PROBLEMA DI OTTIMIZZAZIONE

SI DEFINISCE UN PROBLEMA DI OTTIMIZZAZIONE UN PROBLEMA CON LA SEGUENTE STRUTTURA:

$$\boxed{\min_{x \in S} f(x)}$$

DOVE:

- x SONO LE VARIABILI
- S È L'INSIEME AMMISSIBILE DELLE VARIABILI
- $f(x)$ È LA FUNZIONE DA OTTIMIZZARE TALE CHE

$$f : S \longmapsto \mathbb{R}$$

INOLTRE È POSSIBILE SCRIVERE, IN MANIERA ANALOGA, CHE:

$$\boxed{\min_{x} f(x) = \max_{x} -f(x)}$$

ESEMPIO: (PROBLEMA DEL PORTAFOGLIO)

IL PROBLEMA CONSISTE NELLA SCELTA DI COME INVESTIRE UNA CERTA QUANTITÀ DI DENARO IN UN INSIEME DI POSSIBILI OPZIONI, AL FINE DI MASSIMIZZARE L'OBIETTIVO DESIDERATO.

SI HANNO n ASSET ASSOCIATI AD UN CERTO RITORNO ATTESO μ DOVE L'ASSET i-ESIMO AVRÀ QUINDI IL RITORNO ATTESO μ_i

DATI QUINDI DUE DIVERSI ASSET CHIAMATI i, j SI HA CHE

σ_{ij} È LA COVARIANZA, OVVERO UNA MISURA CHE INDICA LA CORRELAZIONE TRA I DUE DIVERSI ASSET.

IN PARTICOLARE SI HA CHE $\sigma_{ij} \in [\underbrace{-1}_{\text{COMPLETA SCORRELAZIONE TRA GLI ASSET}} ; \underbrace{1}_{\text{COMPLETA CORRELAZIONE TRA GLI ASSET}}]$

INFINE SI HA UN CERTO BUDGET M. DEFINIAMO IL PROBLEMA DI OTTIMIZZAZIONE COME SEGUE:

- VARIABILI: $x_1, \dots, x_n$ DOVE x_i È LA QUANTITÀ DI BUDGET INVESTITA SULL'ASSET i-ESIMO

- VINCOLI: 1) $x_i \geq 0 \quad \forall i = 1, \dots, n$

② $\sum_{i=1} x_i = M$ OVVERO SI DEVE SPENDERE TUTTO IL BUDGET CHE SI HA

$$\bullet\ f(\underline{x}) = \underbrace{-\sum_{i=1}^{m} \mu_i x_i}_{\text{RITORNO, OVVERO IL VALORE ATTESO}} + \underbrace{\lambda}_{\text{PARAMETRO PER PESARE IL RISCHIO/RITORNO}} \underbrace{\sum_{i=1}^{m} \sum_{j=1}^{m} \sigma_{ij} x_i x_j}_{\text{RISCHIO, OVVERO IN BASE ALLA CORRELAZIONE DEGLI ASSET SI VARIA IL LORO VALORE}}$$

QUINDI POSSIAMO SCRIVERE IL PROBLEMA DI OTTIMIZZAZIONE COME:

$$\boxed{\begin{array}{l} \min -\sum_i \mu_i x_i + \lambda \sum_{i,j} \sigma_{ij} x_i x_j \quad \text{TALE CHE} \\ \bullet\ x_i \geq 0 \quad \forall i \\ \bullet\ \sum_i x_i = M \end{array}}$$

UTILIZZANDO LA NOTAZIONE VETTORIALE SI OTTIENE, IN MANIERA PIÙ COMPATTA:

$$\boxed{\begin{array}{l} \min -\mu^t x + \lambda x^t Q x \quad \text{CON } Q_{ij} = \sigma_{ij} \text{ TALE CHE:} \\ \bullet\ x \geq 0 \\ \bullet\ x^t e = M \quad \text{CON } e = (1, \dots, 1) \in \mathbb{R}^m \end{array}}$$

- $x_i = \dots$ dove $i \in \{1, \dots, n\} \subseteq \mathbb{R}$
- $x \in \mathbb{R}^n$

CLASSIFICAZIONE PROBLEMI OTTIMIZZAZIONE

PROBLEMA OTTIMIZZAZIONE

- → Lineare ⇒ $f(x) = c^T x$, $S = \{x \mid Ax \leq b\}$
- → $\boxed{\text{Non Lineare}}$
- → $\boxed{\text{Continui}}$ $(S \subseteq \mathbb{R}^n)$
  - → Misti Interi $(S \subseteq \mathbb{R}^{n-m} \times \mathbb{Z}^m)$
  - → Interi $(S \subseteq \mathbb{Z}^n)$
    - → Binari $(S \subseteq \{0;1\}^n)$

Continui:

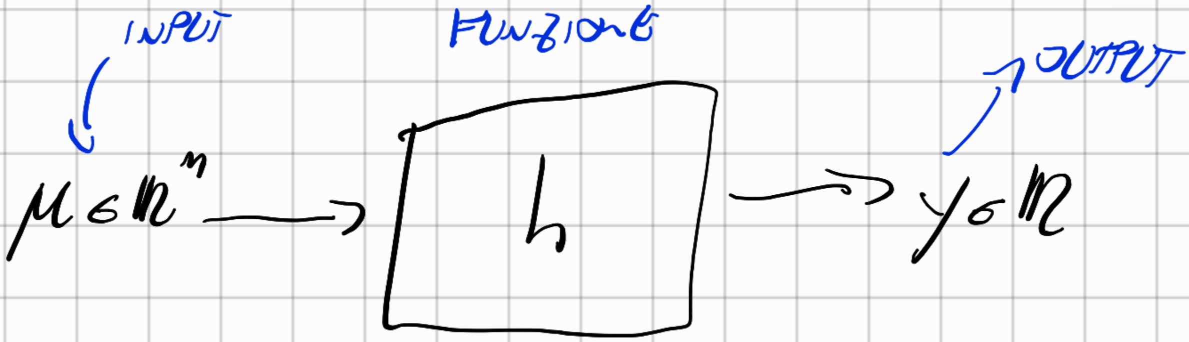
- $\boxed{\text{Vincolati}}$ $(S \subset \mathbb{R}^n)$
- $\boxed{\text{Non Vincolati}}$ $(S = \mathbb{R}^n)$

Non Lineare:

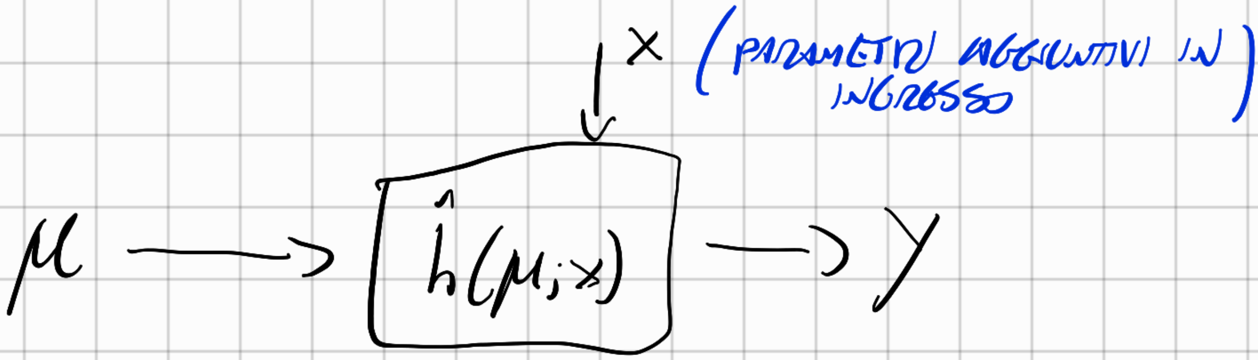
- → $\boxed{\text{Differenziabili}}$ $(f, \nabla f, \nabla^2 f)$
- → Non Differenziabili (f)

ESEMPIO: PROBLEMA NON VINCOLATO

CONSIDERIAMO UN GENERICO SISTEMA FISICO REALE:



VORREMO QUINDI RICAVARE DAL SISTEMA FISICO UN MODELLO MATEMATICO CHE LO APPROSSIMI IL PIÙ POSSIBILE SULLA BASE DELLE OSSERVAZIONI RICAVATE:



QUINDI SI ESEGUONO DELLE MISURE SUL SISTEMA FISICO:

μ	y
μ_1	y_1
$\vdots$	$\vdots$
μ_N	y_N

È, al fine di misurare la differenza tra il modello matematico ottenuto ed il sistema fisico effettivo, si calcola una **funzione di errore**

$$\boxed{e: \mathbb{R}\times\mathbb{R} \longmapsto \mathbb{R}^+}$$

la quale va a valutare la differenza tra $y_i = h(\mu_i)$ e $\hat{h}(\mu_i; x)$

Si va quindi a costruire il problema di ottimizzazione come un problema di **minimizzazione della funzione di errore**:

$$\boxed{\min_x \sum_{i=1}^{N} e\left(\hat{h}(\mu_i, x), y_i\right)}$$

dove e è una funzione di errore che può essere di vari tipi.

Richiami sulla norma

La norma è una funzione $\|\cdot\|: \mathbb{R}^n \longmapsto \mathbb{R}^+$ t.c.:

1) $\|x\| \geq 0 \quad \forall x \in \mathbb{R}^n$

2) $\|x\| = 0 \iff x = 0$

3) $\|x + y\| \leq \|x\| + \|y\| \quad \forall x, y \in \mathbb{R}^n$

4) $\|\alpha x\| = \alpha \|x\| \quad \forall x \in \mathbb{R}^n, \forall \alpha \geq 0$

ESEMPI:

- $\|x\|_2 = \sqrt{\sum_{i=1}^{n} x^2} = \sqrt{x^t x} \rightarrow$ NORMA EUCLIDEA
- $\|x\|_1 = \sum_{i=1}^{n} |x_i| \rightarrow$ NORMA UNO
- $\|x\|_\infty = \max_{i=1,\dots,n} |x_i| \rightarrow$ NORMA INFINITO

QUINDI POSSIAMO RISCRIVERE LA FORMULAZIONE DEL PROBLEMA PRECEDENTE ANDANDO A MINIMIZZARE LA NORMA IN MODO DA POTERNE SFRUTTARE LE PROPRIETÀ:

$$\boxed{\min_x \|E(x)\|^{(2)}}$$ DOVE:

$$E = \begin{pmatrix} e(\hat{h}(\mu_1, x), y_1) \\ \vdots \end{pmatrix}$$

$$E(x) = \begin{pmatrix} \vdots \\ e(\hat{h}(\mu_N, x), y_N) \end{pmatrix}$$

INTRODUZIONE ALL'APPRENDIMENTO SUPERVISIONATO

L'OBIETTIVO È DI TROVARE UNA FUNZIONE DI MODELLO INCOGNITA h:

$$\boxed{h: \mathbb{R}^n \longmapsto Y}$$

DOVE:

- $Y \in \mathbb{R}$ → PROBLEMA DI REGRESSIONE
- $Y \in \mathbb{Z}$ → PROBLEMA DI CLASSIFICAZIONE

OVVIAMENTE SI CERCA DI APPROSSIMARE LA h EFFETTIVA ANDANDO A COSTRUIRE UNA FUNZIONE APPROSSIMANTE

$$\boxed{\hat{h} \in H_W}$$

APPARTENENTE AD UN CERTO SPAZIO DI FUNZIONI APPROSSIMANTI CON PARAMETRI CARATTERIZZANTI W.

QUINDI OTTERREMO CHE:

$$\boxed{\hat{h}(\underbrace{x}_{\text{INPUT DEL PROBLEMA}},\underbrace{w}_{\text{PARAMETRO DA ALLENARE}}) \approx h(x) = y \quad \forall x \in \mathbb{R}^n}$$

SI AVRÀ QUINDI UN CERTO DATASET DI MISURAZIONI

$$\boxed{D = \{(x_i, y_i) \mid y_i = h(x_i)\}}$$

E L'OBIETTIVO È QUINDI QUELLO DI MINIMIZZARE L'ERRORE ATTESO, OVVERO:

$$\boxed{\min_{w} E_{x,y}\left[\underbrace{e(\hat{h}(x,w), y)}_{\text{L'ERRORE COMMESSO DALLA FUNZIONE APPROSSIMATA } \hat{h} \text{ RISPETTO ALL'OUTPUT EFFETTIVO DATO DA } y}\right]}$$

TUTTAVIA NELLA PRATICA L'OBIETTIVO SARÀ QUELLO DI MINIMIZZARE IL RISCHIO EMPIRICO, IL QUALE È UNA

MISURA DI QUANTO BENE IL MODELLO SI COMPORTA SUI DATI DI ADDESTRAMENTO:

$$\boxed{\min_{w} \sum_{i=1}^{N} e\left(\hat{h}(x_i, w), y_i\right)}$$

SI CONSIDERA SUI DATI CHE SONO STATI MISURATI, OVVERO QUELLI PRESENTI NEL DATASET.

DOVE IL RISCHIO EMPIRICO È COSTITUITO DA:

$$\boxed{e = \underbrace{Loss}_{\text{ERRORE DI APPROSSIMAZIONE}} + \underbrace{\Omega(w)}_{\text{PARAMETRO CHE VALUTA I MODELLI PIÙ SEMPLICI E PIÙ ADATTABILI}}}$$

QUINDI L'OBIETTIVO DI UN PROBLEMA DI OTTIMIZZAZIONE È:

$$\boxed{\min_{w} \underbrace{\mathcal{L}(w, X, y)}_{\text{LOSS}} + \underbrace{\lambda\, \Omega(w)}_{\text{REGOLARIZZATORE}}}$$

DOVE $$\boxed{\mathcal{L}(x, X, y) = \frac{1}{N} \sum_{i=1}^{N} \ell(w, x_i, y_i)}$$

NEL CASO PROBLEMI DI REGRESSIONE SI HANNO LE SEGUENTI LOSS:

- ERRORE QUADRATICO: $$\boxed{\ell(w, x_i, y_i) = (h_w(x_i) - y_i)^2}$$
- ERRORE ASSOLUTO: $$\boxed{\ell(w, x_i, y_i) = |h_w(x_i) - y_i|^2}$$

MENTRE, NEL CASO DI PROBLEMI DI CLASSIFICAZIONE SI HANNO LE SEGUENTI LOSS:

$$\begin{cases} 1 & y_i \neq h_w(x_i) \\ 0 & \text{ALTRIMENTI} \end{cases}$$

- 0/1-LOSS: $$\boxed{\mathbb{1}\{y_i \neq h_w(x_i)\}}$$

VALUTA QUANTO VICINO È IL PUNTO RISPETTO ALLA LINEA DI SEPARAZIONE

- HINGE-LOSS: $$\boxed{\max\{0, 1 - y_i h_w(x_i)\}}$$
- LOG-LOSS: $$\boxed{\log(1 + e^{-y_i h_w(x_i)})}$$

DEFINIZIONE DI OTTIMO

RISOLVERE UN PROBLEMA $\min_{x \in S} f(x)$ SIGNIFICA TROVARE UN PUNTO DI OTTIMO GLOBALE, OVVERO UN PUNTO $x^* \in S$ CHE SODDISFA LA PROPRIETÀ $\boxed{f(x^*) \leq f(x)}$ $\forall x \in S$

IL VALORE $f(x^*)$ È DEFINITO OTTIMO GLOBALE.

TUTTAVIA NON È SEMPRE VERO CHE UNA SOLUZIONE AL PROBLEMA ESISTA SEMPRE. IN PARTICOLARE NON SI HA SOLUZIONE QUANDO:

- $f(x)$ È ILLIMITATA INFERIORMENTE
- $S = \emptyset$ => L'INSIEME AMMISSIBILE È VUOTO!

OCCORRE QUINDI DEFINIRE LE CONDIZIONI DI ESISTENZA DELLA SOLUZIONE.

RICHIAMI DI TOPOLOGIA

$B_\rho(x) = \{y \mid \|x-y\| \leq \rho\}$

. UN INSIEME S SI DICE APERTO SE $\boxed{\forall x \in S \ \exists \rho > 0 \text{ t.c. } B_\rho(x) \subseteq S}$

- UN INSIEME S SI DICE CHIUSO SE IL SUO COMPLEMENTARE È APERTO

- UN INSIEME S È LIMITATO SE $\boxed{\exists M>0 \text{ t.c. } \|x\| \le M \ \forall x \in S}$

- UN INSIEME S È COMPATTO SE VALGONO LE SEGUENTI:

  1. $\forall \{x^n\} \subseteq S$ ESISTE UNA SOTTOSEQUENZA CONVERGENTE, OVVERO $\exists K \subseteq \{0,1,\dots\}$ t.c. $\boxed{\lim_{\substack{n\in K \\ n\to\infty}} x^n = \bar{x}}$

  2. $\forall \{x^n\} \subseteq S$ OGNI PUNTO DI ACCUMULAZIONE, CIOÈ OGNI PUNTO LIMITE DI UNA SOTTOSEQUENZA, APPARTIENE AD S, OVVERO SE $K \subseteq \{0,1,\dots\}$ È TALE CHE SE $\boxed{\lim_{\substack{n\in K \\ n\to\infty}} x^n = \bar{x} \text{ ALLORA } \bar{x} \in S}$

VALGONO INOLTRE LE SEGUENTI PROPRIETÀ:

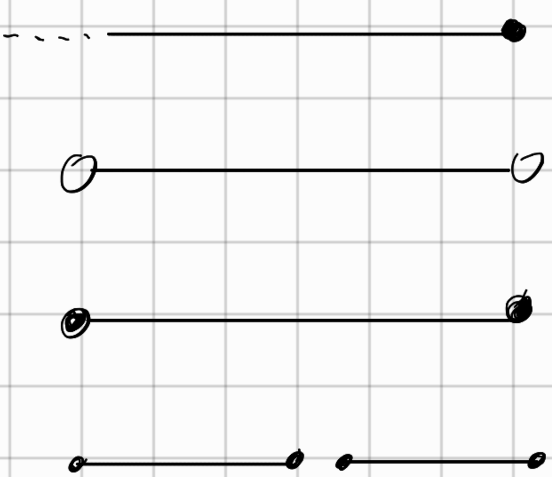
- SE $S \subseteq \mathbb{R}^n$ ALLORA S È CHIUSO $\Leftrightarrow$ VALE (2)

S È LIMITATO $\Leftrightarrow$ VALE (1)

• $S \subseteq \mathbb{R}^n$ è compatto $\Leftrightarrow$ è chiuso e limitato

in $\mathbb{R}$:

insiemi / unioni di intervalli



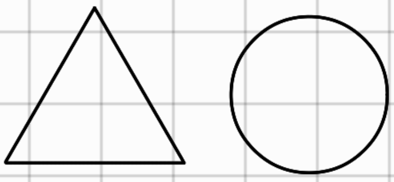
chiuso ma non limitato

limitato ma non chiuso

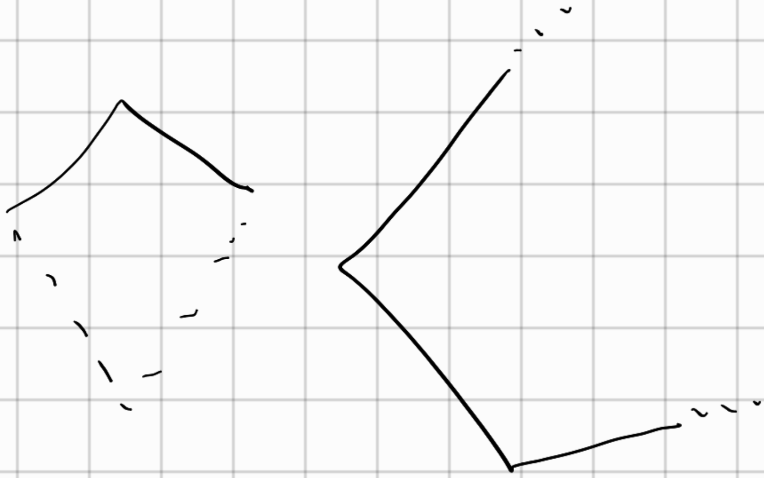
compatto

compatto

in $\mathbb{R}^2$:



compatti



non compatti

TEOREMA DI WEIERSTRASS

sia $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ continua e sia $S \subseteq \mathbb{R}^n$ compatto.

ALLORA f AMMETTE MINIMO SU S

Dim:

L'IMMAGINE DI S TRAMITE f

SIA $Y = f(S) = \{ y \mid \exists x \in S \text{ tc: } f(x) = y \}$

CHIAMO $L = \inf(Y)$, ALLORA $\exists \{y^n\} \subseteq Y$ TALE CHE

$\lim_{\substack{n \in \mathbb{N} \\ n \to \infty}} y^n = L$ PER DEFINIZIONE.

LA SEQUENZA DEVE CONVERGERE ALL'ESTREMO INFERIORE PER DEFINIZIONE

DEFINISCO ALLORA $\{x^n\} \subseteq S$ TALE CHE $\boxed{f(x^n) = y^n}$

MA PER IPOTESI S È COMPATTO E QUINDI È ANCHE LIMITATO

$\Rightarrow \exists \mathbb{N} \in \{0, 1, \dots\}$ TALE CHE $\boxed{\lim_{\substack{n \in \mathbb{N} \\ n \to \infty}} x^n = \bar{x}}$ MA S È

1

OVVIAMENTE ANCHE CHIUSO $\Rightarrow$ $\boxed{\bar{x} \in S}$ 2

ALLORA SE CONSIDERO:

PER LA CONTINUITÀ DI f

$$L = \lim_{\substack{n \in \mathbb{N} \\ n \to \infty}} y^n = \lim_{\substack{n \in \mathbb{N} \\ n \to \infty}} f(x^n) \overset{3}{=} f\left(\lim_{\substack{n \in \mathbb{N} \\ n \to \infty}} x^k \right) = f(\bar{x})$$

$\Rightarrow f(\bar{x}) = l$

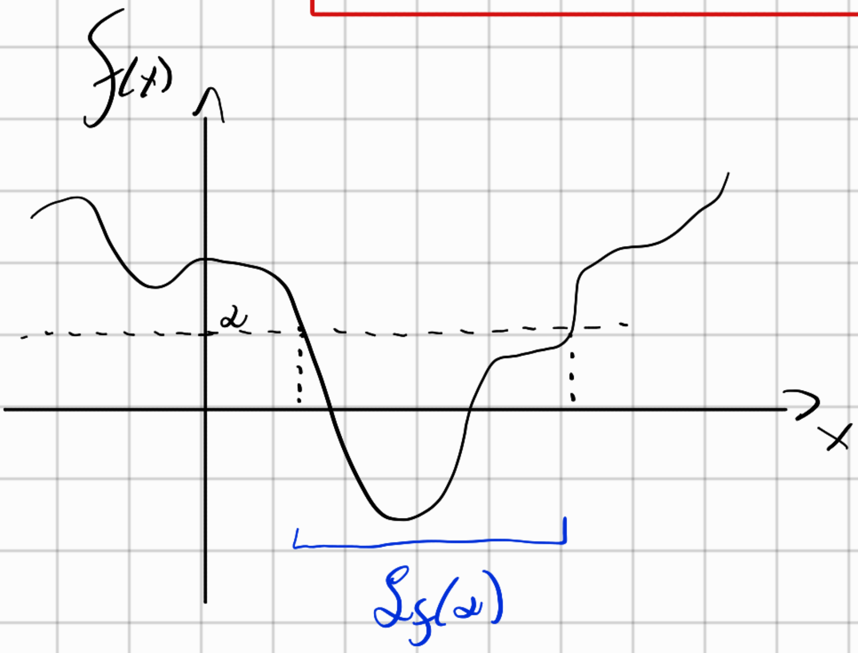
QUINDI $\bar{x}$ È UN PUNTO DI MINIMO GLOBALE POICHÉ

$$\boxed{\bar{x} \in S \text{ e } f(\bar{x}) \leq y \quad \forall y \in Y = f(S)} \qquad \square$$

INSIEME DI LIVELLO

SIA $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ E SIA $\alpha \in \mathbb{R}$. L'INSIEME DI LIVELLO $\mathcal{L}_f(\alpha)$ È DEFINITO COME

$$\boxed{\mathcal{L}_f(\alpha) = \{ x \in \mathbb{R}^n \mid f(x) \leq \alpha \}}$$



SIA $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ CONTINUA E SUPPONIAMO CHE f ABBIA un insieme di livello $\boxed{\mathcal{L}_f(\alpha) \text{ compatto}}$, allora f ammette minimo su $\mathbb{R}^n$

Dim

Considero il problema $\boxed{\min_{x \in \mathcal{L}_f(\alpha)} f(x)}$. Per il teorema di Weierstrass si ha che il problema ha una soluzione ottimale x^*. Considero allora un qualsiasi punto $z \in \mathbb{R}^n$. Si distinguono due casi:

- $z \in \mathcal{L}_f(\alpha) \Rightarrow \boxed{f(x^*) \leq f(z)}$ poiché x^* minimo di f su $\mathcal{L}_f(\alpha)$
- $z \notin \mathcal{L}_f(\alpha) \Rightarrow \boxed{f(z) > \alpha \geq f(x^*)}$ per definizione di insieme di livello

Allora $\forall z \in \mathbb{R}^n$ vale che $\boxed{f(x^*) \leq f(z)}$ □

Definizione funzione coerciva

Sia $f: \mathbb{R}^n \longmapsto \mathbb{R}$ una funzione. f è coerciva se:

$\forall \{x^k\} \subseteq \mathbb{R}^n$ tale che $\|x^k\| \to \infty$ ...

$\forall \{x^k\} \subseteq \mathbb{R}^n$ TALE CHE $\|x^k\| \to \infty$, ALLORA VALE

$$\boxed{\lim_{k\to\infty} f(x^k) = \infty}$$

SE f È CONTINUA ALLORA È COERCIVA SE E SOLO SE TUTTI I SUOI INSIEMI DI LIVELLO SONO COMPATTI

QUINDI SE f È COERCIVA E CONTINUA ALLORA AMMETTE MINIMO SU $\mathbb{R}^n$

$$\boxed{f \text{ COERCIVA E CONTINUA} \Rightarrow \text{AMMETTE MINIMO SU } \mathbb{R}^n}$$

RICHIAMO SULLE PROPRIETÀ MATRICI

SIA $Q \in \mathbb{R}^{n\times n}$ UNA MATRICE QUADRATA. Q È DETTA:

- SEMIDEFINITA POSITIVA

SE $\forall x \in \mathbb{R}^n$ VALE CHE $\boxed{x^T Q x \geq 0}$

- **DEFINITA POSITIVA**

SE $\forall x \neq 0$ VALE CHE $\boxed{x^T Q x > 0}$

INOLTRE, SE Q È SIMMETRICA ALLORA Q È:

- SEMIDEFINITA POSITIVA $\Leftrightarrow$ HA TUTTI GLI AUTOVALORI NON NEGATIVI

- DEFINITA POSITIVA $\Leftrightarrow$ HA TUTTI GLI AUTOVALORI POSITIVI

- $\forall x \in \mathbb{R}^n$ $\boxed{x^T Q x \geq \lambda_{min} \|x\|^2}$ PER Q S.D.P

VALE INOLTRE LA **DISUGUAGLIANZA DI CAUCHY-SCHWARTZ**

$\forall c, x \in \mathbb{R}^n$ SI HA CHE $\boxed{|c^T x| \leq \|c\| \cdot \|x\|}$

DEFINIZIONE PROBLEMI QUADRATICI

SONO DELLA FORMA

$$\boxed{\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{2} x^T Q x - c^T x}$$

DOVE Q È UNA MATRICE SIMMETRICA QUADRATA E CON

SI HA CHE:

$$\boxed{f(x) = \frac{1}{2}x^TQx - c^Tx \text{ È COERCIVA } \Leftrightarrow Q \text{ È DEFINITA POSITIVA}}$$

Dim: SI DEVE DIMOSTRARE LA DOPPIA IMPLICAZIONE:

1) Q DEFINITA POSITIVA $\Rightarrow$ f È COERCIVA

CONSIDERO $f(x) = \frac{1}{2}x^TQx - c^Tx$ E APPLICO 4:

$$f(x) \geq \frac{1}{2}\lambda_{min}\|x\|^2 - c^Tx$$

$$\geq \frac{1}{2}\lambda_{min}\|x\|^2 - \|c\|\cdot\|x\|$$

CONSIDERO ALLORA $\{x^k\}$ TALE CHE $\|x^k\| \to \infty$

E CONSIDERO:

$$\lim_{k\to\infty} f(x^k) \geq \lim_{k\to\infty} \frac{1}{2}\lambda_{min}\|x^k\|^2 - \|c\|\cdot\|x^k\|$$

$$= \lim \|x^k\|\left(\frac{1}{2}\lambda_{min}\|x^k\| - \|c\|\right)$$

≥ 0 poiché φ è definita positiva

costante finita

$\to +\infty$

$\to +\infty$

$$= +\infty$$

$\to$ ovvero f è coerciva

Quindi ottengo che $\boxed{\lim_{k \to \infty} f(x^k) = +\infty}$ $\square$

---

② f è coerciva $\Rightarrow$ φ è definita positiva

Supponiamo, per assurdo, che φ non sia D.P, ovvero $\exists\, y \in \mathbb{R}^n$ con $y \neq 0$ t.c. $\boxed{y^T \varphi y \leq 0}$

Supponiamo allora che $\boxed{c^T y \geq 0}$ e costruiamo una sequenza $\{x_k\}$ tale che $\boxed{x_k = k \cdot y \ \forall k}$

vale che:

$$\boxed{\lim_{k \to \infty} \|x_k\| = \lim_{k \to \infty} k\|y\| = +\infty}$$

Considero $f(x_k) = \frac{1}{2} x_k^T \varphi x_k - c^T x_k$ (I)

$x_k = ky$

$$= \frac{1}{2}(ky)^T Q (ky) - c^T (ky)$$

$$= \frac{1}{2} k^2 y^T Q y - k c^T y$$

$$= k\left(\frac{1}{2} k y^T Q y - c^T y\right)$$

SI FA IL LIMITE

ALLORA:

$$\lim_{k\to\infty} f(x_k) = \lim_{k\to\infty} \underbrace{k}_{+\infty}\Big(\underbrace{\underbrace{\tfrac{1}{2}k}_{+\infty}\ \underbrace{y^T Q y}_{\leq 0}}_{\leq 0} - \underbrace{c^T y}_{\leq 0}\Big) < +\infty$$

OVVERO $\boxed{\lim_{k\to\infty} f(x_k) < +\infty \Rightarrow \text{NON È COERCIVA}}$ $\Rightarrow$ ASSURDO

$\square$

L'ULTIMA PROPRIETÀ IMPORTANTE È:

> SE Q È DEFINITA POSITIVA ALLORA $f(x) = \frac{1}{2} x^T Q x - c^T x$ AMMETTE MINIMO SU $\mathbb{R}^n$

$\hookrightarrow$, POICHÉ ABBIAMO DIMOSTRATO CHE

$\underbrace{f \text{ CA COERCIVA}}$ $\Leftrightarrow$ φ DEF POS

$\Downarrow$

AMMETTE MINIMO

SU $\mathbb{R}^n$