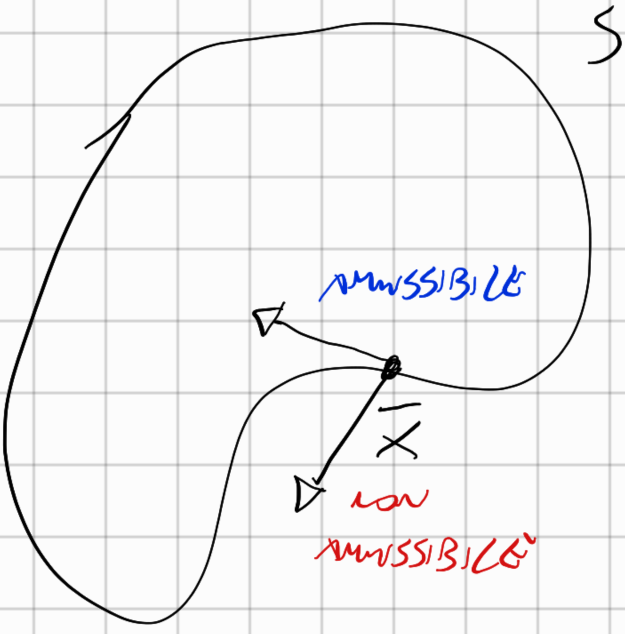
SI AFFRONTANO PROBLEMI DEL TIPO $\boxed{\min_{x\in S\subseteq \mathbb{R}^n} f(x)}$

QUESTO SIGNIFICA CHE QUANDO SI VA A VALUTARE UNA CERTA SOLUZIONE NON SI GUARDA PIÙ ESCLUSIVAMENTE AL SOLO VALORE OBIETTIVO MA OCCORRERÀ ANCHE CONSIDERARE L'AMMISSIBILITÀ DELLA SOLUZIONE TROVATA

DEFINIZIONE: DIREZIONE AMMISSIBILE

UNA DIREZIONE $d\in\mathbb{R}^n$ È AMMISSIBILE IN UN PUNTO $\bar{x}\in S$ SE $\exists\, \bar{t}>0$ TALE CHE:

$$\boxed{\bar{x}+td\in S \quad \forall t\in[0,\bar{t}]}$$



SE x^* È UN PUNTO DI MINIMO LOCALE, ALLORA NON ESISTE $d\in\mathbb{R}^n$ CHE SIA DI DISCESA E AMMISSIBILE IN x^*

CNO I° ORDINE

SE $f \in C^1(\mathbb{R}^n)$ ALLORA SE x^* È MINIMO LOCALE, NON ESISTE $d \in \mathbb{R}^n$ CHE È AMMISSIBILE IN x^* TALE CHE:

$$\boxed{\nabla f(x^*)^T d < 0}$$

SE x^* MINIMO LOCALE d NON PUÒ ESSERE DI DISCESA

MENTRE SE x^* È MINIMO LOCALE, ALLORA $\boxed{\nabla f(x^*)^T d \geq 0 \quad \forall d \text{ AMMISSIBILE}}$

QUESTA È LA NUOVA CONDIZIONE DI STAZIONARIETÀ

INOLTRE, SE $S = \mathbb{R}^n \Rightarrow x^*$ STAZIONARIO SE $\nabla f(x^*)^T d \geq 0 \quad \forall d \in \mathbb{R}^n$ OVVERO

$$\boxed{\nabla f(x^*) = 0}$$

ABBIAMO OTTENUTO LA CONDIZIONE DI STAZIONARIETÀ CHE AVEVAMO NEL CASO NON VINCOLATO

CNO II° ORDINE

SE $f \in C^2(\mathbb{R}^n)$ ALLORA SE x^* È PUNTO DI MINIMO LOCALE, NON PUÒ ESISTERE $d \in \mathbb{R}^n$ AMMISSIBILE IN x^* TALE CHE $\boxed{\nabla f(x^*)^T d = 0}$!

E $\boxed{d^T \nabla^2 f(x^*) d < 0}$

CASO DI INSIEME AMMISSIBILE CONVESSO

SIA S UN INSIEME CONVESSO E SIA $\bar{x} \in S$. ALLORA LA DIREZIONE $(x - \bar{x})$ È AMMISSIBILE $\forall x \in S$

Dim:

S convesso $\Rightarrow \forall \lambda \in [0,1] \Rightarrow (1-\lambda)\bar{x} + \lambda x \in S \quad \forall x \in S$

quindi $\bar{x} - \lambda \bar{x} + \lambda x \in S \Rightarrow \bar{x} + \lambda(x - \bar{x}) \in S \quad \forall \lambda \in [0,1]$ e $\forall x \in S$

allora $\forall x \in S$ la direzione $(x - \bar{x})$ soddisfa la direzione ammissibile in x^* con $\bar{t}=1$ □

Se S è convesso e se x^* è un punto di minimo locale, allora

$$\boxed{\nabla f(x^*)^T (x - x^*) \geq 0 \quad \forall x \in S}$$

Dim:

Supponiamo per assurdo che esista $\hat{x}$ ammissibile tale che:

$\boxed{\nabla f(x^*)^T(\hat{x} - x^*) < 0} \Rightarrow (\hat{x} - x^*)$ è una direzione ammissibile poiché S è convesso. Inoltre è anche di discesa ma questo viola le CNO! □

Proposizione: convessità $\Rightarrow$ globalità

Sia S convesso e $f \in C^1(\mathbb{R}^n)$ convessa. Allora x^* è un punto di

MINIMO GLOBALE $\Leftrightarrow$ VALE CHE:

$$\boxed{\nabla f(x^*)^T(x-x^*) \geq 0 \quad \forall x \in S}$$

Dim:

($\Leftarrow$) GIÀ VISTA SOPRA

($\Rightarrow$) SUPPONIAMO $\nabla f(x^*)^T(x-x^*) \geq 0 \ \forall x \in S$. POICHÉ f È CONVESSA VALE:

$$f(x) \geq f(x^*) + \underbrace{\nabla f(x^*)^T(x-x^*)}_{\geq 0 \ \forall x \in S} \quad \forall x$$

ALLORA $\forall x \in S$ VALE CHE $f(x) \geq f(x^*)$ OVVERO x^* È UN PUNTO DI MINIMO GLOBALE $\square$

VINCOLI POLIEDRALI

UN INSIEME $P \subseteq \mathbb{R}^n$ È CHIAMATO POLIEDRO SE È DEFINITO COME:

$$\boxed{P = \{ x \in \mathbb{R}^n \mid Ax \leq b,\ A \in \mathbb{R}^{m \times n},\ b \in \mathbb{R}^m \}}$$

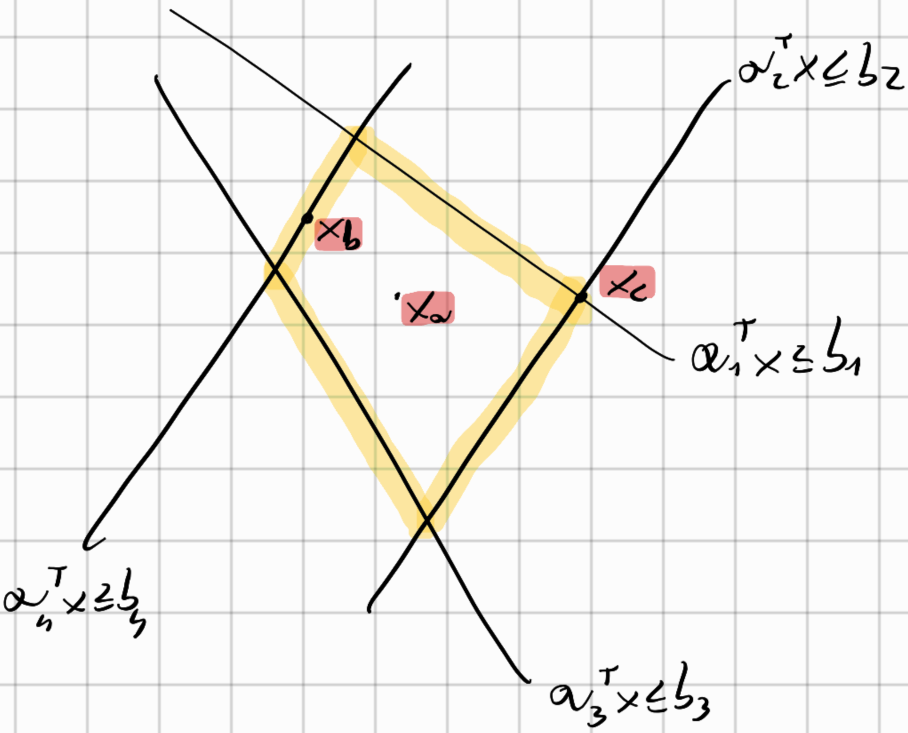
QUINDI I PUNTI DELL'INSIEME SODDISFANO UN SISTEMA DI DISEQUAZIONI LINEARI

$$\boxed{\begin{cases} a_1^T x \le b_1 \\ \vdots \\ a_m^T x \le b_m \end{cases}}$$

=> LE DIREZIONI AMMISSIBILI SONO QUINDI DEFINITE DOVE I VINCOLI VALGONO CON L'UGUAGLIANZA

L'INSIEME $\boxed{I(x)=\{i\in\{1,\dots,m\} \mid a_i^T x = b_i\}}$ È DETTO INSIEME DEI VINCOLI ATTIVI. SE UN CERTO PUNTO $\bar{x}$ È UN PUNTO INTERNO ALLORA $\boxed{I(\bar{x})=\emptyset}$

ESEMPIO:



$I(x_a)=\emptyset$; $I(x_b)=\{4\}$; $I(x_c)=\{1,2\}$

PROPOSIZIONE

SIA $\bar{x}\in P$, UNA DIREZIONE $d\in\mathbb{R}^n$ È AMMISSIBILE IN $\bar{x}$ SE E SOLO SE $\boxed{a_i^T d \le 0 \quad \forall i\in I(\bar{x})}$

Dim:

($\Rightarrow$) SE d È AMMISSIBILE E, PER ASSURDO, SUPPONIAMO CHE $\exists i \in I(\bar{x})$ t.c.: $a_i^T d > 0$

POICHÉ d È AMMISSIBILE, PER $t > 0$ ABBASTANZA PICCOLO VALE $\bar{x} + td \in P$ OVVERO $\boxed{(\bar{x} + td)^T a_i \le b_i}$ PER DEFINIZIONE DI POLIEDRO.

MA $\underbrace{\bar{x}^t a_i}_{= b_i \text{ POICHÉ } i \in I(\bar{x})} + td^T a_i \le \cancel{b_i}$ $\Rightarrow td^T a_i \le 0$ ASSURDO POICHÉ d È AMMISSIBILE. $\square$

($\Leftarrow$)

SUPPONIAMO CHE $a_i^T d \le 0 \ \forall i \in I(\bar{x})$, VOGLIAMO DIMOSTRARE CHE d È AMMISSIBILE, OVVERO $\exists \bar{t} > 0 : A(\bar{x} + td) \le b \ \forall t \in [0, \bar{t}]$

CONSIDERIAMO $\boxed{j \in \{1, \ldots, m\}}$. SI DISTINGUONO 3 CASI:

1. $\boxed{j \in I(\bar{x})}$: $(\bar{x} + td)^t a_j = \underbrace{\bar{x}^t a_j}_{= b_j} + td^t a_j = b_j + td^t a_j \le b_j \Rightarrow \bar{x} + td$ SODDISFA IL VINCOLO $\forall t > 0$ ✓

2. $\boxed{j \notin I(\bar{x}) \text{ E } a_j^T d \le 0}$: $(\bar{x} + td)^t a_j = \bar{x}^t a_j + td^t a_j \le b_j + td^t a_j \le b_j \Rightarrow$ ✓

3. $\boxed{j \notin I(\bar{x}) \text{ E } a_j^T d > 0}$: $(\bar{x} + td)^t a_j = \bar{x}^t a_j + td^t a_j$ QUESTO RISULTA $\le b_j$ SE:

$$t \leq \frac{b_j - \bar{x}^T a_j}{d^t a_j}$$

ALLORA $(\bar{x}+td)^t a_j \leq b_j$ SE $t \in \left[0, \frac{b_j - \bar{x}^T a_j}{d^T a_j}\right]$ E BASTERÀ QUINDI SCEGLIERE

$$\boxed{t = \min_{\substack{j \notin I(\bar{x}) \\ a_j^T d > 0}} \left\{ \frac{b_j - \bar{x}^T a_j}{d^t a_j} \right\}} \quad \square$$

PROPOSIZIONE

SIA $\boxed{\bar{x} \in S = \{x \mid Ax \leq b, \mu_i^T x = c_i \ \forall i = 1, \dots, p\}}$, ALLORA d È AMMISSIBILE IN $\bar{x} \in S$ SE E SOLO SE $\boxed{a_i^T d \leq 0 \ \forall i \in I(\bar{x})}$ E $\boxed{\mu_i^T d = 0 \ \forall i = 1, \dots, p}$

↳ PER I VINCOLI ATTIVI

Dim:

$$\mu_i^T x = b_i \iff \begin{cases} \mu_i^T x \leq b_i \\ -\mu_i^T x \leq 0 \end{cases}$$

SE $\bar{x} \in S$, $\mu_i^T \bar{x} = b$ E QUINDI I DUE VINCOLI SONO ATTIVI, ALLORA: $\begin{cases} \mu_i^T d \geq 0 \\ \mu_i^T d \leq 0 \end{cases} \Rightarrow \mu_i^T d = 0$ $\square$

VINCOLI DI BOX

SONO VINCOLI DEL TIPO $\boxed{S = \{x \mid l_i \leq x_i \leq \mu_i \text{ con } i = 1, \dots, n\}}$

ANALIZZANDO I CASI POSSIBILI, SI HANNO $2n$ VINCOLI:

- n DEL TIPO $x^T e_i \geq l_i$
- n DEL TIPO $x^T e_i \leq \mu_i$

CON e_h CHE È L'i-ESIMO VETTORE DELLA BASE CANONICA DI $\mathbb{R}^n$.

UNA SOLUZIONE OTTIMALE $\bar{x}$ DEVE AVERE $\boxed{\nabla f(\bar{x})^T d \geq 0 \quad \forall d \text{ AMMISSIBILE}}$, QUINDI SI HANNO 3 CASI:

- $\boxed{\bar{x}_i = l_i}$ ALLORA $0 \leq \nabla f(\bar{x}) e_i = \dfrac{\partial f(\bar{x})}{\partial x_i}$ OVVERO $\boxed{\dfrac{\partial f(\bar{x})}{\partial x_i} \geq 0}$

- $\boxed{\bar{x}_i = \mu_i, \ -\nabla f(\bar{x})^T e_i \geq 0}$ ALLORA $0 \geq \nabla f(\bar{x}) e_i = \dfrac{\partial f(\bar{x})}{\partial x_i}$ OVVERO $\boxed{\dfrac{\partial f(\bar{x})}{\partial x_i} \leq 0}$

- $\boxed{l_i < x_i < \mu_i}$ ALLORA VALGONO CONTEMPORANEAMENTE LE ALTRE DUE E QUINDI

$$\boxed{\nabla f(\bar{x}) e_i = \frac{\partial f(\bar{x})}{\partial x_i} = 0}$$

RICAPITOLANDO: SE $\bar{x}$ È OTTIMALE ALLORA:

$$\boxed{\frac{\partial f(\bar{x})}{\partial x_i} = \begin{cases} = 0 & l_i < \bar{x}_i < \mu_i \\ \geq 0 & \bar{x}_i = l_i \\ \leq 0 & \bar{x}_i = \mu_i \end{cases}}$$

VINCOLI DI SIMPLESSO

SONO VINCOLI DEL TIPO $\boxed{S = \{ x \in \mathbb{R}^n \mid x \geq 0,\ e^T x = 1 \}}$

QUINDI SE $\bar{x}$ È OTTIMALE ALLORA VALE CHE: $\boxed{\dfrac{\partial f(\bar{x})}{\partial x_i} \geq \dfrac{\partial f(\bar{x})}{\partial x_j}}$

$\forall i, \forall j$ TALE CHE $\bar{x}_j > 0$

PROIEZIONE SU INSIEMI CONVESSI

DATO IL PROBLEMA $\boxed{\min_{y \in S} \overbrace{\frac{1}{2} \| y - \bar{x} \|^2}^{r(\bar{x})}}$ CON $\bar{x} \in \mathbb{R}^n$ CON S CHIUSO E CONVESSO,

L'OBIETTIVO È TROVARE UN Y AMMISSIBILE PIÙ VICINO POSSIBILE AD $\bar{x}$. INOLTRE:

- $r(\bar{x})$ È COERCIVA ALLORA IL PROBLEMA AMMETTE SICURAMENTE UNA SOLUZIONE!
- $r(x)$ È QUADRATICA STRETTAMENTE CONVESSA E QUINDI LA SOLUZIONE È UNICA!

DETTA QUINDI $P(\bar{x})$ LA SOLUZIONE DEL PROBLEMA, SI DEFINISCE CHE $P(\bar{x})$ È LA PROIEZIONE EUCLIDEA DI $\bar{x}$ SU S

$$\boxed{P(\bar{x}) = \arg\min_{y \in S} \frac{1}{2} \| y - \bar{x} \|^2}$$

PROPOSIZIONE:

SIA $\bar{x} \in \mathbb{R}^n$, ALLORA $P(\bar{x})$ È LA PROIEZIONE DI $\bar{x}$ SU S SE E SOLO SE:

$$\boxed{(\bar{x}-P(\bar{x}))^T(x-P(\bar{x}))\leq 0 \quad \forall x\in S}$$

Dim:

$P(\bar{x})$ È, PER DEFINIZIONE, SOLUZIONE DI $\min_{y\in S} \frac{1}{2}\underbrace{\|y-\bar{x}\|^2}_{r(y)}$. PER LE C.N.S DI OTTIMALITÀ SI HA CHE QUESTO È EQUIVALENTE A DIRE CHE, SE $y=P(x)$:

$$\boxed{\nabla r(y)^T(x-y)\geq 0 \quad \forall x\in S}$$

$\nabla r(y)=y-\bar{x}$ E QUINDI $(y-\bar{x})^T(x-y)\geq 0 \ \forall x\in S$, OVVERO $(P(\bar{x})-\bar{x})^T(x-y)\geq 0$ $\forall x\in S$

$\Rightarrow (\bar{x}-P(\bar{x}))^T(x-P(\bar{x}))\leq 0 \quad \forall x\in S \quad \square$

PROPOSIZIONE:

LA FUNZIONE PROIEZIONE $P(x):\mathbb{R}^n\longrightarrow S$ È UNA FUNZIONE NON ESPANSIVA, OVVERO

$$\boxed{\|P(x)-P(y)\|\leq\|x-y\| \quad \forall x,y\in\mathbb{R}^n}$$

E QUINDI È CONTINUA

SIGNIFICA CHE LA PROIEZIONE RENDE I NUOVI PUNTI PROIETTATI "PIÙ VICINI" RISPETTO ALLA LORO DISTANZA INIZIALE

PROPOSIZIONE:

$\bar{x} \in S$ È STAZIONARIO SE E SOLO SE $\boxed{\bar{x} = P(\bar{x} - \nabla f(\bar{x}))}$

Dim:

$$(\cancel{\bar{x}} - \nabla f(\bar{x}) - \cancel{\bar{x}})^T (x - \bar{x}) \leq 0 \quad \forall x \in S \Rightarrow -\nabla f(\bar{x})^T (x - \bar{x}) \leq 0 \quad \forall x \in S$$

$\Rightarrow \nabla f(\bar{x})^T (x - \bar{x}) \geq 0$ CHE È PROPRIO LA DEFINIZIONE DI STAZIONARIETÀ! □

PROPOSIZIONE:

SE $\bar{x}$ È UN PUNTO DI MINIMO LOCALE, ALLORA $\boxed{\bar{x} = P(\bar{x} - \nabla f(\bar{x}))}$ (C.N.)

SE f È CONVESSA ALLORA $\bar{x}$ È PUNTO DI MINIMO GLOBALE SE E SOLO SE $\boxed{\bar{x} = P(\bar{x} - \nabla f(\bar{x}))}$

CASI PARTICOLARI

1) VINCOLI DI BOX: $S = \{ x \mid x_i \in [l_i, \mu_i] \ \forall i = 1, \dots, N \}$

ALLORA SE $y = P(\bar{x})$ VALE CHE:

$$\boxed{y_i = \begin{cases} x_i & \text{SE } x_i \in (l_i, \mu_i) \\ l_i & \text{SE } x_i \leq l_i \\ \mu_i & \text{SE } x_i \geq \mu_i \end{cases}}$$

2) VINCOLI DI SIMPLESSO: $S = \{x \in \mathbb{R}^n \mid x \geq 0;\ e^T x = 1\}$

IN QUESTO CASO ESISTE UN ALGORITMO CHE RICAVA LA PROIEZIONE IN AL MASSIMO n PASSI.

ALGORITMI ITERATIVI DI OTTIMIZZAZIONE VINCOLATA

SIA UN PROBLEMA VINCOLATO $\boxed{\min_{x \in S} f(x) \text{ CON } S \text{ CONVESSO E COMPATTO.}}$

LA FORMA DEL PROBLEMA È LA STESSA DEL CASO NON VINCOLATO: $\boxed{x^{k+1} = x^k + \alpha_k d_k}$

L'OBIETTIVO È QUINDI QUELLO DI RICAVARE UNA DIREZIONE d_k DI DISCESA AMMISSIBILE ED UN CERTO PASSO RICAVABILE TRAMITE IL METODO DI ARMIJO

SIA $\boxed{d_k = (\hat{x}^k - x^k)}$, È UNA DIREZIONE AMMISSIBILE SE S È CONVESSO E α_k È PASSO TALE CHE:

$$\boxed{x^k + \alpha_k d_k \in S}$$

ALLORA $\boxed{x^{k+1} \in S \text{ SE } \alpha \leq 1}$

QUINDI È UNA BUONA SCELTA PARTIRE CON UNA LINE-SEARCH DI ARMIJO CON VALORE INIZIALE $\alpha_0 \leq 1$.

PROPRIETÀ DELLA LINE-SEARCH DI ARMIJO VINCOLATA

SIA $\{x^k\}$ LA SEQUENZA PRODOTTA DA UN ALGORITMO LINE-SEARCH CON IL VINCOLO CHE d_k SIA DI DISCESA AMMISSIBILE E α_k SIA CALCOLATO CON ARMIJO, ALLORA:

- $\boxed{x^{k+1} \in S}$
- $\boxed{f(x^{k+1}) \leq f(x^k)}$
- $\boxed{\lim_{k\to\infty} \nabla f(x^k)^T d_k = 0}$

TUTTAVIA NON POSSIAMO SCEGLIERE d_k COME L'ANTIGRADIENTE POICHÉ NON PUÒ ESSERE AMMISSIBILE; SCEGLIAMO QUINDI $d_k = \hat{x}^k - x^k$.

VEDIAMO COME SCEGLIERE $\hat{x}^k$:

1) **METODO DEL GRADIENTE PROIETTATO**

SI CONSIDERA $\boxed{\hat{x}^k = P\left(x^k - \nabla f(x^k)\right)}$. SE $\boxed{d_k = \left(\hat{x}^k - x^k\right)}$ ALLORA SI HA CHE QUESTA DIREZIONE È AMMISSIBILE E DI DISCESA:

$$\left(x^k - \nabla f(x^k) - \hat{x}^k\right)^T \left(x^k - \hat{x}^k\right) \leq 0$$

$$\Rightarrow \left(x^k - \hat{x}^k\right)^T\left(x^k - \hat{x}^k\right) - \nabla f(x^k)^T\left(x^k - \hat{x}^k\right) \leq 0$$

$$\Rightarrow \nabla f(x^k)^T(\hat{x}^k - x^k) \leq -\|x^k - \hat{x}^k\|^2 \leq 0$$

$\to < 0$ se $x^k \neq \hat{x}^k$ (1)

$\to = 0$ se $x^k = \hat{x}^k$ (2)

(1) $\nabla f(x^k)^T d_k < 0 \Rightarrow d_k$ è di discesa

(2) $x^k = P(x^k - \nabla f(x^k)) \Longleftrightarrow x^k$ è stazionario

---

L'algoritmo del gradiente proiettato:

→ nel caso non vincolato, se applico l'algoritmo ritrovo l'antigradiente come direzione di ricerca.

⇓

è una generalizzazione!

Dato $x^0 \in S$ e $k=0$

while $(\|x^k - P(x^k - \nabla f(x^k))\| > \varepsilon)$:

- $\boxed{\hat{x}^k = P(x^k - \nabla f(x^k))}$
- $d_k = \hat{x}^k - x^k$
- calcolo α_k lungo d_k con Armijo
- $x^{k+1} = x^k + \alpha_k d_k$
- $k = k+1$

End

Proposizione:

SIA $S \subseteq \mathbb{R}^n$ CONVESSO E COMPATTO E SIA $f : \mathbb{R}^n \longmapsto \mathbb{R}$ TALE CHE $f \in C^1(\mathbb{R}^n)$. SIA $\{x^k\}$ LA SEQUENZA PRODOTTA DAL METODO DEL GRADIENTE PROIETTATO, ALLORA x^k HA PUNTI DI ACCUMULAZIONE, CIASCUNO DEI QUALI È STAZIONARIO

Dim:

$\{x^k\} \in S$ COMPATTO PER LE ISTRUZIONI DELL'ALGORITMO, QUINDI $\{x^k\}$ HA PUNTI DI ACCUMULAZIONE.

SIA $K \subseteq \{0, 1, \dots\}$ TALE CHE $\boxed{\lim_{\substack{k \to \infty \\ k \in K}} x^k = \bar{x}}$

POICHÉ P È FUNZIONE NON ESPANSIVA ⇓ CONTINUA

POSSIAMO ALLORA CONSIDERARE $\boxed{\hat{x}^k = P(x^k - \nabla f(x^k))}$. PER LA CONTINUITÀ DI P E DI ∇f SI HA CHE

$$\boxed{\lim_{\substack{k \to \infty \\ k \in K}} \hat{x}^k = \lim_{\substack{k \to \infty \\ k \in K}} P(x^k - \nabla f(x^k)) = P(\bar{x} - \nabla f(\bar{x})) = \hat{x}}$$

E PER LE PROPRIETÀ DELLA RICERCA DI ARMIJO VALE CHE:

$$\boxed{0 = \lim_{\substack{k \to \infty \\ k \in K}} \nabla f(x^k)^T d_k = \lim_{\substack{k \to \infty \\ k \in K}} \nabla f(x^k)^T (\hat{x}^k - x^k) = \nabla f(\bar{x})^T (\hat{x} - \bar{x})}$$

PER LE PROPRIETÀ DI P SI HA CHE:

$$(\bar{x} - \nabla f(\bar{x}) - \hat{x})^T (\bar{x} - \hat{x}) \leq 0$$

$$\Rightarrow -\nabla f(\bar{x})^T(\bar{x}-\hat{x}) \le -\|\bar{x}-\hat{x}\|^2, \quad 0 = \nabla f(\bar{x})^T(\bar{x}-\hat{x}) \le -\|\bar{x}-\hat{x}\|^2$$

$\hat{x}$ è punto stazionario!

Quindi $\|\bar{x}-\hat{x}\|^2 = 0$ se e solo se $\bar{x} = \hat{x}$ ma $\hat{x} = P(\bar{x} - \nabla f(\bar{x}))$

ovvero $\bar{x}$ è stazionario! □

---

2) $\boxed{\text{Metodo di Frank-Wolfe}}$

z^k = minimo

$\hat{x}^k = x^k$ che realizza il minimo

Sia S compatto e sia $\boxed{z^k := \min_{x \in S} \nabla f(x^k)^T(x - x^k)}$

Poniamo $\boxed{\hat{x}^k := \arg\min_{x \in S} \nabla f(x^k)^T(x - x^k)}$, allora $\boxed{z^k = 0}$

se e solo se $\boxed{\nabla f(x^k)^T(x - x^k) \ge 0 \ \forall x \in S}$, ossia x^k è stazionario. Se infatti $\boxed{z^k < 0 \Rightarrow \nabla f(x^k)^T(\hat{x}^k - x^k) < 0}$ allora $d_k = \hat{x}^k - x^k$ è di discesa e quindi l'ammissibilità di d_k segue dalla definizione di $\hat{x}^k \in S$

Quindi, se $\{x^k\}$ è la sequenza prodotta dal metodo FW, allora $\{x^k\}$ ha punti di accumulazione, ognuno dei quali è stazionario!

L'algoritmo di Frank-Wolfe:

DATI $x^0 \in S$, $k=0$

WHILE $\left(\| \hat{x}^k - x^k \| \leq \varepsilon \right)$

- $\boxed{\hat{x}^k \in \arg\min_{x \in S} \nabla f(x^k)^T (x - x^k)}$
- CALCOLO α_k LUNGO $d_k = \hat{x}^k - x^k$ SECONDO ARMIJO
- $x^{k+1} = x^k + \alpha_k d_k$
- $k = k+1$

END

PROPOSIZIONE:

SIA S CONVESSO E COMPATTO, $f \in C^1(\mathbb{R}^n)$ E SIA LA SEQUENZA GENERATA DAL METODO FW $\{x^k\}$, ALLORA $\{x^k\}$ AMMETTE PUNTI DI ACCUMULAZIONE, OGNUNO DEI QUALI È STAZIONARIO.

Dim:

PER COSTRUZIONE $\{x^k\} \subseteq S$ È COMPATTO, QUINDI $\{x^k\}$ HA PUNTI DI ACCUMULAZIONE.

SIA $K \subseteq \{0, 1, \dots\}$ TALE CHE $\boxed{\lim_{\substack{k \to \infty \\ k \in K}} x^k = \bar{x}}$

$d_k = \hat{x}^k - x^k \Rightarrow \boxed{\| d_k \| = \| \hat{x}^k - x^k \| \leq \| \hat{x}^k \| + \| x^k \|}$

ALLORA $\hat{x}^k, x^k \in S$ COMPATTO $\Rightarrow \boxed{\| x^k \| \leq M, \ \| \hat{x}^k \| \leq M \text{ PER QUALCHE } M > 0}$

$\Rightarrow \boxed{\|d_k\| \le 2M}$ $\forall k$, cioè $\{d_k\}$ è limitata. Quindi $\boxed{\exists K_1 \subseteq K \text{ tale che } \lim_{\substack{k \in K_1 \\ k \to \infty}} d_k = \bar{d}}$

Per le proprietà della ricerca di Armijo:

$$\boxed{\lim_{\substack{k \to \infty \\ k \in K_1}} \nabla f(x^k)^T d_k = \nabla f(\bar{x})^T \bar{d} = 0}$$

Sia allora $y \in S$, per definizione di $\hat{x}^k$ vale che $\forall k \in K_1$:

$$\boxed{\nabla f(x^k)^T (\hat{x}^k - x^k) \le \nabla f(x^k)^T (y - x^k)}$$

E passando al limite:

$$\lim_{\substack{k \in K_1 \\ k \to \infty}} \nabla f(x^k)^T d_k \le \lim_{\substack{k \to \infty \\ k \in K_1}} \nabla f(x^k)^T (y - x^k)$$

$$\longrightarrow \boxed{0 = \nabla f(\bar{x})^T \bar{d} \le \nabla f(\bar{x})^T (y - \bar{x})}$$

E per l'arbitrarietà di y otteniamo che:

$$\boxed{\nabla f(\bar{x})^T (x - \bar{x}) \ge 0 \quad \forall x \in S}$$

$\Rightarrow \bar{x}$ è stazionario! $\square$

Problemi con vincoli in forma analitica

SONO PROBLEMI DEL TIPO $\min\limits_{h, g} f(x)$. CON VINCOLI IN FORMA ANALITICA

DEL TIPO:

$$\left.\begin{array}{l} \cdot\ h_i(x) = 0 \quad \forall i = 1, \dots, p \\ \cdot\ g_i(x) \leq 0 \quad \forall i = 1, \dots, m \end{array}\right\} \in C^1(\mathbb{R}^n)$$

OVVERO SI HANNO m VINCOLI DI DISUGUAGLIANZA E P VINCOLI DI UGUAGLIANZA.

CONDIZIONI DI FRITZ-JOHN

SIA $x^* \in \mathbb{R}^n$ SE x^* È PUNTO DI MINIMO LOCALE $\min\limits_{\substack{x \in \mathbb{R}^n \\ h \\ g}} f(x)$, ALLORA ESISTONO

$\lambda_0 \in \mathbb{R}, \lambda \in \mathbb{R}^p, \mu \in \mathbb{R}^m$ (MOLTIPLICATORI) TALI CHE:

1) $h_i(x^*) = 0 \quad \forall i = 1, \dots, p$

2) $g_i(x^*) \leq 0 \quad \forall i = 1, \dots, m$

} CONDIZIONI DI AMMISSIBILITÀ

3) $\mu_i g_i(x^*) = 0 \quad \forall i = 1, \dots, m$ => CONDIZIONE DI COMPLEMENTARIETÀ
DOVE IL VINCOLO NON È ATTIVO IL RELATIVO MOLTIPLICATORE DEVE ESSERE NULLO

4) $\lambda_0, \mu_i \geq 0 \quad \forall i = 1, \dots, m$

5) $(\lambda_0, \lambda, \mu) \neq (0, 0, 0)$

} AMMISSIBILITÀ DEI MOLTIPLICATORI

6) $\lambda_0 \nabla f(x^*) + \sum\limits_{i=1}^{p} \lambda_i \nabla h_i(x^*) + \sum\limits_{i=1}^{m} \mu_i \nabla g_i(x^*) = 0$

↳ ANNULLAMENTO DEL GRADIENTE DELLA FUNZIONE LAGRANGIANA

OSSERVAZIONE

LA FUNZIONE LAGRANGIANA APPLICATA AL PROBLEMA È:

$$\boxed{\mathcal{L}(x,\lambda_0,\lambda,\mu)=\lambda_0 f(x)+\sum_{i=1}^{m}\mu_i g_i(x)+\sum_{i=1}^{p}\lambda_i h_i(x)}$$

QUESTA FUNZIONE QUINDI PENALIZZA I VINCOLI NON RISPETTATI. IL SUO GRADIENTE RISPETTO A X

$$\boxed{\nabla_x \mathcal{L}(x,\lambda_0,\lambda,\mu)=\lambda_0\nabla f(x)+\sum_{i=1}^{m}\mu_i\nabla g_i(x)+\sum_{i=1}^{n}\lambda_i\nabla h_i(x)}$$

E QUINDI QUELLO CHE CHIEDIAMO NELLA CONDIZIONE 6 È CHE:

$$\boxed{\nabla_x \mathcal{L}(x^*,\lambda_0^*,\lambda^*,\mu^*)=0}$$

ESEMPIO:

SIA $\min\limits_{y^2=0} f(x,y)$

SAPPIAMO CHE SE (x^*,y^*) È OTTIMALE, ALLORA ESISTONO $\lambda_0,\lambda_1\in\mathbb{R}$ tc:

$$\begin{cases} (y^*)^2 = 0 \\ \lambda_0 \geq 0 \\ (\lambda_0, \lambda_1) \neq (0,0) \\ \lambda_0 \nabla f(x^*, y^*) + \lambda_1 \begin{pmatrix} 0 \\ 2y^* \end{pmatrix} = 0 \end{cases}$$

DATO CHE $\nabla h(x) = (0, 2y)^T$

SE SCEGLIAMO $\lambda_0 = 0$ E $\lambda_1 = 1$ OTTENIAMO CHE LE CONDIZIONI FJ SONO SODDISFATTE PER OGNI PUNTO $(x, 0)$ $\forall x$. SI NOTA QUINDI COME SI SIA PERSA LA DIPENZA DALLA FUNZIONE OBIETTIVO f E TUTTO VADA SOLO A DIPENDERE DAI VINCOLI, OVVERO DA S

QUINDI PER FARE IN MODO CHE RIMANGA LA DIPENDENZA DA f, OCCORRE CHE λ_0 NON SIA NULLO

CONDIZIONI KKT

SE x^* È UN PUNTO DI MINIMO LOCALE ED INOLTRE VIENE SODDISFATTA UNA CONDIZIONE DI REGOLARITÀ DEI VINCOLI IN x^*, ALLORA VALGONO LE FJ CON $\lambda_0^* \neq 0$, CIOÈ $\exists \lambda \in \mathbb{R}^p$, $\mu \in \mathbb{R}^m$ TALI CHE:

- $g_i(x^*) \leq 0 \quad \forall i = 1, \ldots, m$
- $h_i(x^*) = 0 \quad \forall i = 1, \ldots, p$

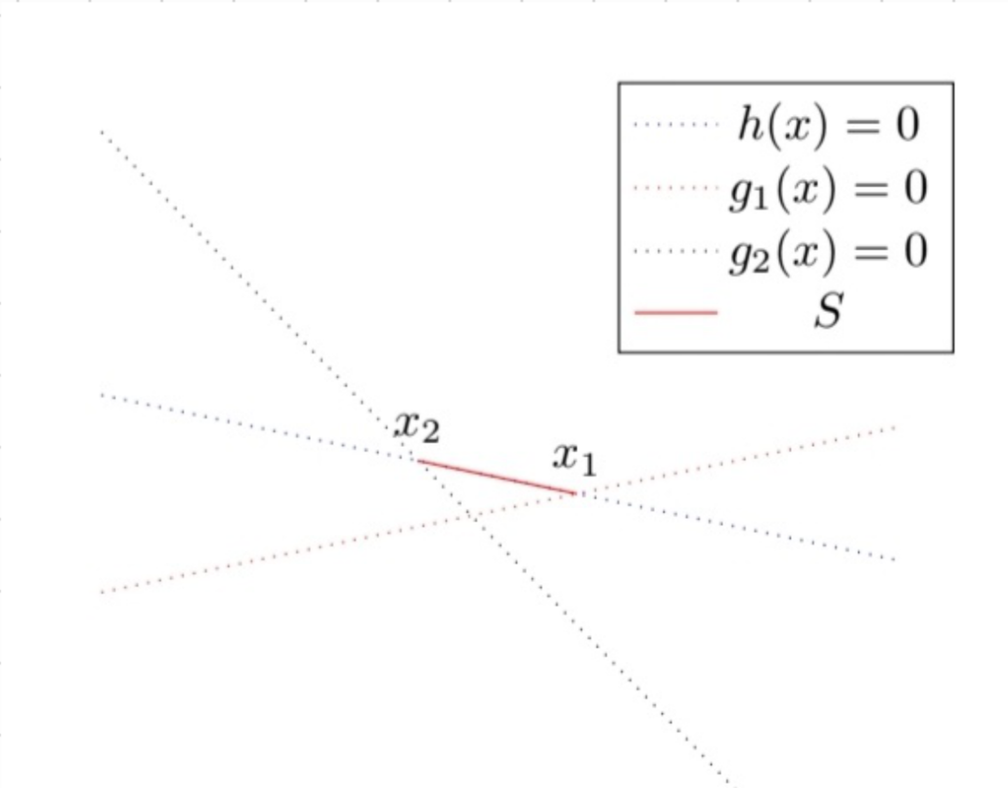
- $g_i(x^*)\mu_i = 0 \quad \forall i=1,\dots,m$
- $\mu_i \geq 0 \quad \forall i=1,\dots,m$
- $\nabla f(x^*) + \sum_{i=1}^{m} \mu_i \nabla g_i(x^*) + \sum_{i=1}^{p} \lambda_i \nabla h_i(x^*) = 0$

DOVE QUEST'ULTIMA PUÒ ANCHE ESSERE RISCRITTA COME:

$$\boxed{-\nabla f(x^*) = \sum_{i=1}^{m} \mu_i \nabla g_i(x^*) + \sum_{i=1}^{p} \lambda_i \nabla h_i(x^*)}$$

DUNQUE L'ANTIGRADIENTE PUÒ ESSERE SCRITTO COME COMBINAZIONE LINEARE DEI GRADIENTI DEI VINCOLI CON $\mu_i \geq 0$ ($\mu_i = 0$ SE g_i NON È ATTIVO)

INTERPRETAZIONE GEOMETRICA



IN QUESTO ESEMPIO SI HA UN INSIEME DEFINITO DA DUE VINCOLI DI DISUGUAGLIANZA ED UNO DI UGUAGLIANZA. SCRIVIAMO L'ANTIGRADIENTE:

$$-\nabla f(x^*) = \sum_{i=1}^{2} \mu_i \nabla g_i(x^*) + \lambda \nabla h(x^*)$$

QUINDI L'ANTIGRADIENTE È COMBINAZIONE LINEARE DEI GRADIENTI DEI VINCOLI NON ATTIVI (SE SONO DISUGUAGLIANZE). QUINDI L'ANTIGRADIENTE È $\perp$ AL VINCOLO DI UGUAGLIANZA.

NELL'ESEMPIO, SE NON CI TROVIAMO IN x_1 O x_2 ALLORA L'ANTIGRADIENTE PUÒ SOLO ESSERE ORTOGONALE AD $h(x)$. IN x_2 IL $\nabla f(x_1)$ È COMBINAZIONE LINEARE DI $\nabla h(x_1)$ E $\nabla g_1(x_1)$ CON $\mu_1 \geq 0$, QUINDI PUÒ STARE SOLO NEL SEMIPIANO INFERIORE DEFINITO DAL VINCOLO g_1. ANALOGO PER x_1

CONDIZIONI DI REGOLARITÀ DEI VINCOLI

- **LINEAR CONSTRAINT QUALIFICATION (LCQ)**

  IMPONE CHE TUTTI I VINCOLI SIANO LINEARI

- **LINEAR INDEPENDENCE CONSTRAINT QUALIFICATION (LICQ)**

  I VETTORI $\nabla h_i(x^*)\ \forall i=1,\dots,p$ E $\nabla g_i(x^*)\ \forall i \in I(x^*)$ SONO LINEARMENTE INDIPENDENTI

- Mangasarian - Fromovitz CQ (MFCQ)

I vettori $\nabla h_i(x^*)$ $\forall i=1,\dots,p$ sono linearmente indipendenti ed esiste una direzione $d \in \mathbb{R}^n$ tale che $\nabla h_i(x^*)^T d = 0$ $\forall i=1,\dots,p$ e $\nabla g_i(x^*)^T d < 0$ $\forall i \in I(x^*)$

- Slater CQ

Se f, h, g sono funzioni convesse ed esiste un punto $x^* \in \mathbb{R}^n$ t.c. $h_i(x^*) = 0$ $\forall i=1,\dots,p$ e $g_i(x^*) < 0$ $\forall i=1,\dots,m$ allora i vincoli sono regolari

Dim (LICQ + FJ => KKT)

punto di minimo

Sia x^* t.c. le FJ sono soddisfatte, allora $\exists\ \lambda_0, \lambda, \mu$ tali che:

$$h(x^*) = 0$$

$$g(x^*) \le 0$$

$$\lambda_0 \mu \ge 0$$

$$\mu_i g_i(x^*) = 0$$

$$(\lambda_0, \lambda, \mu) \neq (0,0,0)$$

$$\lambda_0 \nabla f(x^*) + \sum \lambda_i \nabla h_i(x^*) + \sum \mu_i \nabla g_i(x^*) = 0$$

SUPPONIAMO PER ASSURDO CHE $\boxed{\lambda_0=0}$, ALLORA $\boxed{\sum_{i=1}^{p} \lambda_i \nabla h_i(x^*) + \sum_{i=1}^{m} \mu_i \nabla g_i(x^*) = 0}$

PER I VINCOLI DI COMPLEMENTARIETÀ SI PUÒ DIRE CHE $\mu_i = 0 \ \forall i$ t.c. $g_i(x^*) < 0$, CIOÈ $\forall i \notin I(x^*)$

ALLORA $\boxed{\sum_{i=1}^{p} \lambda_i \nabla h_i(x^*) + \sum_{i \in I(x^*)} \mu_i \nabla g_i(x^*) = 0}$

MA PER LA LICQ SI HA CHE $\nabla h_i(x^*)$ CON $i = 1, \dots, p$ E $\nabla g_i(x^*)$ CON $i \in I(x^*)$ SONO LINEARMENTE INDIPENDENTI E QUINDI $\lambda_i = 0 \ \forall i$ E $\mu_i = 0 \ \forall i \in I(x^*)$

MA QUESTO SIGNIFICA $(\lambda_0, \lambda, \mu) = (0, 0, 0)$ MA È ASSURDO POICHÉ VIOLA LE FJ □

OSSERVIAMO INOLTRE CHE KKT SONO CN DI OTTIMALITÀ, MA SE f È CONVESSA ALLORA LE KKT DIVENTANO CNS DI OTTIMALITÀ GLOBALE

VINCOLI DI SIMPLESSO (PT. 2)

$$S = \{ x \mid e^T x = 1, \ x \geq 0 \}.$$ LI POSSIAMO RISCRIVERE IN FORMA STANDARD:

$$\begin{cases} e^T x - 1 = 0 \rightsquigarrow \lambda \\ -x_h \leq 0 \quad \forall h = 1, \dots, n \rightsquigarrow \mu_1, \dots, \mu_n \end{cases}$$

SCRIVO ORA LE KKT:

- $\mu_1,\dots,\mu_n \geq 0$
- $\mu_i(-x_i)=0 \quad \forall i=1,\dots,n$
- $\dfrac{\partial \mathcal{L}(x,\lambda,\mu)}{\partial x_i} = \dfrac{\partial f(x)}{\partial x_i} + \lambda - \mu_i = 0 \quad \forall i$

MA SE $x_j>0 \Rightarrow \mu_j=0$ PER LA COMPLEMENTARIETÀ

ALLORA $\dfrac{\partial f(x)}{\partial x_j}+\lambda=0 \Rightarrow \dfrac{\partial f(x)}{\partial x_j} = -\lambda \leq \lambda + \overbrace{\mu_i}^{\geq 0} \overset{\Delta}{=} \dfrac{\partial f(x)}{\partial x_i}$

$\Delta \quad -\lambda+\mu_i = \dfrac{\partial f(x)}{\partial x_i}$

QUINDI $\forall j \quad x_j>0$ E $\forall i=1,\dots,n$ $$\boxed{\dfrac{\partial f(x)}{\partial x_j} \leq \dfrac{\partial f(x)}{\partial x_i}}$$