lambda_i$$

$$w = \sum_{i=1}^{\ell} \lambda_i y^i x^i \to L = -\frac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell} y^i y^j (x^i)^T x^j \lambda_i\lambda_j + \sum\lambda_i$$

$$\varphi(\lambda) = \begin{cases} -\infty & \text{if } \sum\lambda_i y^i \neq 0 \\ -\frac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell} y^i y^j (x^i)^T x^j \lambda_i\lambda_j + \sum\lambda_i & \text{if } \sum\lambda_i y^i = 0 \end{cases}$$

DUAL of linear SVM

$$\begin{cases} \max_\lambda & -\frac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell} y^i y^j (x^i)^T x^j \lambda_i\lambda_j + \sum\lambda_i \\ & \sum\lambda_i y^i = 0 \\ & \lambda \geq 0 \end{cases}$$

positive semi-definite

$$\begin{cases} \max_\lambda & -\frac{1}{2}\lambda^T X^T X\lambda + e^T\lambda \\ & \sum\lambda_i y^i = 0 \\ & \lambda \geq 0 \end{cases}$$

with $X = (y^1 x^1, \dots, y^\ell x^\ell)$ $n \times \ell$ matrix, $e^T = (1, \dots, 1) \in \mathbb{R}^\ell$

sufficient and necessary

- a KKT multiplier λ^* of the primal optimum (w^*, b^*) is a dual optimum
- if $\lambda_i^* > 0 \implies x^i$ is said a support vector
- if λ^* is a dual optimum $\implies w^* = \sum\lambda_i^* y^i x^i$ (from KKT since P is convex, w^* is optimal)
- b^* is obtained by : $\lambda_i^*[1 - y^i(w^{*T}x^i + b^*)] = 0 \to$ if $\lambda_i > 0$: $b^* = \frac{1}{y^i} - w^{*T}x^i$

  $\to$ non mi interessa conoscere b^* per valori che non sono boundary

What if A and B are not linearly separable?

- $1 - y^i(w^T x^i + b) \leq 0 \quad \forall i$ has no solution

We need to introduce slack variables $\varepsilon_i \geq 0$: $\begin{cases} 1 - y^i(w^T x^i + b) \leq \varepsilon_i \\ \varepsilon_i \geq 0 \end{cases}$

If x^i is misclassified $\implies \varepsilon_i > 1 \to \sum\varepsilon_i$ is an upper-bound of the number of misclassified points

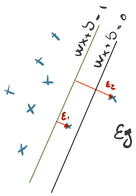
1) If $x^i \in A$ with $w^T x^i + b < 0$ then: $1 < 1 - (w^T x^i + b) = 1 - y^i(w^T x^i + b) \leq \varepsilon_i$

2) If $x^i \in B$ with $w^T x^i + b > 0$ then: $1 < 1 + (w^T x^i + b) = 1 + y^i(w^T x^i + b) \leq \varepsilon_i$

I can define a linear SVM with soft margin:

constant > 0 trade-off between misclassification and a larger gap

$$\begin{cases} \min\limits_{w,b,\varepsilon} \frac{1}{2}\|w\|^2 + C \sum\limits_{i=1}^{l} \varepsilon_i \\ 1 - y^i(w^T x^i + b) \leq \varepsilon_i \\ \varepsilon_i \geq 0 \end{cases}$$



$\varepsilon_j = \max(0, 1 - y_j(wx^j + b))$

To find the dual $L = \frac{1}{2} w^T w + C \sum \varepsilon_i + \sum \lambda_i [1 - y^i(w^T x^i + b) - \varepsilon_i] - \sum \mu_i \varepsilon_i$

$$\begin{cases} \nabla_w L = w - \sum \lambda_i y^i x^i = 0 \\ \nabla_b L = -\sum \lambda_i y^i = 0 \\ \nabla_\varepsilon L = C - \lambda_i - \mu_i = 0 \quad i = 1, \dots, l \end{cases}$$

$\rightarrow$ otherwise un-bounded from below

Since P is convex $\Rightarrow$ a global minimum is a stationary point of L $\Rightarrow$ solution of the system

$$\Rightarrow L = \frac{1}{2} w^T w + C \cancel{\sum} \varepsilon_i + \sum \lambda_i [1 - y^i(w^T x^i \cancel{+b}) - \cancel{\varepsilon_i}] - \sum \cancel{\mu_i} \varepsilon_i$$

$$= \frac{1}{2}\left(\sum \lambda_i y^i x^i\right)^T \left(\sum \lambda_j y^j x^j\right) + \sum \lambda_i (1 - y^i w^T x^i)$$

$$= -\frac{1}{2}\left(\sum \lambda_i y^i x^i\right)^T \left(\sum \lambda_j y^j x^j\right) + \sum \lambda_i$$

DUAL

$$\begin{cases} \max\limits_{\lambda} -\frac{1}{2} \sum\sum y^i y^j (x^i)^T x^j \lambda_i \lambda_j + \sum \lambda_i \\ \sum \lambda_i y^i = 0 \\ C - \lambda_i - \mu_i = 0 \\ \lambda \geq 0, \ \mu \geq 0 \end{cases} \rightarrow \boxed{0 \leq \lambda_i \leq C} \quad \forall i \in \{1, \dots, l\}$$

Since $C - \lambda_i = \mu_i \geq 0 \Rightarrow$

$\hookrightarrow$ the only difference with the previous dual

- $w^* = \sum \lambda_i^* y^i x^i \rightarrow$ only support vectors (margin and over) contribute

- b : $0 < \lambda_i^* < C$ $\Rightarrow$ $\begin{cases} \lambda_i^* [1 - y^i(w^{*T} x^i + b^*) - \varepsilon_i^*] = 0 \\ \mu_i^* \varepsilon_i^* = (C - \lambda_i^*) \varepsilon_i^* = 0 \end{cases}$ complementary KKT conditions

  $\hookrightarrow$ points on the margin $\rightarrow$ MATLAB

$$\Rightarrow b^* = \frac{1}{y^i} - w^{*T} x^i$$

REMARKS

1) $0 < \lambda_i^* < C \Rightarrow \varepsilon_i^* = 0$

2) $0 < \lambda_i^* \leq C \Rightarrow 1 - y^i(w^{*T} x^i + b^*) - \varepsilon_i^* = 0$ $\quad$ can find ε_i if $\lambda_i^* = C$

3) $\varepsilon_i^* > 0 \Rightarrow \lambda_i^* = C$

1) POINTS on the margin : $0 < \lambda_i^* < C \Rightarrow \varepsilon_i = 0 \wedge y^i [w^T x^i + b] = 1$

2) POINTS inside the margin : $\lambda_i^* = 0 \Rightarrow \varepsilon_i^* = 0$

3) POINTS between margin and boundary : $0 < \varepsilon_i^* \leq 1$ (exactly the decision boundary) $\Rightarrow \lambda_i^* = C$

4) POINTS outside the boundary : $\varepsilon_i^* > 1 \Rightarrow \lambda_i^* = C$
↳ misclassified

$\varepsilon_i^* = 1 - y^i (w^{*T} x^i + b^*)$
↓
for $\lambda_i > 0$

- It's more efficient to solve the dual since the dimensionality is lower and also it has a lower number of inequalities

- Found an hyperplane with soft-margin

NON LINEAR SVM

Maybe the two sets are linearly separable in another space

$\phi : \mathbb{R}^n \to \mathcal{H}$ is an higher dimensional space (maybe infinite)
↳ features space

- Then I can consider the dataset as $\{ \phi(x^i) \; i = 1, \dots, \ell \}$

on the features space

$$\begin{cases} \min\limits_{w,b,\varepsilon} \frac{1}{2} \|w\|^2 + C \sum \varepsilon_i \\ 1 - y^i (w^T \phi(x^i) + b) \leq \varepsilon_i \\ \varepsilon_i \geq 0 \end{cases}$$

$$\begin{cases} \max\limits_{\lambda} -\frac{1}{2} \sum \sum y^i y^j \phi(x^i) \phi(x^j) \lambda_i \lambda_j + \sum \lambda_i \\ \sum \lambda_i y^i = 0 \\ 0 \leq \lambda_i \leq C \end{cases}$$

still defined on $\mathbb{R}^\ell$

cannot be retrieved since depends on $\phi(x)$

$$\begin{cases} w^* = \sum \lambda_i y^i \phi(x^i) \\ b^* = \frac{1}{y^i} - \sum\limits_j \lambda_j y^j \phi(x^j)^T \phi(x^i) \end{cases}$$

↓ $0 < \lambda_i < C$

$$f(x) = \text{sign}(w^{*T} \phi(x) + b^*) = \text{sign}\left(\sum_{i=1}^{\ell} \lambda_i^* y^i \phi(x^i)^T \phi(x) + b^* \right)$$

↳ depends only on scalar product in the features space

KERNEL FUNCTION

A function $K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ (original space) is called kernel if $\exists$ a map $\phi : \mathbb{R}^n \to \mathcal{H}$: $K(x, y) = \langle \phi(x), \phi(y) \rangle$

$K(x,y) = x^T y$

$K(x,y) = (x^T y + 1)^p \qquad p \geq 1 \qquad$ POLINOMIAL

$K(x,y) = e^{-\gamma \|x - y\|^2} \qquad \gamma > 0 \qquad$ GAUSSIAN

Th: if K is the kernel and $x^i \in \mathbb{R}^n \Rightarrow K_{ij} = K(x^i, x^j)$ is positive semidefinite

→ In this way the dual is the dual of something (because it must be convex)

$$f(x) = \text{sign}\left(\sum^{\ell} \lambda_i^* y^i K(x^i, x) + b^* \right)$$

↳ non-linear

- $f(x) = 0$ that is the separating surface that is no more an hyperplane

REGRESSION PROBLEMS

We have ℓ experimental data $y_1, \dots, y_\ell \in \mathbb{R}$ corresponding to the observations $x_1, \dots, x_\ell \in \mathbb{R}$

- polynomial regression $\longrightarrow$ want to find the best polynomial with $n \leq \ell$ (errors) where $n-1$ is the degree of the polynomial

$$p(x) = z_1 + z_2 x + \dots + z_n x^{n-1}$$

The residual is defined as $r_i = p(x_i) - y_i$

- I want to minimize the norm of the residual $\|r\|$ $\longrightarrow$ unconstrained minimization

$$\begin{cases} \min\limits_z \|Az - y\| \\ z \in \mathbb{R}^n \end{cases} \qquad A = \begin{pmatrix} 1 & x_1 & \dots & x_1^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_\ell & \dots & x_\ell^{n-1} \end{pmatrix} \in \mathbb{R}^{\ell \times n}$$

- since any norm is convex and composed with a linear function $\Rightarrow$ convex but not any type of norm is differentiable

$(Az)^T y = y^T A z$ because they are scalar

$\|\cdot\|_2$ $$\min \frac{1}{2}\|Az - y\|_2^2 = \frac{1}{2}(Az-y)^T(Az-y) = \frac{1}{2} z^T A^T A z - z^T A^T y + \frac{1}{2} y^T y$$

can be proved that $\operatorname{rank}(A) = n \rightarrow A^T A$ is positive definite then exists a unique solution

stationary point of: $A^T A z = A^T y$

since the inverse exists $z = (A^T A)^{-1} A^T y$

$\|\cdot\|_1$ $$\min \|Az - y\|_1 = \min \sum_{i=1}^{\ell} |A_i z - y_i|$$

this is equivalent to:

$$\begin{cases} \min\limits_{z,u} \sum\limits_{i=1}^{\ell} u_i \\ u_i = |A_i z - y_i| = \max\{A_i z - y_i, -A_i z + y_i\} \end{cases} \Rightarrow \begin{cases} \min\limits_{z,u} \sum u_i \\ u_i \geq \max\{\dots\} \end{cases}$$

$$\Rightarrow \begin{cases} \min\limits_{z,u} \sum u_i \\ u_i \geq A_i z - y_i & \forall i \\ u_i \geq y_i - A_i z & \forall i \end{cases} \Rightarrow \begin{cases} \min\limits_{z,u} e_e^T u \\ Az - u \leq y \\ -Az - u \leq -y \end{cases}$$

- This is a linear programming problem

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

$$d = \begin{pmatrix} y \\ -y \end{pmatrix} \longrightarrow \begin{cases} \min\limits_{z,u} (0_n^T, e_e^T) \begin{pmatrix} z \\ u \end{pmatrix} \\ D \begin{pmatrix} z \\ u \end{pmatrix} \leq d \end{cases}$$

$\|\cdot\|_\infty$

$$\min \|Az - y\|_\infty = \min_z \max_i |A_i z - y_i|$$

$$\begin{cases} \min\limits_{z,u} u \\ u = \max\limits_i |A_i z - y_i| \end{cases} \Rightarrow \begin{cases} \min\limits_{z,u} u \\ u \geq A_i z - y_i & \forall i \\ u \geq y_i - A_i z & \forall i \end{cases}$$

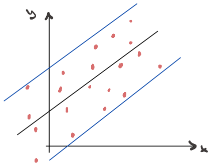
$$D = \begin{pmatrix} A & -e_e \\ -A & -e_e \end{pmatrix}$$

$$d = \begin{pmatrix} y \\ -y \end{pmatrix} \rightarrow \begin{cases} \min\limits_{z,u} (\bar{0}_n, 1) \begin{pmatrix} z \\ u \end{pmatrix} \\ D \begin{pmatrix} z \\ u \end{pmatrix} \leq d \end{cases}$$

ε - SV

given a training dataset $\{(x_i, y_i) \quad i \in 1, \dots \ell\}$ where $x_i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}$ and a tolerance $\varepsilon > 0$, we want to find a function $f : \mathbb{R}^n \to \mathbb{R}$

$$|f(x_i) - y_i| < \varepsilon \quad \forall i$$



In a linear regression $f(x) = w^T x + b$

- if I want the function flat :

$$\begin{cases} \min_{w,b} \frac{1}{2}\|w\|^2 \\ y_i \leq w^T x + b + \varepsilon \\ y_i \geq w^T x + b - \varepsilon \end{cases}$$

→ convex not strongly convex because the solution isn't unique since I can vary b

not positive definite

$$\begin{cases} \min_{w,b} \frac{1}{2}(w^T \ b)\, Q \begin{pmatrix} w \\ b \end{pmatrix} \\ D \begin{pmatrix} w \\ b \end{pmatrix} \leq d \end{cases}$$

$$Q = \begin{pmatrix} I_n & 0_n \\ 0_n^T & 0 \end{pmatrix}$$

$$D = \begin{pmatrix} -X & -e_\ell \\ X & e_\ell \end{pmatrix} \qquad d = \begin{pmatrix} \varepsilon e_\ell - y \\ \varepsilon e_\ell + y \end{pmatrix}$$

- If ε is too small the P could be infeasible so I can add slack variables

$$\begin{cases} \min_{w,b,\varepsilon^+,\varepsilon^-} \frac{1}{2}\|w\|^2 + C\sum^{\ell}(\varepsilon_i^+ + \varepsilon_i^-) \\ y_i \leq w^T x_i + b + \varepsilon + \varepsilon_i^+ & \forall i \\ y_i \geq w^T x_i + b - \varepsilon - \varepsilon_i^- & \forall i \\ \varepsilon_i^+, \varepsilon_i^- \geq 0 & \forall i \end{cases}$$

$$\begin{cases} \min \frac{1}{2}(w^T, b, \varepsilon^{+T}, \varepsilon^{-T})\, Q_1 \begin{pmatrix} w \\ b \\ \varepsilon^+ \\ \varepsilon^- \end{pmatrix} + C^T \begin{pmatrix} w \\ b \\ \varepsilon^+ \\ \varepsilon^- \end{pmatrix} \\ D_1 \begin{pmatrix} w \\ b \\ \varepsilon^+ \\ \varepsilon^- \end{pmatrix} \leq d_1 \\ \varepsilon^+ \geq 0 \quad \varepsilon^- \geq 0 \end{cases}$$

with $$Q_1 = \begin{pmatrix} I_n & 0_n & 0_{n\times 2\ell} \\ 0_n^T & 0 & 0_{2\ell}^T \\ 0_{2\ell \times n} & 0_{2\ell} & 0_{2\ell \times 2\ell} \end{pmatrix} \qquad C^T = (0_n^T, 0, Ce_\ell^T, Ce_\ell^T)$$

$$D_1 = \begin{pmatrix} -X & -e_\ell & -I_\ell & 0_{\ell\times\ell} \\ X & e_\ell & 0_{\ell\times\ell} & -I_\ell \end{pmatrix} \qquad d_1 = \begin{pmatrix} \varepsilon e_\ell - y \\ \varepsilon e_\ell + y \end{pmatrix}$$

The dual formulation is interesting because it can extends the model to a non-linear formulation

→ constraints must be $a \le 0$

$L(w, b, \varepsilon^+, \varepsilon^-, \lambda^+, \lambda^-, \eta^+, \eta^-)$

$$L = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{\ell}(\varepsilon_i^+ + \varepsilon_i^-) + \sum_{i=1}^{\ell}\lambda_i^+[y_i - w^T x_i - b + \varepsilon + \varepsilon_i^+] + \sum \lambda_i^-[w^T x_i + b - \varepsilon - \varepsilon_i^- - y_i] - \sum \eta_i^+ \varepsilon_i^+ - \sum \eta_i^- \varepsilon_i^-$$

$$= \frac{1}{2}\|w\|^2 - w^T\sum_{i=1}^{\ell}(\lambda_i^+ - \lambda_i^-)x_i \cancel{- b\sum(\lambda_i^+ - \lambda_i^-)} + \sum \varepsilon_i^+(\cancel{C - \lambda_i^+ - \eta_i^+}) + \sum \varepsilon_i^-(\cancel{C - \lambda_i^- - \eta_i^-}) - \varepsilon\sum(\lambda_i^+ + \lambda_i^-) + \sum y_i(\lambda_i^+ - \lambda_i^-)$$

Since it is unconstrained, to find the minim wrt $w, b, \varepsilon^+, \varepsilon^-$ I can take $\nabla = 0$

$$\nabla_w L = w - \sum(\lambda_i^+ - \lambda_i^-)x_i = 0 \qquad \nabla_b L = -\sum(\lambda_i^+ - \lambda_i^-) = 0$$

$$\nabla_{\varepsilon^+} L = \sum(C - \lambda_i^+ - \eta_i^+) = 0 \qquad \nabla_{\varepsilon^-} L = \sum(C - \lambda_i^- - \eta_i^-) = 0$$

$$\begin{cases} \max\limits_{\lambda^+, \lambda^-, \cancel{\eta^+, \eta^-}} -\frac{1}{2}\sum\sum(\lambda_i^+ - \lambda_i^-)(\lambda_j^+ - \lambda_j^-)x_i^T x_j - \varepsilon\sum(\lambda_i^+ + \lambda_i^-) + \sum y_i(\lambda_i^+ - \lambda_i^-) \\ \sum(\lambda_i^+ - \lambda_i^-) = 0 \\ C - \lambda_i - \eta_i^+ = 0 \\ C - \lambda_i - \eta_i^+ = 0 \\ \lambda^+, \lambda^-, \eta^+, \eta^- \geq 0 \end{cases} \quad \Bigg\} \; 0 \le \lambda_i \le C$$

→ do not appear in the max function

$$X = [x_i^T x_j] \qquad Q = \begin{pmatrix} X & -X \\ -X & X \end{pmatrix}$$

→ less variables and constrains

$$\begin{cases} \max\limits_{\lambda^+, \lambda^-} -\frac{1}{2}(\lambda^{+T}, \lambda^{-T})\, Q \begin{pmatrix} \lambda^+ \\ \lambda^- \end{pmatrix} + [-\varepsilon(e_\ell^T, e_\ell^T) + (y^T, -y^T)] \begin{pmatrix} \lambda^+ \\ \lambda^- \end{pmatrix} \\ (e_\ell^T, -e_\ell^T)\begin{pmatrix} \lambda^+ \\ \lambda^- \end{pmatrix} = 0 \\ \lambda_i^+ \in [0, C] \\ \lambda_i^- \in [0, C] \end{cases}$$

from KKT conditions:

$$\begin{cases} \lambda_i^-[w^T x^i + b - \varepsilon - \varepsilon_i^- - y_i] = 0 \\ \varepsilon_i^-(\overbrace{C - \lambda_i^-}^{\eta_i}) = 0 \end{cases}$$

if there is error ⇒ $\lambda_i = C$

⇒ if $0 < \lambda_i^- < C$ I get $\varepsilon_i^- = 0$ and $b = -w^T x^i + \varepsilon + y_i$

⇒ if $0 < \lambda_i^+ < C$ ⇒ $\varepsilon_i^+ = 0$ and $b = -w^T x^i - \varepsilon + y_i$

- Dual problem is convex
- If $\lambda_i^+ > 0 \lor \lambda_i^- > 0$ $\rightarrow$ x_i is called support vector $\left(w = \sum (\lambda_i^+ - \lambda_i^-) x_i \right)$

$\rightsquigarrow$ eg. find the errors given f

$\phi: \mathbb{R}^n \rightarrow \mathcal{H}$ (high dimensional space) $\rightarrow$ I want a linear regression of
$D' = \{ \phi(x_i), y_i \} \subseteq \mathcal{H} \times \mathbb{R}$

$$\begin{cases} \min \frac{1}{2} \|w\|^2 + C \sum (\varepsilon_i^+ + \varepsilon_i^-) \\ y_i \leq w^T \phi(x_i) + b + \varepsilon + \varepsilon_i^+ & \forall i \\ y_i \geq w^T \phi(x_i) + b - \varepsilon - \varepsilon_i^- & \forall i \end{cases}$$

- we do not know ϕ but the dual is solvable

$$\begin{cases} \max_{\lambda_i^+, \lambda^-} -\frac{1}{2} \sum \sum (\lambda_i^+ - \lambda_i^-)(\lambda_j^+ - \lambda_j^-) \phi(x_i)^T \phi(x_j) - \varepsilon \sum (\lambda_i^+ + \lambda_i^-) + \sum y_i (\lambda_i^+ - \lambda_i^-) \\ \sum (\lambda_i^+ - \lambda_i^-) = 0 \\ \lambda_i^+, \lambda_i^- \in [0, C] \end{cases}$$

$$w^T \phi(x) = \sum_{i=1}^{\ell} (\lambda_i^+ - \lambda_i^-) \phi(x_i) \phi(x)$$

support vector but with $\lambda_i^+ \neq C$

$$b = y_i - \varepsilon - \sum_{j=1}^{\ell} (\lambda_j^+ - \lambda_j^-) \phi(x_i) \phi(x_j) \qquad i: \quad 0 < \lambda_i^+ < C$$

or

$$b = y_i + \varepsilon - \sum_{j=1}^{\ell} (\lambda_j^+ - \lambda_j^-) \phi(x_i) \phi(x_j) \qquad i: \quad 0 < \lambda_i^- < C$$

In this way we can compute the regression function

$$f(x) = \underbrace{w^T \phi(x) + b}_{\text{linear in the feature space}} = \underbrace{\sum (\lambda^+ - \lambda_i^-) K(x_i, x_j) + b}_{\text{non linear in the input space}}$$

CLUSTERING → unsupervised machine learning problem

Given a set S of patterns and $K \in \mathbb{N}$. A clustering problem consists in finding a partition S in K subsets $S_1, \dots, S_K$ that are well separate and homogeneous

$\hookrightarrow$ CLUSTERS

$$S = \{ p_1, \dots, p_n \} \quad \text{with} \quad p_i \in \mathbb{R}^n$$

- distance will be defined as $p: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ $\quad d(x,y) = \|x-y\|_2^2$ or $d(x,y) = \|x-y\|_1$
- for each of the K clusters S_j I introduce $x_j \in \mathbb{R}^n$ as a centroid
- Define clusters: each p_i is associated with the nearest centroid

$$\begin{cases} \min\limits_{x} \sum\limits_{i=1}^{\ell} \min\limits_{j=1,\dots,K} d(p_i, x_j) \\ x_j \in \mathbb{R}^n \end{cases} \quad \overset{\|\cdot\|_2^2}{\Longrightarrow} \quad \begin{cases} \min \sum \min\limits_{j=1,\dots,K} \| p_i - x_j \|_2^2 \\ x_j \in \mathbb{R}^n \end{cases}$$

the sup is convex

If $K=1 \Rightarrow$

$$\begin{cases} \min \sum\limits_{i=1}^{\ell} (x-p_i)^T (x-p_i) \\ x \in \mathbb{R}^n \end{cases}$$

→ convex and quadratic problem

The gradient wrt x is: $\nabla x = \sum\limits_{i=1}^{\ell} [(x-p_i) + (x-p_i)] = 2\ell x - 2\sum\limits_{i=1}^{\ell} p_i = 0$

$$\Rightarrow x = \frac{\sum p_i}{\ell}$$ ← mean / baricenter

- If $K>1 \Rightarrow$ P is non differentiable and not convex

- for fixed p_i and x_j

$$\min_{j=1,\dots,K} \| p_i - x_j \|_2^2 = \begin{cases} \min\limits_{\alpha} \sum\limits_j \alpha_{ij} \| p_i - x_j \|_2^2 \\ \sum\limits_j \alpha_{ij} = 1 \\ \alpha_{ij} \geq 0 \quad \forall j \end{cases}$$

$\alpha_{ij} \in \{0,1\}$ but it the same as $\alpha_{ij} \in [0,1]$ since its linear in α_{ij} the solution is in a vertex of the simplex (vertex) with all zeros except one

$$\alpha_{ij}^* = \begin{cases} 1 & \text{if } j \text{ first index: } \| p_i - x_j \|_2^2 = \min\limits_{a} \| p_i - x_a \|_2^2 \\ 0 & \text{otherwise} \end{cases}$$

So the initial problem is: product of convex doesn't imply convexity

$$\begin{cases} \min_{x,\alpha} f(x,\alpha) \doteq \sum_i \sum_j \alpha_{ij} \|p_i - x_j\|_2^2 \\ \sum_j \alpha_{ij} = 1 \quad \forall i \\ \alpha_{ij} \geq 0 \quad \forall i, \forall j \\ x_j \in \mathbb{R}^n \end{cases}$$

→ non convex but differentiable (so I can write KKT)

→ simple contraints

→ The dual will not be equivalent, but since it is differentiable I can write KKT. The Algorithm will stop at a stationary point

① If x_j are fixed I can decompose P into e simple LP problems

$$\forall i : \alpha_{ij}^* = \begin{cases} 1 & \text{if } \|p_i - x_j\|_2 = \min_h \|p_i - x_h\|_2 \\ 0 & \text{otherwise} \end{cases}$$

② If α_{ij} are fixed → decomposable in k convex QP prob.

$$\forall j \begin{cases} \min_{x_j} \sum_i \alpha_{ij} \|p_i - x_j\|_2^2 = \min_{x_j} \sum_i^{e} \alpha_{ij} (x_j - p_i)^T (x_j - p_i) \\ x_j \in \mathbb{R}^n \end{cases}$$

minimize each term of the initial summation

The optimal solution $x_j^* = \dfrac{\sum_{i=1}^{e} \alpha_{ij} p_i}{\sum_{i=1}^{e} \alpha_{ij}}$

K-MEANS ALGORITHM

$$f(x,\alpha) = \sum_i^{e} \sum_j^{k} \alpha_{ij} \|p_i - x_j\|_2^2$$

0) $t=0$ choose centroids $x_1^0, \dots, x_k^0 \in \mathbb{R}^n$ and assign pattern to each clusters:

$$\alpha_{ij}^0 = \begin{cases} 1 & \text{if } j: \|p_i - x_j^0\|_2 = \min_h \|p_i - x_h^0\|_2 \\ 0 & \text{otherwise} \end{cases}$$

UPDATE CENTROIDS

1) $\forall j \in \{1, \dots, k\}$ $\quad x_j^{t+1} = \dfrac{\sum \alpha_{ij}^t p_i}{\sum \alpha_{ij}^t}$ just points on the cluster x_j

UPDATE CLUSTERS

2) $\forall i \in \{1, \dots, e\}$ $\quad \alpha_{ij}^{t+1} = \begin{cases} 1 & j: \|p_i - x_j^{t+1}\|_2 = \min_h \|p_i - x_h^{t+1}\|_2 \\ 0 & \text{otherwise} \end{cases}$

3) If $f(x^{t+1}, \alpha^{t+1}) = f(x^t, \alpha^t)$ ⟹ STOP
else $t = t+1$, go to step 1)

• This method doesn't give the optimal solution but

Th: K-means stops after a finite # of iterations at a solution (α^*, x^*) of the KKT system:
↳ necessary for optimality

1) $f(x^*, \alpha^*) \leq f(x^*, \alpha) \quad \forall \alpha \geq 0 : \sum_{j=1}^{k} \alpha_{ij} = 1 \quad \forall i$

2) $f(x^*, \alpha^*) \leq f(x, \alpha^*) \quad \forall x \in \mathbb{R}^{kn}$

→ a local optimum

→ es. prova con + punti di inizio!

$\|\cdot\|_1$

$d(x,y) = \|x-y\|_1 \Rightarrow \begin{cases} \min \sum_{i=1}^{\ell} \min_{j=1,\dots,k} \|p_i - x_j\|_1 \\ x_j \in \mathbb{R}^n \end{cases}$

If $k=1$ the problem is convex decomposable in n convex P:

$$\min \sum_{i=1}^{\ell} \|p_i - x\|_1 = \min \sum_{i=1}^{\ell} \sum_{a=1}^{n} |x_a - (p_i)_a| = \min \sum_{a=1}^{n} \underbrace{\sum_{i=1}^{\ell} |x_a - (p_i)_a|}_{f_a(x_a)}$$

Given $a_1 < \dots < a_\ell \in \mathbb{R}$ what is the solution of:

$\begin{cases} \min \sum_{i=1}^{\ell} |x - a_i| = f(x) \\ x \in \mathbb{R} \end{cases}$ → sum of convex function is convex

→ $x_a^* = \text{median}((p_i)_a)$

if $x < a_1$ $\quad f(x) = -\ell x + \sum a_i$

if $x \in [a_1, a_2]$ $\quad f(x) = x - a_1 + \sum (a_i - x) = x - a_1 + \sum_{i=2} a_i - (\ell-1)x = (\ell-2)x - a_1 + \sum_{i=2}^{\ell} a_i$

...

We can prove that $x^* = \text{median}(a_1, \dots, a_\ell) = \begin{cases} a_{\frac{\ell+1}{2}} & \text{if } \ell \text{ odd} \\ \frac{a_{\ell/2} + a_{\ell/2+1}}{2} & \text{if } \ell \text{ even} \end{cases}$

If $k>1$ problem is equivalent to:

$\begin{cases} \min \sum \min |p_i - x_j| \\ x_j \in \mathbb{R}^n \end{cases}$

$\begin{cases} \min_{x,\alpha} \sum_{i=1}^{\ell} \sum_{j=1}^{k} \alpha_{ij} \|p_i - x_j\|_1 \\ \sum_{j=1}^{k} \alpha_{ij} = 1 \quad \forall i \\ \alpha_{ij} \geq 0 \\ x_j \in \mathbb{R}^n \quad \forall j \end{cases}$

NON CONVEX and not DIFFERENTIABLE

$$f(x,\alpha) = \sum_{i=1}\sum_{j=1} \alpha_{ij} |p_i - x_j| = \sum_{i=1}^{\ell}\sum_{j=1}^{k}\sum_{h=1}^{\hat{}} \alpha_{ij} u_{ijh}$$

where $u_{ijh} = |(x_j)_h - (p_i)_h| = \max\{(x_j)_h - (p_i)_h \,;\, (p_i)_h - (x_j)_h\}$

$$\begin{cases} \min\limits_{x,\alpha,u} \sum_{i=1}^{\ell}\sum_{j=1}^{k}\sum_{h=1}^{n} \alpha_{ij} u_{ijh} \\ u_{ijk} \geq (p_i)_h - (x_j)_h \\ u_{ijk} \geq (x_j)_h - (p_i)_h \\ \sum \alpha_{ij} = 1 \\ \alpha_{ij} \geq 0 \\ x_j \in \mathbb{R}^n \end{cases}$$

$\longrightarrow$ DIFFERENTIABLE but not convex

- If x_j are fixed $\Rightarrow$ ℓ simple LP problem $\forall i$

$$\alpha_{ij}^* = \begin{cases} 1 & |p_i - x_j| = \min \|p_i - x_h\| \\ 0 \end{cases}$$

- If α_{ij} is fixed $\rightarrow$ k simple convex problems

points associated to x_j

$$\begin{cases} \min \sum_{i=1}^{\ell} \alpha_{ij} |p_i - x_j| = \min \sum_{i=1}^{\ell}\sum_{h=1}^{n} \alpha_{ij} |(p_i)_h - (x_j)_h| \\ x_j \in \mathbb{R} \end{cases} \Rightarrow x_j^* = \text{median}(p_i : \alpha_{ij} = 1)$$

$\rightarrow$ Alternative minimization of $f(x,\alpha)$

Th: k-median alg stops at (x^*, α^*) :

$$f(x^*,\alpha^*) \leq f(x^*,\alpha) \qquad \forall \alpha \geq 0 \;:\; \sum^{k} \alpha_{ij} = 1 \quad \forall i$$
$$f(x^*,\alpha^*) \leq f(x,\alpha^*) \qquad \forall x \in \mathbb{R}^{kn}$$

$\rightarrow$ doesn't guarantee to find a global optimum

- Alg is equal to the ONE ABOVE

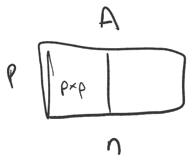
CONSTRAINED OPTIMIZATION

$$\begin{cases} \min\limits_x f(x) \\ g_i(x) \leq 0 & \forall i \\ h_j(x) = 0 & \forall j \end{cases}$$

$X \doteq \{ x \in \mathbb{R}^n : g(x) \leq 0 \land h(x) = 0 \}$

↳ feasible set of P

1) Primal solution with changing variables



$$\begin{cases} \min f(x) \\ Ax = b \end{cases} \quad \text{with } A \in \mathbb{R}^{p \times n}, \ \text{rank}(A) = p$$

Is equivalent to an unconstrained problem:

$A = (A_B, A_N)$ with $\det(A_B) \neq 0$, where $A_B \in \mathbb{R}^{p \times p}$

$x = \begin{pmatrix} x_B \\ x_N \end{pmatrix} \implies Ax = A_B x_B + A_N x_N = b \implies x_B = A_B^{-1}(b - A_N x_N)$

$\implies$ eliminating x_B $\quad \begin{cases} \min f(A_B^{-1}(b - A_N x_N), x_N) \\ x_N \in \mathbb{R}^{n-p} \end{cases}$

2) Penalty Methods

$$\begin{cases} \min f(x) \\ g(x) \leq 0 \end{cases}$$

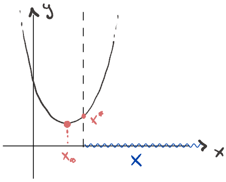
$p(x) \doteq \sum_{i=1}^{m} (\max\{0, g_i(x)\})^2$ — più mi allontano da X più è costoso

↳ quadratic penalty function

$$\begin{cases} \min f(x) + \frac{1}{\varepsilon} p(x) \doteq P_\varepsilon(x) \\ x \in \mathbb{R}^n \end{cases}$$

note that $P_\varepsilon(x) = \begin{cases} f(x) & x \in X \\ > f(x) & x \notin X \end{cases}$

- If f and g_i are continuously differentiable $\implies P_\varepsilon(x)$ is continuously differentiable
  $\nabla_x P_\varepsilon(x) = \nabla f(x) + \frac{2}{\varepsilon} \sum_{i=1}^{m} \max\{0, g_i(x)\} \nabla g_i(x)$
- If f and g_i are convex $\implies P_\varepsilon(x)$ is convex
- Any P_ε is a relaxation of P, since $v(P_\varepsilon) \leq v(P) \quad \forall \varepsilon > 0$ — bigger feasible set
- If x_ε^* solves P_ε $\land$ $x_\varepsilon^* \in X \implies x_\varepsilon^*$ is optimal for P
- $0 < \varepsilon_2 < \varepsilon_1 \implies v(P_{\varepsilon_1}) \leq v(P_{\varepsilon_2})$



1) set $\varepsilon_0 > 0$, $\tau \in (0,1)$, $k = 0$
2) find the optimal solution x^k for P_{ε_k}
3) if $x^k \in X \implies$ STOP
   else $\varepsilon_{k+1} = \tau \varepsilon_k$, $k = k+1$

→ generate non-decreasing values: $v(P_\varepsilon) \geq v(P_{\varepsilon'})$ with $\varepsilon < \varepsilon'$

- If f is coercive $\Rightarrow$ $\{x^k\}$ is bounded and any of its cluster points is an optimal solution
- If $\{x^k\} \to x^*$ $\wedge$ gradients of active constraints at x^* are linearly independent
  $\Rightarrow$ x^* optimal for P $\wedge$ $\{\lambda^k\} \to \lambda^*$ vector of KKT mult associated with x^*
  $\lambda_i^k \doteq \frac{2}{\varepsilon_k} \max\{0, g_i(x^k)\}$

$$0 = \nabla P_{\varepsilon_k}(x^k) = \nabla f(x^k) + \sum_{i=1}^{m} \frac{2}{\varepsilon_k} \max\{0, g_i(x^k)\} \nabla g_i(x^k) \quad \forall x$$

$$\nabla f(x^*) + \sum \lambda_i^* \nabla g_i(x^*) = 0$$

3) exact penalty method

$$\begin{cases} \min f(x) \\ g(x) \le 0 \end{cases}$$ with P convex

$$\tilde{p}(x) = \sum_{i=1}^{m} \max\{0, g_i(x)\} \Rightarrow \begin{cases} \min \tilde{P}_\varepsilon(x) \doteq f(x) + \frac{1}{\varepsilon}\tilde{p}(x) \\ x \in \mathbb{R}^n \end{cases}$$

NON SMOOTH

$$\tilde{P}_\varepsilon(x) = \begin{cases} f(x) & x \in X \\ > f(x) & x \notin X \end{cases}$$

$\longrightarrow$ we do not need $\varepsilon^k \to 0$ to approximate the solution because $\exists \varepsilon^k : v(P) = v(P_{\varepsilon^k})$

Th: Suppose $\exists x^*$ for P and λ^* KKT multipliers $\Rightarrow$ $\{x^* \text{ for } P\} \doteq \{x_\varepsilon^* \text{ for } P_\varepsilon\}$
given that $\varepsilon \in \left(0, \frac{1}{\|\lambda^*\|_\infty}\right)$

$\to$ the method stop at a finite number of iterations at an optimal for P

4) BARRIER METHODS

$$\begin{cases} \min f(x) \\ g(x) \le 0 \end{cases}$$

if f, g_i convex and twice continuously differentiable on $\text{int}(X)$
$\exists x^*$ (eg f coercive or X bounded)
Slater constraints qualification holds ($\exists \bar{x} : g_i(\bar{x}) < 0 \ \forall i$)

$\Rightarrow$ strong duality holds

On the interior of X we can approximate the problem with:

$$\begin{cases} \min \psi_\varepsilon(x) \doteq f(x) - \varepsilon \sum_{i=1}^{m} \log(-g_i(x)) \\ x \in \text{int}(X) \end{cases}$$

logarithmic barrier function

$$B(x) \doteq -\sum_{i=1}^{m} \log(-g_i(x))$$

compositions of concave

$$\psi_\varepsilon(x) = f(x) + \varepsilon B(x)$$

as $x \to$ the boundary of X, the $\psi(x) \to +\infty$
as $\varepsilon \to 0$, $\psi(x) \to f(x)$

domain

- $\mathrm{dom}(B) = \mathrm{int}(X)$
- B is convex
- B is smooth $\qquad \nabla B(x) = -\sum_{i=1}^{m} \frac{\nabla g_i(x)}{g_i(x)}$

$$\nabla^2 B(x) = \sum_{i=1}^{m} \frac{1}{g_i^2(x)} \nabla g_i(x) \nabla g_i(x)^T - \sum_{i=1}^{m} \frac{1}{g_i(x)} \nabla^2 g_i(x)$$

$$\begin{cases} \min \ f(x) - \varepsilon \sum_{i=1}^{m} \log(-g_i(x)) \\ x \in \mathrm{int}(X) \end{cases}$$

If x_ε^* is an optimal solution of this problem $\Rightarrow$

$$\nabla f(x_\varepsilon^*) + \sum_{i=1}^{m} \frac{\varepsilon}{-g_i(x_\varepsilon^*)} \nabla g_i(x_\varepsilon^*) = 0$$

Defining $\lambda_\varepsilon^* = \left(\frac{\varepsilon}{-g_1(x_\varepsilon^*)}, \dots, \frac{\varepsilon}{-g_m(x_\varepsilon^*)} \right) > 0$ because x_ε^* is in $\mathrm{int}(X)$ $\wedge$ $g_i(x_\varepsilon^*) < 0$

The Lagrange function L associated with the original P is:

$$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$$

but $L(x, \lambda^*) = f(x) = f(x) + \sum_{i=1}^{m} (\lambda_\varepsilon^*)_i \, g_i(x) \longrightarrow$ CONVEX

so $\nabla_x L(x^*, \lambda^*) = 0$ for above statements $\Rightarrow$ x^* is a global minimum because convex

Since P is convex and strong duality holds $\quad v(P) = \max_{\lambda \in \mathbb{R}_+^m} \inf_{x \in \mathbb{R}^n} L(x, \lambda)$

$$v(P) \geq \min_x L(x, \lambda_\varepsilon^*) = L(x_\varepsilon^*, \lambda_\varepsilon^*)$$

OPTIMALITY GAP

$$\Rightarrow \quad f(x_\varepsilon^*) \geq v(P) \geq L(x_\varepsilon^*, \lambda_\varepsilon^*) = f(x_\varepsilon^*) - m\varepsilon$$

$$f(x_\varepsilon^*) - m\varepsilon \leq v(P) \leq f(x_\varepsilon^*)$$

For $\varepsilon \to 0 \qquad f(x_\varepsilon^*) \longrightarrow v(P)$

KKT of P
$$\begin{cases} \nabla f(x) + \sum \lambda_i \nabla g_i(x) = 0 \\ -\lambda_i g_i(x) = 0 \quad \forall i \\ \lambda \geq 0 \\ g(x) \leq 0 \end{cases}$$

$x_\varepsilon^*, \lambda_\varepsilon^*$ solves
$$\begin{cases} \nabla f(x) + \sum \lambda_i \nabla g_i(x) = 0 \\ -\lambda_i g_i(x) = \varepsilon \quad \forall i \\ \lambda \geq 0 \\ g(x) \leq 0 \end{cases}$$

APPROXIMATION OF the original KKT

1) set tolerance $\delta > 0$, $\tau \in (0,1)$, $\varepsilon_1 > 0$, $x^0 \in \text{int}(X)$ $\quad K=1$

2) find the optimal of $\begin{cases} \min f(x) - \varepsilon B(x) \\ x \in \text{int}(X) \end{cases}$ using x^{k-1} as starting point

exist because there is coercivity on the boundary $\to +\infty$

3) If $m\varepsilon_k < \delta \to$ STOP

else $\varepsilon_{k+1} = \tau \varepsilon_k$, $k = k+1$

In order to find an interior problem $x^0 \in \text{int}(X)$ we can consider:

$$\begin{cases} \min\limits_{s,x} s \\ g_i(x) \leq s \quad \forall i \end{cases}$$

$\rightarrow$ NON LO CHIEDE

take any $\tilde{x} \in \mathbb{R}^n$, find $\tilde{s} > \max\limits_{i=1,\dots,m} g_i(\tilde{x}) \Rightarrow (\tilde{x}, \tilde{s})$ is in the interior of the feasible set of the auxiliary problem

(x^*, s^*) find the optimal solution for the auxiliary problem using barrier method from $(\tilde{x}, \tilde{s})$

If $s^* < 0 \Rightarrow x^* \in \text{int}(X)$ else $\text{int}(X) = \emptyset$

MULTI-OBJECTIVE OPTIMIZATION

$$\begin{cases} \min f(x) = (f_1(x), \dots, f_s(x)) \\ x \in X \end{cases}$$

→ s objectives to be simultaneously minimized

- We need to define an order on $\mathbb{R}^s$ (comparing vectors)

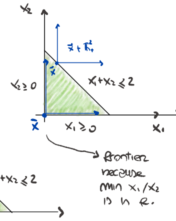
Pareto Order $x, y \in \mathbb{R}^s$, $x \geq y \iff x_i \geq y_i \quad \forall i \in \{1, \dots, s\}$

- it's a partial order, not any two elements are ordered
  - reflexive $x \geq x$
  - asymmetric $x \geq y \wedge y \geq x \Rightarrow x = y$
  - transitive $x \geq y \wedge y \geq z \Rightarrow x \geq z$

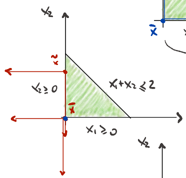
Given a subset $A \subseteq \mathbb{R}^s$

- $\bar{x} \in A$ is a Pareto ideal minimum of A if $y \overset{\text{component wise}}{\geq} \bar{x} \quad \forall y \in A$ (rarely it doesn't exist)
- $\bar{x} \in A$ is a Pareto minimum (efficient point) of A if $\nexists y \in A$, $y \neq \bar{x}$: $\bar{x} \geq y$
  $(\bar{x} \geq y \wedge \exists j : x_j > y_j$ → not equal$)$ → cannot reduce at least one components
- $\bar{x} \in A$ is a Pareto weak minimum of A if $\nexists y \in A$: $\bar{x} > y$ → cannot reduce all components

$$\mathrm{IMin}(A) \subseteq \mathrm{Min}(A) \subseteq \mathrm{WMin}(A)$$

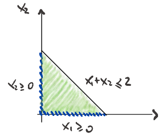


- $\bar{x} \in A$ is an ideal minimum if $A \subseteq (\bar{x} + \mathbb{R}^s_+)$ ↱ something positive in all the components



- $\bar{x} \in A$ is a Pareto minimum if $A \cap \underbrace{(\bar{x} - \mathbb{R}^s_+)}_{y} = \{\bar{x}\}$

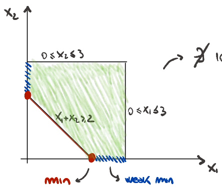
- $\bar{x} \in A$ is a weak minimum if $A \cap (\bar{x} - \mathrm{int}(\mathbb{R}^s_+)) = \emptyset$



Th: If $\mathrm{IMin} \neq \emptyset \Rightarrow \mathrm{IMin}(A) = \mathrm{Min}(A) = \{\bar{x}\}$

if $x^1 \in \mathrm{IMin}(A)$, $x^2 \in \mathrm{Min}(A)$, $x^2 \neq x^1$

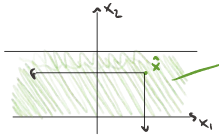
- $x^1 \in \mathrm{IMin}(A) \Rightarrow x_1 \leq x_2$
- $x^2 \in \mathrm{Min}(A) \Rightarrow x_1 = x_2$ a contradiction



↗ $\nexists$ ideal minimum

Th: If $\exists \hat{x} \in A$: $\quad A \cap (\hat{x} - \mathbb{R}^s_+)$ is $\underbrace{\text{compact}}_{\text{closed and bounded}}$ $\Rightarrow$ $Min(A) \neq \emptyset$

(esempio di prima togliendo $x_1 \leq 3 \wedge x_2 \leq 3$, trovo $\hat{x} = 3$ e sono a cavallo)



not compact $\forall \hat{x}$

$$P: \begin{cases} \min\ f(x) = (f_1(x), \dots, f_s(x)) \\ x \in X \end{cases}$$

- if $x^* \in X$ is a pareto ideal minimum of P if $f(x) \geq f(x^*) \quad \forall x \in X$ — global minimum for each function
- if $x^* \in X$ is a minimum $\Rightarrow$ $\nexists x \in X$: $\begin{array}{l} f_i(x^*) \geq f_i(x) \\ f_j(x^*) > f_j(x) \quad \text{for at least one } j \end{array}$
- if $x^* \in X$ is a weak minimum of P $\Rightarrow$ $\nexists x \in X$: $f_i(x^*) > f_i(x) \quad \forall i$

EXAMPLE
$\min(x_1 - x_2, -2x_1 + x_2)$

$$f(X) = \{ (y_1, y_2) : y_1 = x_1 - x_2,\ y_2 = -2x_1 + x_2,\ x \in X \}$$

substitute the inequalities: $\quad x_1 = -y_1 - y_2 \qquad x_2 = -2y_1 - y_2$

$$\Rightarrow f(X) = \{ (y_1, y_2) : -y_1 - y_2 \leq 1,\ y_1 + y_2 \leq 0,\ -y_1 \leq 2,\ -y_2 \leq 0 \}$$

solve for the minimum and then resubstitute and add constraints

$$Min \rightarrow (x_1, x_2) : \begin{cases} x_1 = 1 \\ -x_1 \leq 0 \\ -x_1 + x_2 \leq 2 \\ 2x_1 - x_2 \leq 0 \end{cases}$$

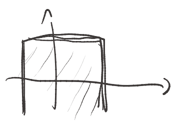
Remark In linear programs the minima\weak are the union of polyedra faces of X

Th: If f_i is continuous $\forall i$ $\wedge$ X is compact $\rightarrow$ $\exists$ minimum of P

Proof: $f(X)$ is compact if f is continuous and X compact

Th: If f_i is continuous $\forall i$, X is closed $\wedge$ $\exists \bar{v} \in \mathbb{R}$, $\exists j \in \{1, \dots, s\}$: $\{x \in X : f_j(x) \leq \bar{v}\}$ is non-empty $\wedge$ bounded $\rightarrow$ $\exists$ minima

↳ SUB-LEVEL SET (closed because X is closed $\wedge$ f continuous)



Corollary f_i continous $\forall i$, X is closed, f_j is coercive (bounded lw. set) for some j $\Rightarrow$ $\exists$ minimum

coercive / strongly convex

$$\begin{cases} \min \ (x_1 + x_2^2, \ (x_1-1)^2 + (x_2-1)^2) \\ x \in X \doteq \mathbb{R}^2_+ \end{cases}$$

· First thing I should do is check if the minimum $\exists$

Th: $x^* \in X$ is a minimum of P $\Leftrightarrow$

$$\begin{cases} \max\limits_{x,\varepsilon} \ \sum\limits_{i=1}^{s} \varepsilon_i \\ f_i(x) + \varepsilon_i \leq f_i(x^*) \quad \forall i \\ x \in X \\ \varepsilon \geq 0 \end{cases}$$

If I find such i with $\varepsilon_i > 0$ $\Rightarrow$ I improve

The optimal value must be $\varepsilon = 0$

Th: $x^* \in X$ is a weak minimum of P $\Leftrightarrow$

$$\begin{cases} \max \ v \\ v \leq \varepsilon_i \quad \forall i \\ f_i(x) + \varepsilon_i \leq f_i(x^*) \quad \forall i \\ x \in X \\ \varepsilon \geq 0 \end{cases}$$

I want the improvement in all i

The optimal value is equal to 0

UNCONSTRAINED MULTIOBJECTIVE PROBLEM

$$\begin{cases} \min \ f(x) = (f_1(x), \dots, f_s(x)) \\ x \in \mathbb{R}^n \end{cases}$$

where $f_i(x)$ is continously differentiable $\forall i$

If x^* is a weak minimum $\Rightarrow$ $\begin{cases} \nabla f_i(x^*)^T d < 0 \quad \forall i \\ d \in \mathbb{R}^n \end{cases}$ is impossible

Th (necessary condition): x^* is a weak minimum $\Rightarrow \exists \theta^* \in \mathbb{R}^s$: (x^*, θ^*) is a solution of:

$$\begin{cases} \sum_{i=1}^{s} \theta_i \nabla f_i(x) = 0 \\ \theta_i \geq 0 \quad \forall i \\ \sum \theta_i = 1 \\ x \in \mathbb{R}^n \end{cases}$$

Th (sufficient condition) P_u is convex (f_i convex $\forall i$) $\wedge$ (x^*, θ^*) solution of the prev. system $\Rightarrow$ 1) x^* is a weak minimum of P_u
2) $\theta^* > 0 \Rightarrow x^*$ is a minimum of P_u

$\rightarrow$ If the problem is convex $\wedge$ differentiable $\Rightarrow$ the system is necessary and sufficient condition

CONSTRAINED MULTIOBJECTIVE PROBLEM

$$\begin{cases} \min f(x) = (f_1(x), f_2(x), \dots, f_s(x)) \\ x \in X \doteq \{x \in \mathbb{R}^n : g_j(x) \leq 0,\ h_k(x) = 0\} \end{cases}$$

If x^* is a weak minimum of P $\wedge$ ACQ holds at x^* $\Rightarrow$ $\exists \theta^* \in \mathbb{R}^s, \lambda^* \in \mathbb{R}^m, \mu^* \in \mathbb{R}^p$ $(x^*, \theta^*, \lambda^*, \mu^*)$ solves the system:

$$\begin{cases} \sum \theta_i \nabla f_i + \sum \lambda_j \nabla g_j(x) + \sum \mu_k \nabla h_k(x) = 0 \\ \theta \geq 0 \quad \sum \theta_i = 1 \\ \lambda \geq 0 \quad \lambda_j g_j(x) = 0 \\ g(x) \leq 0 \quad h(x) = 0 \end{cases}$$

$\rightarrow$ if $X = \mathbb{R}^n$ (unconstrained) the system reduce to the previous one

Th: If x^* is a weak minimum $\Rightarrow$ $\begin{cases} \nabla f_i(x^*)^T d < 0 \quad \forall i \\ d \in T_X(x^*) \end{cases}$ has no solution

Corollary: x^* weak minimum $\wedge$ ACQ holds at x^* $\Rightarrow$

$$\begin{cases} v^T \nabla f_i(x^*) < 0 & \forall i \\ v^T \nabla g_j(x^*) \leq 0 & j \in A(x^*) \\ v^T \nabla h_k(x^*) = 0 & \forall k \\ v \in \mathbb{R}^n \end{cases}$$ Has no solution

$T_X = D_X$ since ACQ

Th: If P is convex

becomes sufficient for optimality

1) if $(x^*, \theta^*, \lambda^*, \mu^*)$ solves KKT $\to$ x^* is a weak minimum
2) if $(x^*, \theta^*, \lambda^*, \mu^*)$, $\theta^* > 0$ solves KKT $\Rightarrow$ x^* is a minimum

Th If x^* is the unique global minimum of f_k on X $\Rightarrow$ x^* is a minimum of P

↳ altrimenti cambio minimo e potrei diminuire una delle altre f

SCALARIZATION METHOD

$$\begin{cases} \min \ f(x) = (f_1(x), \dots, f_s(x)) \\ x \in X \end{cases}$$

$\alpha = (\alpha_1, \dots, \alpha_s)$, $\alpha_i \geq 0$, $\sum_i \alpha_i = 1$

consider the problem $\begin{cases} \min \ \sum \alpha_i f_i(x) \\ x \in X \end{cases}$

Let S_α be the set of optimal solution of P_α

Th: $\bigcup_{\alpha \geq 0} S_\alpha \subseteq \{\text{weak minima of } P\}$

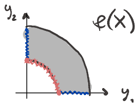
$\bigcup_{\alpha > 0} S_\alpha \subseteq \{\text{minima of } P\}$

Th: If X is convex set $\wedge$ f_i are convex on X $\Rightarrow$ $\{\text{weak minima}\} = \bigcup_{\alpha \geq 0} S_\alpha$

EXAMPLE

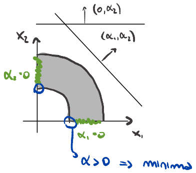
$$\begin{cases} \min \ (x_1, x_2) \\ x_1^2 + x_2^2 - 4 \leq 0 \\ -x_1^2 - x_2^2 + 1 \leq 0 \\ x_1 \geq 0, \quad x_2 \geq 0 \end{cases}$$

$\to$ not convex



$f(X) = X$ the image

$$\begin{cases} \min \ \alpha_1 x_1 + \alpha_2 x_2 \\ (x_1, x_2) \in X \end{cases}$$



$\bigcup_{\alpha \geq 0} S_\alpha = \{(0,1), (1,0)\}$

Th P linear (f_i linear $\wedge$ X a polyhedron) $\Rightarrow$ $\{\text{minima of } P\} = \bigcup_{\alpha > 0} S_\alpha$

- per $\alpha = 0$ non è detto che non abbia minimo, potrebbe essere

→ in LP unique solution is associated a dual solution non-degenerative
↓
of >0 is $>$ to the dimension of the space of Primal

dual → $\lambda \geq 0$ x the inequalities

If I found one that is not unique I can write the accessory conditions

$$\begin{aligned} \cancel{\lambda_1}^{1/2}(-2x_1 + x_2) &= 0 \\ \lambda_2^{0}(-x_1 - x_2) &= 0 \\ \cancel{\lambda_3}^{0}(5x_1 - x_2 - 6) &= 0 \end{aligned} \Rightarrow \begin{cases} x_2 = 2x_1 \\ -x_1 - x_2 \leq 0 \\ 5x_1 - x_2 - 6 \leq 0 \end{cases}$$

Th: If x^* is the unique global minimum of P_α for $\alpha \in [0,1]^s$ $\Rightarrow$ x^* is a minimum of P

IDEAL POINT $\quad z_i = \min\limits_{x \in X} f_i(x) \quad \forall i$

Often $z \notin f(X)$ $\Rightarrow$ I want to find the point as close as possible:

$$\begin{cases} \min \ \|f(x) - z\|_q \\ x \in X \end{cases} \qquad q \in [1, +\infty]$$

Th

1) If $q \in [1, +\infty)$ $\Rightarrow$ any optimal solution of G is a minimum of P
2) If $q = +\infty$ $\Rightarrow$ any optimal solution of G is a weak min of P

$\|\ \|_2$ and LP
GOAL METHOD

$$\begin{cases} \min \ Cx \\ Ax \leq b \end{cases} \qquad \begin{aligned} &C \in \mathbb{R}^{s \times n} \\ &A \in \mathbb{R}^{m \times n} \\ &b \in \mathbb{R}^{m} \end{aligned}$$

$$\begin{cases} \min \ \frac{1}{2}\|Cx - z\|_2^2 = \frac{1}{2}x^T C^T C x - x^T C^T z + \frac{1}{2} z^T z \\ Ax \leq b \end{cases}$$

↳ convex quadratic problem

GAME THEORY

is concerned with the analysis of conflictual situations involving different decision makers, called players

- The decision of each player is called strategy
- GT study the possibility to forecast the strategies that will be chosen by each player in order to minimize his cost

two players cannot agree on anything

A non-cooperative game, is set of N player, where each player has a cost f_i associated with a set of strategies X_i $f_i : X_1 \times \cdots \times X_N \longrightarrow \mathbb{R}$

$$\begin{cases} \min\limits_{x^i} f_i(x^1, \ldots, x^i, \ldots x^N) \\ x^i \in X_i \end{cases}$$

NASH EQUILIBRIUM

$$\begin{cases} \min f_1(x,y) \\ x \in X \end{cases} \qquad \begin{cases} \min f_2(x,y) \\ y \in Y \end{cases}$$

A pair of strategies $(\bar{x},\bar{y})$ is a Nash Equilibrium if

$f_1(\bar{x},\bar{y}) = \min\limits_{x \in X} f_1(x,\bar{y})$ $f_2(\bar{x},\bar{y}) = \min\limits_{y \in Y} f_2(\bar{x},y)$

- $\bar{x}$ is the best strategy of P1 in responce to strategy $\bar{y}$ of P2

A matrix game is a non-cooperative two-player game with:

1) X and Y are finite
2) $f_2 = -f_1$ (zero sum game)

Can be represented by a $m \times n$ matrix C where $f_1(i,j) = c_{ij}$ is the amount of money that P_1 pays to P_2 if P_1 choose strategy i and P_2 strategy j

Given a two-player non-coop game, a strategy $x \in X$ is strictly dominated by $\bar{x} \in X$ if:

$f_1(x,y) > f_1(\bar{x},y) \quad \forall y \in Y$

similar y is strictly dominated by $\bar{y}$ if: $f_2(x,y) > f_2(x,\bar{y}) \quad \forall x \in X$

they are both to be minimized

I can find NE by mapping the best strategy for each of the other player. If overlapping $\Rightarrow$ NE

Th if (i,j) and (p,q) are NE of a matrix game $\Rightarrow$

1) $c_{ij} = c_{pq}$
2) also (i,q) $\wedge$ (p,j) are NE

Dim

$$c_{ij} = \min f(x,j) = \max -f(i,y) = f(i,j)$$

$$c_{ij} = \max_{y} f(i,y) \geq c_{iq} \geq c_{pq} = \min_{x} f(x,q) \geq c_{pj} \geq c_{ij}$$

$$\begin{tikzcd}
c_{ij} \arrow[r, "\text{max on row}"] & c_{iq} \arrow[d, "\text{min on col}"] \\
c_{pj} \arrow[u, "\text{min on col}"] & c_{pq} \arrow[l, "\text{max on row}"']
\end{tikzcd}$$

$\rightarrow$ in odd and even

	O	E
O	1	-1
E	-1	1

NE doesn't exist

In a matrix game C, a mixed strategy for P1 is a m-vector of probabilities $X = \{ x \in \mathbb{R}^m : x \geq 0, \ \sum x_i = 1 \}$

The vertex of $X = \{0, \dots, 1, \dots, 0\}$ is a pure strategy for P1

The expected cost are $f_1(x,y) = x^T C y$ $\qquad f_2(x,y) = -x^T C y$

$$x^T C y = \sum_{i=1}^{m} \sum_{j=1}^{n} x_i c_{ij} y_j$$

- peso la strategia per la probabilità che venga giocata

- Pure strategies are considered in expected cost

- $(\bar{x}, \bar{y}) \in X \times Y$ is a mixed strategy NE if $\rightarrow$ SADDLE POINT

$$\max_{y \in Y} \bar{x}^T C y = \bar{x}^T C \bar{y} = \min_{x \in X} x^T C \bar{y}$$

SADDLE POINT: $F(\bar{x}, y) \leq F(\bar{x}, \bar{y}) \leq F(x, \bar{y})$ $\qquad \phi(y) \leq F(x,y) \leq \psi(x) \quad \forall x, y$

$\psi(x) \doteq \sup_{y \in Y} F(x,y) \quad x \in X$ $\qquad \phi(y) \doteq \inf_{x \in X} F(x,y) \quad y \in Y$

Th: $(\bar{x},\bar{y}) \in X \times Y$ is a saddle point $\iff$
1) $\bar{x}$ is an optimal sol of $\min_x \psi(x)$
2) $\bar{y}$ is an optimal sol of $\max_y \phi(y)$
3) $\psi(\bar{x}) = \phi(\bar{y})$

$$\Rightarrow \quad \min_{x\in X} \sup_{y \in Y} F(x,y) = \max_{y\in Y} \inf_{x\in X} F(x,y) = F(\bar{x},\bar{y})$$

EXISTANCE of SP

$X \subseteq \mathbb{R}^m$, $Y \subseteq \mathbb{R}^n$

If
1) X and Y are non-empty compact convex sets
2) $F(\cdot, y)$ is continous and convex (quasi) $\forall y$
3) $F(x, \cdot)$ is continous and concave (quasi) $\forall x$

$\Rightarrow$ F admits a saddle point on $X \times Y$

$x^T C y = F(x,y)$
- that is linear fixed or x or y (convex and concave)
- also continuity is fulfilled
- also X and Y are compact (simplex)

$\Rightarrow$ MSNE always exist (at least one)

$\bar{x}$ is an optimal sol of $\min_{x\in X} \max_{y\in Y} x^T C y$

$\bar{y}$ is an optimal sol of $\max_{y\in Y} \min_{x\in X} x^T C y$

- with optimal value that coincide to $\bar{x} C \bar{y}$

Th: $\min_{x\in X} \max_{y\in Y} x^T C y \iff$ LP $\max \min x^T C y \iff$ LP

$$P1 \begin{cases} \min_{v,x} \; v \\ v \geq \sum_{i=1}^{m} c_{ij} x_i & \forall j \\ x \geq 0 \\ \sum_{i=1}^{m} x_i = 1 \end{cases} \qquad P2 \begin{cases} \max \; w \\ w \leq \sum_{j=1}^{n} c_{ij} y_j & \forall i \\ y \geq 0 \\ \sum y_j = 1 \end{cases}$$

P1 is the dual of P2

$\Rightarrow$ $\bar{v} = \bar{w}$ (since is a linear programming problem)

BIMATRIX GAME : two person game , $X = \{ x \in \mathbb{R}^n, x \geq 0 \; \sum x_i = 1 \}$ — finite set of strategies

but $f_1 \neq -f_2$ (non zero-sum game) $\qquad C_1, C_2 \in \mathbb{R}^{m \times n}$

$f_1 = x^T C_1 y$ → bilinear cost functions

Th : A bimatrix game has at least one mixed strategy game

PRISONER DILEMMA → they are giulty of a robbery and a severe crime (evidence for robbery but not for the severe crime)

S1 : confess $\qquad$ S2 : stay quiet

- if both stay quiet ⇒ 2years each
- if both confess ⇒ 5 years each
- if only one confess and the other no ⇒ one 1y the other 10y

$$
C_1 = \begin{array}{c} \\ \\ \end{array}\begin{array}{cc} S_1 & S_2 \end{array}
$$

$$
C_1 = \begin{pmatrix} 5 & 1 \\ \cancel{10} & \cancel{2} \end{pmatrix} \begin{array}{c} S_1 \\ S_2 \end{array} \qquad C_2 = \begin{pmatrix} 5 & \cancel{10} \\ 1 & \cancel{2} \end{pmatrix}
$$

- There is a NE that is both to confess since if P_1 choose S_1 ⇒ P_2 choose S_1

- NE is optimal since I don't know what the other will do If i confess I'm sure to get or 1 or 5 but not 10
  → there is not cooperation

EXAMPLE $\qquad C_1 = \begin{pmatrix} -5 & 0 \\ 0 & -1 \end{pmatrix} \qquad C_2 = \begin{pmatrix} -1 & 0 \\ 0 & -5 \end{pmatrix}$

- No strictly dominated strategy
- $\begin{pmatrix} -5 & 0 \\ 0 & -1 \end{pmatrix} \quad \begin{pmatrix} -1 & 0 \\ 0 & -5 \end{pmatrix}$ → they correspond so $(1,1)$, $(2,2)$ are both nash equilibrium

Th (best response mapping) $\quad B_1 : Y \to X$ and $B_2 : X \to Y$

$$B_1(y) = \{ \text{optimal solution of } \min_{x \in X} x^T C_1 y \}$$

$$B_2(x) = \{ \text{optimal solution of } \min_{y \in Y} x^T C_2 y \}$$

given move x I want to know my best move

$(\bar{x},\bar{y})$ is a Nash equilibrium $\iff$ $\bar{x}\in B_1(\bar{y}) \wedge \bar{y}\in B_2(\bar{x})$

EX CONTINUED

given $y\in Y$ we need to solve $\begin{cases}\min x^T C_1 y = -5x_1y_1 - x_2y_2 \\ x\in X\end{cases}$

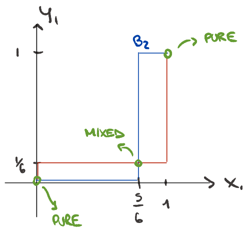
$x_2 = 1-x_1$
$y_2 = 1-y_1$

$$\iff \begin{cases}\min\limits_{x} \overbrace{(1-6y_1)}^{\text{if } >0} x_1 + y_1 + 1 \quad (=0) \\ 0\le x_1 \le 1\end{cases}$$

→ Bello per l'altro

$$B_1(y_1) = \begin{cases} 0 & \text{if } y_1\in[0,\frac{1}{6}) \\ [0,1] & \text{if } y_1 = 1/6 \\ 1 & \text{if } y_1\in(\frac{1}{6},1]\end{cases} \qquad B_2(x_1) = \begin{cases} 0 & \text{if } x_1\in[0,\frac{5}{6}) \\ [0,1] & \text{if } x_1 = \frac{5}{6} \\ 1 & \text{if } x_1\in(\frac{5}{6},1]\end{cases}$$

x_1



$\Rightarrow$ There are 3 nash equilibrium

- $[(0,1),(0,1)] \wedge [(1,0),(1,0)]$ is pure
- $[(\frac{5}{6},\frac{1}{6}),(\frac{1}{6},\frac{5}{6})]$ is mixed

KKT conditions for BIMATRIX GAMES

$$P_1(y) = \begin{cases}\min\limits_x x^T C_1 y \\ \sum x_i = 1 \\ x\ge 0\end{cases} \qquad P_2 = \begin{cases}\min\limits_y x^T C_2 y \\ \sum y_i = 1 \\ y\ge 0\end{cases}$$

KKT for P_1: $\quad L(x,\lambda,\mu) = x^T C_1 y + \mu_1(\sum x_i - 1) - \overbrace{\sum \lambda_i x_i}^{x^T\lambda}$

$$\begin{cases}\nabla_x L(x,\lambda,\mu) = C_1 y + \mu_1 \bar{e}_m - \lambda = 0 & \longrightarrow \lambda = C_1 y + \mu_1 e_m \\ -x^T\lambda = 0 & \Rightarrow x_i(C_1 y + \mu_1 e_m)_i \quad \forall i \\ x\ge 0 \quad \sum x_i = 1 & \\ \lambda \ge 0 & \longrightarrow C_1 y + \mu_1 e_m \ge 0\end{cases}$$

$$\begin{cases} C_1 y + \mu_1 e_m \geq 0 \\ x \geq 0, \quad \sum x_i = 1 \\ x_i (C_1 y + \mu_1 e_m)_i = 0 \quad \forall i \end{cases} \quad \wedge \quad \begin{cases} C_2^T x + \mu_2 e_n \geq 0 \\ y \geq 0 \quad \sum y_i = 1 \\ y_j (C_2^T x + \mu_2 e_n)_j = 0 \quad \forall j \end{cases}$$

complementarity conditions

· Since given x or y the two are linear → KKT necessary and sufficient

⟹ I can solve it together to obtain Nash Equilibrium

Th: $(\bar{x}, \bar{y})$ is a NE $\iff \exists (\mu_1, \mu_2) \in \mathbb{R}^2$:

$\begin{cases} KKT_1 \\ KKT_2 \end{cases}$ are solution for the system

$m+n+2$ variables

PREV EXAMPLE

$$C_1 = \begin{pmatrix} -5 & 0 \\ 0 & -1 \end{pmatrix} \qquad C_2 = \begin{pmatrix} -1 & 0 \\ 0 & -5 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & -5 \end{pmatrix}^T$$

$$KKT_1 \begin{cases} \begin{pmatrix} -5y_1 \\ -y_2 \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_1 \end{pmatrix} \geq 0 \\ x \in X \quad (x \geq 0 \ \wedge \ \sum x_i = 1) \\ x_1 (-5y_1 + \mu_1) = 0 \\ x_2 (-y_2 + \mu_1) = 0 \end{cases} \qquad KKT_2 \begin{cases} \begin{pmatrix} -x_1 \\ -5x_2 \end{pmatrix} + \begin{pmatrix} \mu_2 \\ \mu_2 \end{pmatrix} \geq 0 \\ y \in Y \\ y_1 (-x_1 + \mu_2) = 0 \\ y_2 (-5x_2 + \mu_2) = 0 \end{cases}$$

Can be solved considering the cases $\begin{cases} x_1 = 0 \\ x_1 = 1 \\ 0 < x_1 < 1 \end{cases}$

How to solve it in matlab → I need an optimization problem

$$\begin{cases} C_1 \bar{y} + \mu_1 e_m \geq 0 \\ \bar{x} \geq 0, \quad \sum x_i = 1 \\ \underbrace{\bar{x}^T}_{\geq 0} \underbrace{(C_1 \bar{y} + \mu_1 e_m)}_{\geq 0} = 0 \\ C_2^T \bar{x} + \mu_2 e_n \geq 0 \\ \bar{y} \geq 0 \quad \sum y_i = 1 \\ \bar{y}^T (C_2^T \bar{x} + \mu_2 e_n) = 0 \end{cases}$$

→ then just a scalar product

All the feasible solution x and y imply that $C_1 y + \mu_1 e_m \geq 0$
$x \geq 0$

Then I can minimize $\Psi(x,y,\mu_1,\mu_2) = x^T(C_1 y + \mu_1 e_m) + y^T(C_2^T x + \mu_2 e_n)$

QUADRATIC PROBLEM

with the constraints :
$$\begin{cases} C_1\bar{y} + \mu_1 e_m \geq 0 \\ \bar{x} \geq 0, \quad \sum x_i = 1 \\ C_2^T\bar{x} + \mu_2 e_n \geq 0 \\ \bar{y} \geq 0 \quad \sum y_i = 1 \end{cases}$$

I know that a solution, if exists, will have $\Psi(\bar{x},\bar{y},\mu_1,\mu_2) = 0$

↳ will exists for sure since ∃ nash equilibriums

$$\nabla\Psi(x,y,\mu_1,\mu_2) = \begin{pmatrix} C_1 y + \mu_1 e_m + C_2 y \\ C_1^T x + C_2^T x + \mu_2 e_n \\ e_m^T x \\ e_n^T y \end{pmatrix} \begin{array}{l} \text{wrt } x \\ \text{wrt } y \\ \text{wrt } \mu_1 \\ \text{wrt } \mu_2 \end{array}$$

Since we know that in $f(x) = \frac{1}{2}x^T H x$ $\quad \nabla f = Hx$

$$H = \begin{pmatrix} 0_{m\times n} & C_1+C_2 & e_m & 0_{m\times 1} \\ C_1^T + C_2^T & 0_{n\times n} & 0_{n\times 1} & e_n \\ e_m^T & 0_{1\times n} & 0 & 0 \\ 0_{1\times m} & e_n^T & 0 & 0 \end{pmatrix}$$

$$A_{in} = \begin{pmatrix} -C_2^T & 0_{n\times n} & 0_{n\times 1} & -e_n \\ 0_{m\times m} & -C_1 & -e_m & 0_{m\times 1} \end{pmatrix} \qquad b_{in} = \begin{pmatrix} 0_{n\times 1} \\ 0_{m\times 1} \end{pmatrix}$$

$$A_{eq} = \begin{pmatrix} 1 & \dots & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & & 0 & 1 & \dots & 1 & 0 & 0 \end{pmatrix} \qquad b_{eq} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

→ not quadprog since it's not convex
- not sure how many there are → MULTI-START APPROACH

EXAMPLE

$$C_1 = \begin{pmatrix} 3 & 3 \\ 4 & 1 \\ 6 & 0 \end{pmatrix} \qquad C_2 = \begin{pmatrix} 3 & 4 \\ 4 & 0 \\ 3 & 5 \end{pmatrix}$$

rats(solution) → from $0.\bar{3}$ → $\frac{1}{3}$

CONVEX PROBLEMS

general functions

$P_1: \begin{cases} \min_x f_1(x,y) \\ g_i^1(x) \leq 0 \quad \forall i \end{cases}$

$P_2: \begin{cases} \min_y f_2(x,y) \\ g_j^2(y) \leq 0 \quad \forall j \end{cases}$

- f_1, f_2, g_1, g_2 are continously diff
- f_1, f_2 convex wrt X and Y

X is the set of feasible strategies → closed, convex, bounded

$f_1(\cdot, y)$ is convex $\forall y \in Y \wedge f_2(x, \cdot)$ is convex $\forall x \in X$

Th If $(\bar{x}, \bar{y})$ is NE $\wedge$ ACQ holds $\Rightarrow \exists \lambda^1 \in \mathbb{R}^p, \lambda^2 \in \mathbb{R}^q$:

$$\begin{cases} \nabla_x f_1(\bar{x}, \bar{y}) + \sum \lambda_i^1 \nabla g_i^1(\bar{x}) = 0 \\ \lambda^1 \geq 0 \qquad g^1(\bar{x}) \leq 0 \\ \lambda_i^1 g_i^1(\bar{x}) = 0 \qquad \forall i \\ \nabla_y f_2(\bar{x}, \bar{y}) + \sum \lambda_j^2 \nabla g_j^2(\bar{y}) = 0 \\ \lambda^2 \geq 0 \qquad g^2(\bar{y}) \leq 0 \\ \lambda_j^2 g_j^2(\bar{y}) = 0 \qquad \forall j \end{cases}$$

- Convexity is necessary so that KKT are both necessary and sufficient

→ It's a generalization of bimatrix games