Condizioni di minimo per problemi di ottimo vincolato

Obiettivo: affrontare problemi di ottimo vincolato posti nella forma:

$$\left\langle \begin{array}{ll} & \min f(x) \\ \underbrace{\text{s.v.}}_{\text{soggetto ai vincoli}} & g_i(x) \geq 0 \;, \; i = 1, \dots, m \end{array} \right\rangle \quad (1)$$

con $f, g_1, \dots, g_m : \mathbb{R}^n \to \mathbb{R}$, tutte di classe $C^{(1)}$

$\Omega = \{ x \mid x \in \mathbb{R}^n, \; g_i(x) \geq 0 \; \forall i = 1, \dots, m \}$

Nota. In alcuni casi possono essere presenti vincoli di uguaglianza $g_i(x) = 0$. Essi possono essere sostituiti dalla coppia di vincoli

$\begin{cases} g_i(x) \geq 0 \\ g_i(x) \leq 0 \end{cases}$ cioe' $\begin{cases} g_i(x) \geq 0 \\ -g_i(x) \geq 0 \end{cases}$, ritornando alla formulazione (1)

Esamineremo i problemi per ordine crescente di difficolta'

$$\begin{array}{ll} \min f(x) \\ \underbrace{a^{(i)T} x = b_i \;, \; i = 1, \dots, m}_{\text{vincoli lineari di uguaglianza}} \end{array} \qquad \qquad \begin{array}{ll} \min f(x) \\ \underbrace{a^{(i)T} x \geq b_i \quad i = 1, \dots, m}_{\text{vincoli lineari di disuguaglianza}} \end{array}$$

I problemi con vincoli di uguaglianze sono stati studiati da Lagrange nel '700. Solo verso la meta' del 1900 si e' trovato il modo di affrontare quelli con vincoli di uguaglianza.

Teoremi delle alternative

"Una famiglia" di teoremi i cui enunciati hanno una forma comune: vengono definiti due sistemi di equazioni e disequazioni lineari che sono "alternativi" nel senso che uno e solo uno di essi ammette soluzione (equivalentemente, i due sistemi non possono essere entrambi ammissibili e non possono essere entrambi inammissibili)

Teorema di Gordan.

Siano dati m vettori $a^{(1)}, \dots, a^{(m)}$ di $\mathbb{R}^n$. I due sistemi

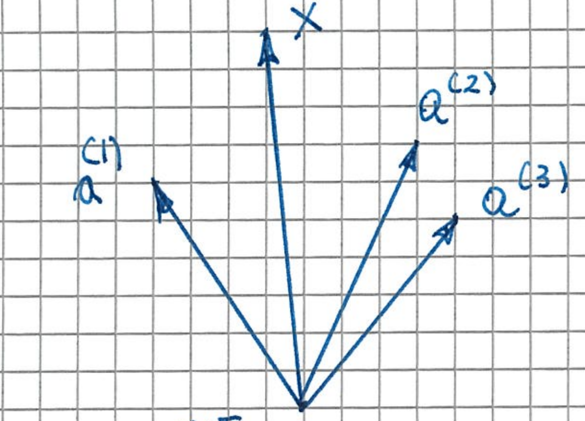
I) $a^{(i)T} x > 0, \quad i = 1, \dots, m$ II) $\sum_{i=1}^{m} a^{(i)} y_i = 0, \quad y_i \geq 0 \ \forall i$ con almeno un $y_i > 0$

sono alternativi.

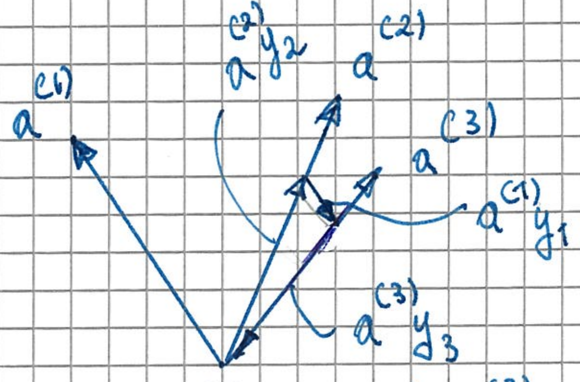
Significato geometrico.

Sistema I) : determinare un vettore $x \in \mathbb{R}^n$ che formi un angolo acuto con tutti gli $a^{(i)}$

Sistema II) : Ottenere il vettore nullo come combinazione lineare, non banale e con coefficienti non-negativi, dei vettori $a^{(i)}$.



$a^{(i)T} x > 0, \quad i = 1, \dots, 3$



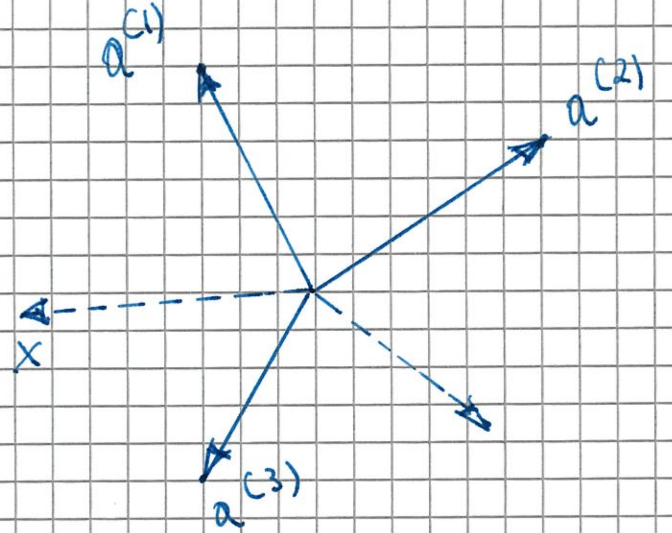
$0 = a^{(1)} y_1 + a^{(2)} y_2 + a^{(3)} y_3$

$y_1 < 0$ $\quad$ $y_3 < 0$

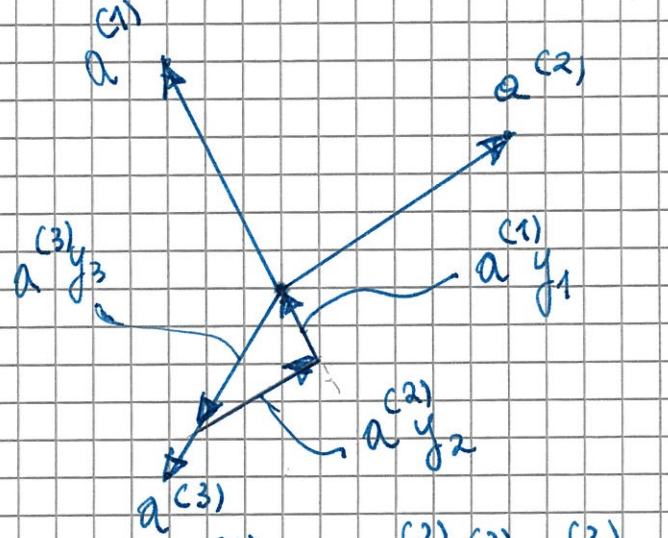
Caso a):

I ammissibile

II inammissibile



Qualsiasi x formerà un angolo ottuso con almeno un vettore



$0 = a^{(1)} y_1 + a^{(2)} y^{(2)} + a^{(3)} y_3$

$y_1, y_2, y_3 > 0$

Caso b):

I) inammissibile

II) ammissibile.

Prova (solo della impossibilità che i due sistemi siano ammissibili)

Definita la matrice $A = \begin{pmatrix} - a^{(1)T} - \\ - a^{(2)T} - \\ \vdots \\ a^{(m)T} \end{pmatrix}$, i sistemi I) e II)

possono essere riscritti come

I) $Ax > 0$ II) $A^T y = 0$, $y \geq 0$, $y \neq 0$ $\left(y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \right)$

Supponiamo, per assurdo, che I) e II) siano ammissibili, con soluzioni $\bar{x}$ e $\bar{y}$, rispettivamente. Dovrebbe essere:

$$0 < \underbrace{\bar{y}^T}_{\substack{\geq 0 \\ \neq 0}} \underbrace{A\bar{x}}_{> 0} = (\underbrace{A^T \bar{y}}_{0})^T \bar{x} = 0$$ Contraddizione: $0 > 0$!!!

Esercizio. Dimostrare che il sistema II) $A^T y = 0$, $y \geq 0$, $y \neq 0$ è equivalente al sistema $A^T \lambda = 0$, $e^T \lambda = 1$, $\lambda \geq 0$
$(\lambda \in \mathbb{R}^m,\ e^T = (1, 1, ---, 1))$

(cioè $0 \in \operatorname{Conv}\{a^{(1)}, --, a^{(m)}\}$)

Esercizio Dimostrare (utilizzando il teorema di Gordon e l'esercizio precedente) la condizione di separabilità lineare vista nella classificazione)

Teorema di Farkas. Dati $(m+1)$ vettori di $\mathbb{R}^n$ $a^{(1)}, a^{(2)}, \dots, a^{(m)}, g$ i seguenti due sistemi sono alternativi

$$\text{I)} \begin{cases} g^T d < 0 \\ a^{(i)T} d \geq 0 \quad \forall i = 1, \dots, m \end{cases} \qquad \text{II)} \begin{cases} g = \sum_{i=1}^{m} \lambda_i a^{(i)} \\ \lambda_i \geq 0, \quad i = 1, \dots, m \end{cases}$$

(Significato geometrico di I): trovare un vettore d che formi un angolo ottuso con g e non ottuso con tutti gli $a^{(i)}$)

(Significato geometrico di II): esprimere g come combinazione lineare a coefficienti non negativi degli $a^{(i)}$.)

Prova (solo del fatto che i due sistemi non possono essere entrambi ammissibili)

Dimostrazione per assurdo. Siano $\bar{d}$ e $\bar{\lambda}_i$, $i = 1, \dots, m$ soluzioni dei due sistemi. Vale

$$0 > g^T \bar{d} = \left(\sum_{i=1}^{m} \bar{\lambda}_i a^{(i)} \right)^T \bar{d} = \sum_{i=1}^{m} \underbrace{\bar{\lambda}_i}_{\geq 0} \underbrace{\left(a^{(i)T} \bar{d}\right)}_{\geq 0} \geq 0$$

$0 > 0$!! Contraddizione

Vincoli lineari di uguaglianza (Condizioni necessarie di minimo)

$$\begin{cases} \min f(x) \\ a^{(i)T}x - b_i = 0 \quad i=1,\dots,m \end{cases}$$

$\left(g_i(x) = a^{(i)T}x - b_i = 0 \right)$

Nota $\nabla g_i(x) = a^{(i)}$, $i = 1,\dots,m$

Supponiamo che x^* sia un punto di minimo. Naturalmente x^* è ammissibile, quindi è $a^{(i)T}x^* = b_i$, $i = 1,\dots,m$

Gli spostamenti ammissibili a partire da x^* sono tutti i vettori $d \in \mathbb{R}^n$ per cui risulti $a^{(i)T}(x^*+d) = b_i$, $i = 1,\dots,m$

cioè $\underbrace{a^{(i)T}x^* - b_i}_{=0} + a^{(i)T}d = 0 \Rightarrow a^{(i)T}d = 0$, $i=1,\dots,m$

$D = \{d \mid d \in \mathbb{R}^n, \ a^{(i)T}d = 0, \ i = 1,\dots,m\}$ Insieme degli spostamenti ammissibili

Ricordiamo che se d soddisfa $\nabla f(x^*)^T d < 0$, d è una direzione di discesa

Se x^* è un punto di minimo non può esistere una d che soddisfi $\nabla f(x^*)^T d < 0$ (sia cioè una direzione di discesa) e sia al tempo stesso ammissibile. Se esistesse, infatti, sarebbe possibile "muoversi" da x^* riducendo f e garantendo l'ammissibilità

Possiamo quindi dire:

x^* punto di minimo $\Longrightarrow$ Il sistema (I) $\left\langle \begin{array}{l} \nabla f(x^*)^T d < 0 \\ a^{(i)T} d = 0,\ i = 1, \dots, m \end{array} \right\rangle$ Non ammette soluzione

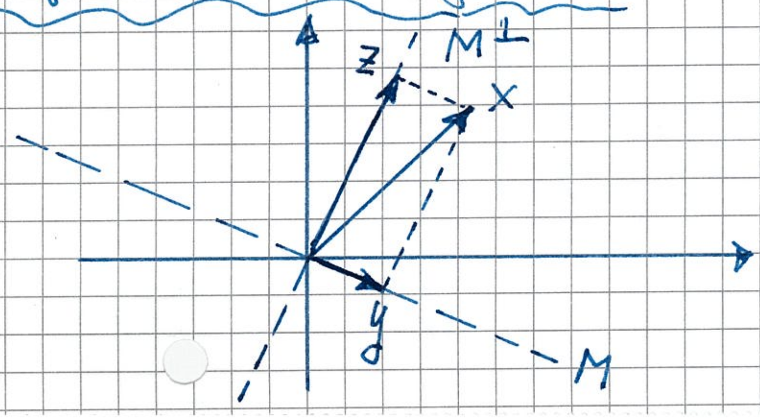
(Se esistesse una soluzione $\bar{d}$, da $\nabla f(x^*)^T \bar{d} < 0$ seguirebbe che $\exists \bar{\alpha} > 0 : \forall \alpha \in (0, \bar{\alpha}]$ $f(x^* + \alpha \bar{d}) < f(x^*)$ con $x^* + \alpha \bar{d}$ ammissibile $\left(a^{(i)T}(x^* + \alpha \bar{d}) = \underbrace{a^{(i)T} x^*}_{b_i} + \alpha \underbrace{a^{(i)T} \bar{d}}_{0} - b_i = 0 \ \forall i \right)$)

Teorema dei moltiplicatori di Lagrange:

Sistema (I) è inammissibile $\Longleftrightarrow$ $\exists \lambda_1, \dots, \lambda_m : \nabla f(x^*) = \sum_{i=1}^{m} \lambda_i a^{(i)}$ (II)

Moltiplicatori di Lagrange

$\Rightarrow$ Supponiamo (I) inammissibile. È noto che, dato un sottospazio $M \subset \mathbb{R}^n$, ogni vettore $x \in \mathbb{R}^n$ può essere espresso come $x = y + z$ con $y \in M$ e $z \in M^\perp$ ($M^\perp$ è il complemento ortogonale di M, cioè $\{z \mid z^T y = 0 \ \forall y \in M\}$)



Consideriamo il sottospazio M "generato" dai vettori $a^{(1)}, \dots, a^{(m)}$

$M = \{ y \mid y = \sum_{i=1}^{m} \lambda_i a^{(i)},\ \lambda_i \in \mathbb{R},\ i = 1, \dots m \}$, $\quad M^\perp = \{ z \mid z^T y = 0 \ \forall y \in M \}$

Possiamo quindi sempre esprimere $\nabla f(x^*) = y + z$, con

$y = \sum_{i=1}^{m} \lambda_i a^{(i)}$ e $z \in M^\perp$, e quindi, in particolare $a^{(i)T} z = 0 \ \forall i = 1, \dots, m$

Supponiamo, per assurdo, che il sistema (II) sia inammissibile

$\Rightarrow z \neq 0$. Poniamo quindi $d = -z$.

Vale, ovviamente, $a^{(i)T} d = 0 \ \forall i = 1, \dots, m$ con $d \neq 0$

Calcoliamo $\nabla f(x^*)^T d = \left(\sum_{i=1}^{m} \lambda_i a^{(i)} + z \right)^T (-z) =$

$$= \sum_{i=1}^{m} \lambda_i \underbrace{\left(-a^{(i)T} z\right)}_{0} - \|z\|^2 = -\|z\|^2 < 0$$

Abbiamo cioè dimostrato che $d = -z$ è ammissibile per (I), che è una contraddizione

$\Leftarrow$ Supponiamo che (II) sia ammissibile, cioè $\exists \lambda_i$ tali che $\nabla f(x^*) = \sum_{i=1}^{m} \lambda_i a^{(i)}$, e consideriamo una qualsiasi d tale che $a^{(i)T} d = 0$, $i = 1, \dots, m$.

Calcoliamo $\nabla f(x^*)^T d = \left(\sum \lambda_i a^{(i)} \right)^T d = \sum_{i=1}^{m} \lambda_i \underbrace{a^{(i)T} d}_{0} = 0$.

Cioè per nessuna $d : a^{(i)T} d = 0 \ \forall i$ può essere $\nabla f(x^*)^T d < 0 \Rightarrow$ Il sistema (I) è inammissibile

In definitiva possiamo dire:

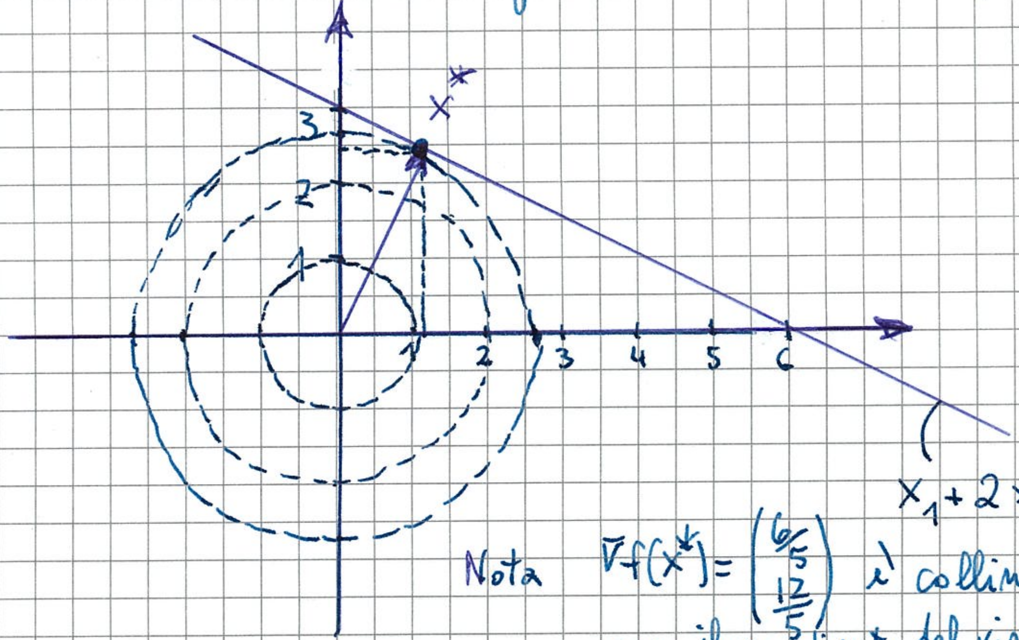
$$x^* \text{ punto di minimo} \Longrightarrow \text{Sistema Inammissibile} \begin{cases} \nabla f(x^*)^T d < 0 \\ a^{(i)T} d = 0, \ i=1,..,m \end{cases} \Longleftrightarrow \text{Sistema ammissibile} \left\langle \nabla f(x^*) = \sum_{i=1}^{m} \lambda_i a^{(i)} \right\rangle$$

Quindi

$$x^* \text{ punto di minimo} \Longrightarrow \nabla f(x^*) = \sum_{i=1}^{m} \lambda_i a^{(i)}$$

Ricordando che $a^{(i)} = \nabla g_i(x^*)$ $\quad (g_i(x) = a^{(i)T} x - b_i)$, $i=1,...,m$,

Possiamo enunciare il Teorema dei moltiplicatori di Lagrange nella forma:
"All'ottimo di $f(x)$ il gradiente di f è esprimibile come combinazione lineare dei gradienti dei vincoli, mediante coefficienti detti moltiplicatori di Lagrange"



$$\min \frac{1}{2}(x_1^2 + x_2^2)$$

$$\text{s.v.} \quad x_1 + 2x_2 = 6 \qquad a = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

$$\nabla f(x) = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \qquad b = 6$$

All'ottimo x^* sono verificate le seguenti equazioni

$$\begin{cases} \begin{pmatrix} x_1^* \\ x_2^* \end{pmatrix} = \lambda \begin{pmatrix} 1 \\ 2 \end{pmatrix} \Rightarrow \begin{matrix} x_1^* = \lambda \\ x_2^* = 2\lambda \end{matrix} \\ x_1^* + 2x_2^* = 6 \end{cases}$$

$$\Downarrow$$

$$\lambda = \frac{6}{5} \Rightarrow x^* = \begin{pmatrix} \frac{6}{5} \\ \frac{12}{5} \end{pmatrix}$$

Nota $\nabla f(x^*) = \begin{pmatrix} 6/5 \\ \frac{12}{5} \end{pmatrix}$ è collineare con il gradiente del vincolo $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$

Esercizio. Quali sarebbero state le soluzioni ottime dei problemi ottenuti trasformando i vincoli di uguaglianza in vincoli di disuguaglianza nel problema appena svolto?

$$\begin{cases} (P_1): \min \frac{1}{2}(x_1^2+x_2^2) \\ \text{s.v. } x_1+2x_2 \leq 6 \end{cases} \qquad \begin{cases} (P_2): \min \frac{1}{2}(x_1^2+x_2^2) \\ \text{s.v. } x_1+2x_2 \geq 6 \end{cases}$$

Vincoli lineari di disuguaglianza (Condizioni necessarie di minimo)

$$\begin{cases} \min f(x) \\ a^{(i)T}x - b_i \geq 0, \ i=1,\dots,m \end{cases} \qquad \left.\begin{aligned} g_i(x) &= a^{(i)T}x - b_i \geq 0 \\ \nabla g_i(x) &= a^{(i)} \end{aligned}\right\} i=1,\dots,m$$

Sia x^* un punto (ammissibile!) di minimo

Definizione Insieme dei vincoli attivi: vincoli soddisfatti per uguaglianza

$$I(x^*) = \{ i \mid a^{(i)T}x^* - b_i = 0, \ i=1,\dots,m \}$$

Solo i vincoli attivi avranno un ruolo nella definizione delle condizioni di ottimalità (ciò è evidente se si riflette sul fatto che esiste un intorno di x^* in cui tutti i vincoli non attivi sono soddisfatti. È possibile quindi spostarsi "almeno per un po'" nell'intorno di x^* senza che nessuno di essi venga violato).

Definizione. Una direzione $d \in \mathbb{R}^n$, $d \neq 0$, è ammissibile in x^* se esiste $\bar{t} > 0$ tale che $x^* + td$ è ammissibile $\forall t \in [0, \bar{t}]$

Teorema $d \in \mathbb{R}^n$, $d \neq 0$ ammissibile $\iff a^{(i)^T} d \geq 0 \quad \forall i \in I(x^*)$

$\Rightarrow$ Se d è ammissibile, per qualche $t > 0$ vale

$a^{(i)^T}(x^* + td) - b_i \geq 0 \quad \forall i = 1, \dots, m$ e quindi, anche $\forall i \in I(x^*)$

Segue $\cancel{\underbrace{a^{(i)^T} x^*}_{= b_i}} + \underbrace{t}_{> 0} a^{(i)^T} d - \cancel{b_i} \geq 0 \Rightarrow a^{(i)^T} d \geq 0 \quad \forall i \in I(x^*)$

$\Leftarrow$ Supponiamo che $d \neq 0$ soddisfi $a^{(i)^T} d \geq 0 \quad \forall i \in I(x^*)$.

L'insieme $I = \{1, \dots, m\}$ è partizionato in $I(x^*)$ e $\bar{I}(x^*)$

con $\bar{I}(x^*) = \{ i \mid a^{(i)^T} x^* > b_i \}$ ($\bar{I}(x^*)$ insieme dei vincoli non attivi in x^*)

Esaminiamo separatamente $I(x^*)$ e $\bar{I}(x^*)$ sulla semiretta

$$X_d = \{ x \mid x = x^* + td, \ t \geq 0 \}$$

a) $i \in I(x^*)$. $a^{(i)^T}(x^* + td) = \underbrace{a^{(i)^T} x^*}_{= b_i} + \underbrace{t}_{\geq 0} \underbrace{a^{(i)^T} d}_{\geq 0} \geq b_i \Rightarrow$ vincoli soddisfatti su tutta la semiretta ($\forall t \geq 0$

b) $i \in \bar{I}(x^*)$. In questo caso è $a^{(i)^T} x^* > b_i \quad \forall i \in \bar{I}(x^*)$

Per $i \in \bar{I}(x^*)$ può accadere $\underbrace{a^{(i)T}d \geq 0}_{\text{caso } b_1}$ oppure $\underbrace{a^{(i)T}d < 0}_{\text{caso } b_2}$

Caso b_1 $\quad a^{(i)T}(x^* + td) = \underbrace{a^{(i)T}x^*}_{> b_i} + \underbrace{t}_{\geq 0}\, \underbrace{a^{(i)T}d}_{\geq 0} > b_i$ (vincoli soddisfatti $\forall t \geq 0$)

Caso b_2 $\quad a^{(i)T}(x^* + td) = \underbrace{a^{(i)T}x^*}_{> b_i} + \underbrace{t}_{\geq 0}\, \underbrace{a^{(i)T}d}_{< 0}$. Imponiamo

che questa quantità sia $\geq b_i$

$a^{(i)T}x^* + t\, a^{(i)T}d \geq b_i \implies t\underbrace{(-a^{(i)T}d)}_{> 0} \leq \underbrace{a^{(i)T}x^* - b_i}_{> 0} \implies$

$$0 < t \leq \frac{a^{(i)T}x^* - b_i}{-a^{(i)T}d}$$

Se si pone $0 < \bar{t} = \min\limits_{\{i \in \bar{I}(x^*),\, a^{(i)T}d < 0\}} \dfrac{a^{(i)T}x^* - b_i}{-a^{(i)T}d}$,

tutti i vincoli $i \in \bar{I}(x^*)$ di tipo b_2 risultano soddisfatti per $0 \leq t < \bar{t}$

In conclusione per $0 \leq t \leq \bar{t}$ tutti i vincoli (sia di $I(x^*)$ che di $\bar{I}(x^*)$) vengono soddisfatti e quindi d è una direzione di discesa.

Se x^* è un punto di minimo, non esiste nessuna direzione in x^* che sia contemporaneamente ammissibile e di discesa.

Il sistema (I)

$$(\underline{I}) \left\langle \begin{array}{ll} \nabla f(x^*)^T d < 0 & \\ a^{(i)T} d \geq 0 & \underset{\sim}{i \in I(x^*)} \end{array} \right\rangle \text{ è inammissibile}$$

(se, infatti esistesse una soluzione $\bar{d}$ di tale sistema, sarebbe possibile "muoversi" lungo $\bar{d}$ riducendo f e conservando l'ammissibilità)

Il teorema di Farkas ci dice che il sistema (I) è inammissibile se e solo se il sistema $\underline{II}$ è ammissibile

$$\underline{II} \left\langle \begin{array}{l} \nabla f(x^*) = \sum\limits_{i \in I(x^*)} \lambda_i a^{(i)} \\ \lambda_i \geq 0, \ i \in I(x^*) \end{array} \right\rangle$$

Ricordando che $a^{(i)} = \nabla g_i(x^*) = \nabla g_i(x) \ \forall x$, vale il seguente teorema

Teorema (Condizione necessaria di ottimalità) Se x^* è un punto di minimo esistono dei moltiplicatori λ_i, associati ai vincoli attivi, non negativi, che consentono di esprimere il gradiente di f in x^* come combinazione lineare dei gradienti dei vincoli.

Nota Il teorema precedente definisce una condizione necessaria ma non sufficiente di ottimalità. Supponiamo infatti che il sistema

$$\begin{cases} \nabla f(x^*)^T d < 0 \\ a^{(i)T} d \geq 0, \ i \in I(x^*) \end{cases}$$

sia inammissibile. Ciò non esclude che fosse esistere una direzione di discesa $\bar{d}$ con $\nabla f(x^*)^T d = 0$ e $a^{(i)T} d \geq 0$, $i \in I(x^*)$ (Vedi esempio a pag 118)

Un modo alternativo di enunciare il teorema è:

Teorema x^* ottimo
$\Rightarrow$ sistema

$$\begin{cases} \nabla f(x^*) = \sum_{i=1}^{m} \lambda_i a^{(i)} \\ \lambda_i \geq 0, \quad i = 1, \dots, m \\ \lambda_i (a^{(i)T} x^* - b_i) = 0, \quad i = 1, \dots, m \end{cases}$$

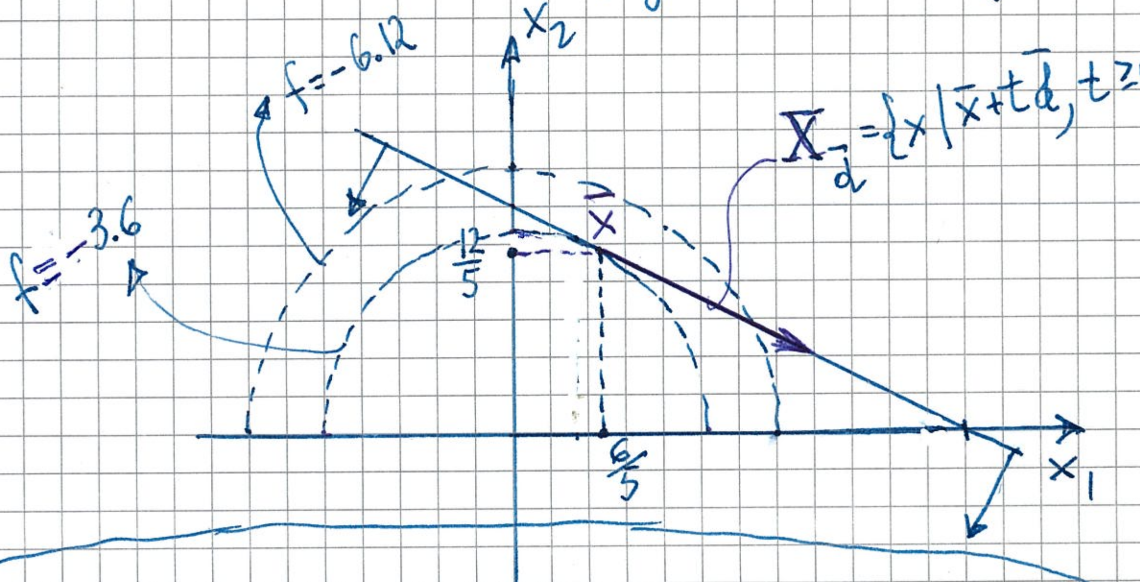
Condizione di complementarità: gli unici moltiplicatori che possono essere diversi da 0 sono quelli associati ai vincoli attivi in x^*

Esempio Il soddisfacimento delle condizioni studiate **non** è **sufficiente** a garantire l'ottimalità

Consideriamo (vedi esempio a pag. 112)

$$\begin{cases} \min -\frac{1}{2}(x_1^2+x_2^2) \\ -x_1-2x_2 \geq -6 \end{cases}$$

(Il grafico della funzione è una specie di "bicchiere rovesciato" $f \to -\infty$ sulla regione ammissibile)



Consideriamo il punto

$$\bar{x} = \begin{pmatrix} 6/5 \\ 12/5 \end{pmatrix} \qquad \nabla f(\bar{x}) = \begin{pmatrix} -\frac{6}{5} \\ -\frac{12}{5} \end{pmatrix}$$

$$a^{(1)} = \begin{pmatrix} -1 \\ -2 \end{pmatrix} \qquad I(\bar{x}) = \{1\}$$

$\Rightarrow$ $\nabla f(\bar{x}) = \lambda a^{(1)}$ è soddisfatta per $\lambda = \frac{6}{5} > 0$ $\Rightarrow$ Quindi il punto $\bar{x}$ soddisfa il sistema

(II) $\nabla f(\bar{x}) + \sum \lambda_i a^{(i)}$, $\lambda_i \geq 0$

$\Updownarrow$ Teor. Farkas

Il sistema (I)

$$\begin{cases} \nabla f(\bar{x})^T d < 0 \\ a^{(i)T} d \geq 0 \quad i \in I(\bar{x}) \end{cases}$$

è inammissibile !!!

Però il punto $\bar{x}$ **non** è un punto di minimo

Infatti la direzione $\bar{d} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$ è una **direzione di discesa** e vale

$$\begin{cases} \nabla f(\bar{x})^T \bar{d} = \left(-\frac{6}{5}, -\frac{12}{5}\right)\begin{pmatrix} 2 \\ -1 \end{pmatrix} = 0 \\ a^{(1)T} d = (-1, -2)\begin{pmatrix} 2 \\ -1 \end{pmatrix} = 0 \end{cases}$$

Condizioni necessarie di minimo vincolato (caso generale)

$$\begin{cases} \min f(x) \\ g_i(x) \geq 0 \,, \quad i = 1, \dots, m \end{cases}$$

sia x^* un punto di minimo e sia $I(x^*) = \{ i \mid g_i(x^*) = 0 \}$ l'insieme dei vincoli attivi in x^*.

Costruiamo un problema ausiliario, dipendente da x^* e ottenuto linearizzando i vincoli a partire da x^*. Sostituiamo alle $g_i(x)$ le loro linearizzazioni $g_i(x^*) + \nabla g_i(x^*)^T (x - x^*)$, $i \in I(x^*)$

Otteniamo il Problema $P(x^*)$

$$\begin{cases} \min f(x) \\ \underbrace{g_i(x^*)}_{0} + \nabla g_i(x^*)^T (x - x^*) \geq 0 \quad i \in I(x^*) \end{cases}$$

che è del tipo
$$\begin{cases} \min f(x) \\ a^{(i)T} x - b_i \geq 0 \quad i \in I(x^*) \end{cases}$$

dove si è posto $a^{(i)} = \nabla g_i(x^*)$ e $b_i = \nabla g_i(x^*)^T x^*$

Facciamo l'ipotesi (tutta da verificare!!) che il punto x^* sia punto di minimo anche per il problema $P(x^*)$.

Poichè $P(x^*)$ è un problema con vincoli lineari di ~~di~~ disuguaglianza

se x^* è un punto di minimo valgono le condizioni precedentemente studiate, cioè risulta soddisfatto il sistema

$$\begin{cases} \nabla f(x^*) = \sum_{i=1}^{m} \lambda_i a^{(i)} \\ \lambda_i \geq 0 \quad , \; i = 1, \dots, m \\ \lambda_i (a^{(i)T} x^* - b_i) = 0 \quad i = 1, \dots, m \end{cases}$$

Che possiamo riscrivere

$$\begin{cases} \nabla f(x^*) = \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) \\ \lambda_i \geq 0 \quad i = 1, \dots, m \\ \lambda_i \, g_i(x^*) = 0 \quad i = 1, \dots, m \end{cases} \qquad (KKT)$$

Queste condizioni furono scoperte (o inventate?...) intorno al 1940 e sono note come condizioni di Karush-Kuhn-Tucker (KKT)

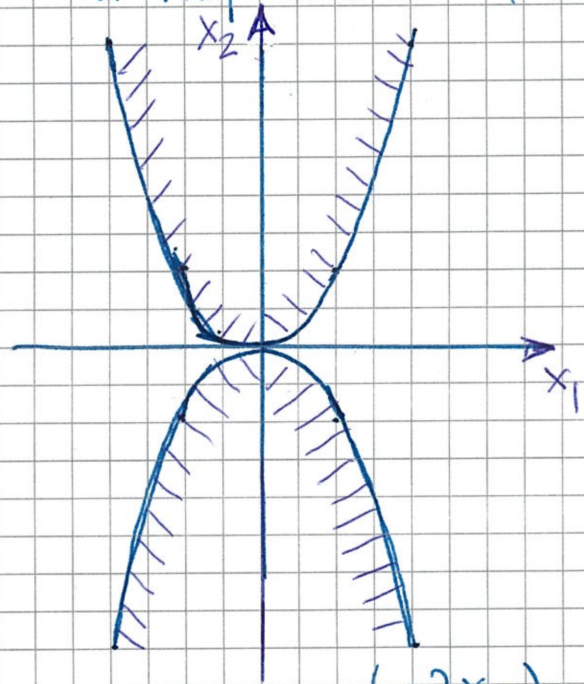
Quindi:

se x^* è un punto di minimo e l'ipotesi fatta è valida in x^* vengono soddisfatte le condizioni KKT

Ma l'ipotesi è valida? La risposta è "non sempre"

Esempio 1

$$\min x_1$$
$$x_2 \geq x_1^2$$
$$x_2 \leq -x_1^2$$



Questo "strano" problema ha solo un punto ammissibile in $\mathbb{R}^2$, cioè $\begin{pmatrix}0\\0\end{pmatrix}$

Essendo l'unico punto ammissibile, $x^* = \begin{pmatrix}0\\0\end{pmatrix}$ è anche il punto di minimo

Riscriviamo il problema nella forma generale

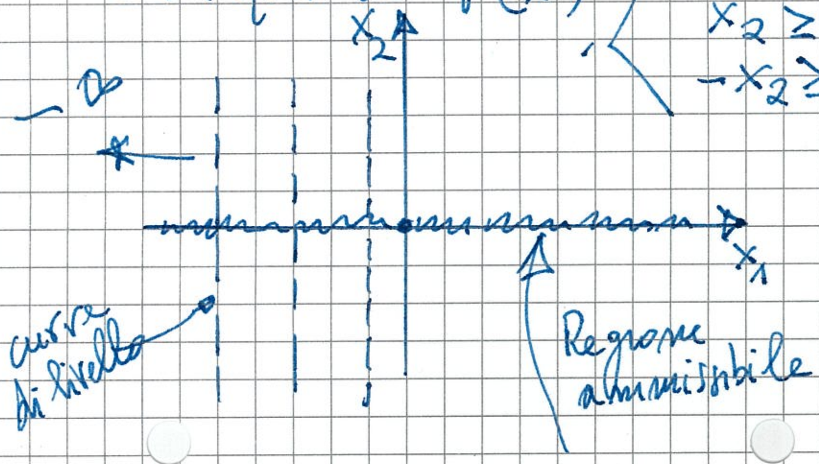
$$\min f(x_1, x_2) = \min x_1$$
$$g_1(x) = -x_1^2 + x_2 \geq 0$$
$$g_2(x) = -x_1^2 - x_2 \geq 0$$

$f(x^*) = 0$, $I(x^*) = \{1, 2\}$ $\quad \nabla g_1(x) = \begin{pmatrix}-2x_1\\1\end{pmatrix}$, $\nabla g_2(x) = \begin{pmatrix}-2x_1\\-1\end{pmatrix}$, $\nabla f(x) = \begin{pmatrix}1\\0\end{pmatrix}$

Quindi $a^{(1)} = \nabla g_1(x^*) = \begin{pmatrix}0\\1\end{pmatrix}$, $a^{(2)} = \nabla g_2(x^*) = \begin{pmatrix}0\\-1\end{pmatrix}$, $\nabla f(x^*) = \begin{pmatrix}1\\0\end{pmatrix}$, $b_1 = b_2 = 0$

Si ha quindi $P(x^*)$ $\left\langle \begin{array}{l} \min x_1 \\ x_2 \geq 0 \\ -x_2 \geq 0 \end{array} \right\rangle$ cioè $\left\langle \begin{array}{l} \min x_1 \\ x_2 = 0 \end{array} \right\rangle$

x^*, punto di minimo, non è punto di minimo del problema linearizzato



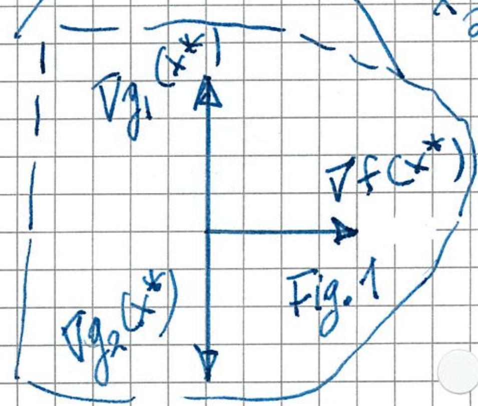
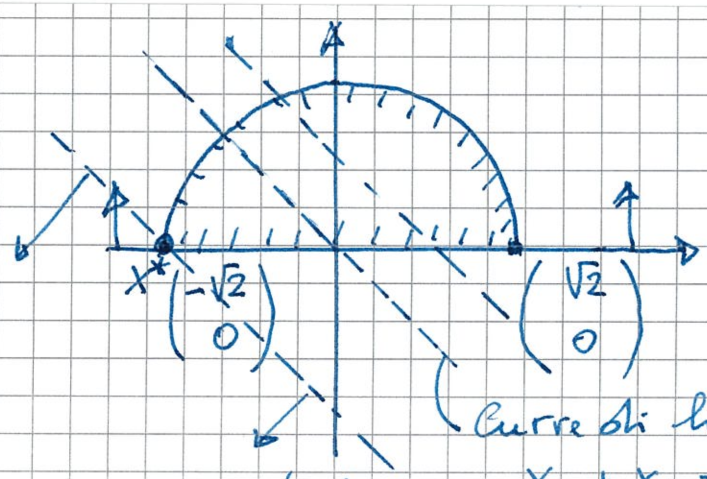


Fig. 1

KKT non valgono: ∇f non è esprimibile come combinazione nonnegativa dei ∇g_i

Esempio 2 $\begin{cases} \min \; x_1 + x_2 \\ -x_1^2 - x_2^2 \geq -2 \\ x_2 \geq 0 \end{cases}$



Curve di livello $x_1 + x_2 = K$

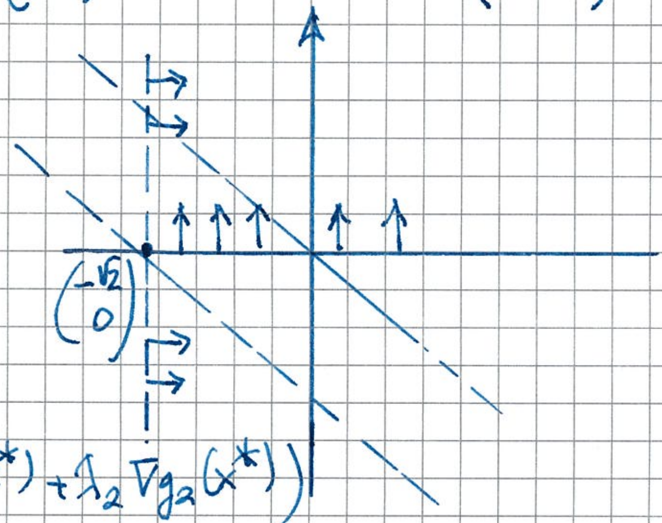
Chiaramente il punto di minimo è $x^* = \begin{pmatrix} -\sqrt{2} \\ 0 \end{pmatrix}$

$\nabla f(x) = \nabla f(x^*) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \nabla g_1(x) = \begin{pmatrix} -2x_1 \\ -2x_2 \end{pmatrix}, \quad \nabla g_2(x) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

$I(x^*) = \{1, 2\}$; $\nabla g_1(x^*) = \begin{pmatrix} 2\sqrt{2} \\ 0 \end{pmatrix}, \quad \nabla g_2(x^*) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ $\quad b_1 = (2\sqrt{2}, 0)\begin{pmatrix} -\sqrt{2} \\ 0 \end{pmatrix} = -4$

$b_2 = (0, 1)\begin{pmatrix} -\sqrt{2} \\ 0 \end{pmatrix} = 0$

$P(x^*) \quad \begin{cases} \min \; x_1 + x_2 \\ 2\sqrt{2}\, x_1 \geq -4 \\ x_2 \geq 0 \end{cases}$ Cioè $\begin{cases} \min \; x_1 + x_2 \\ x_1 \geq -\sqrt{2} \\ x_2 \geq 0 \end{cases}$



x^* è punto di minimo anche per $P(x^*)$

$\Longrightarrow$ Quindi valgono KKT $\left(\nabla f(x^*) = \lambda_1 \nabla g_1(x^*) + \lambda_2 \nabla g_2(x^*) \right)$

con $\lambda_1 = \dfrac{1}{2\sqrt{2}}$ e $\lambda_2 = 1$

Fig 2

In definitiva quando valgono le condizioni di KKT?

Dipende dalla qualità dei vincoli ("Constraint qualification")

Nel primo esempio (vedi Fig. 1 a pag 121) i gradienti dei vincoli attivi sono linearmente dipendenti, mentre nel secondo esempio (vedi Fig. 2 a pag 122) essi sono linearmente indipendenti

La indipendenza lineare dei gradienti dei vincoli è una "constraint qualification" (C.Q.)

Un altra C.Q. è l'esistenza di almeno un punto x in cui sia $g_i(x) > 0 \quad \forall i = 1, .., m$ (Condizione di Slater)

Un punto che soddisfa almeno una C.Q. è detto "regolare per i vincoli"

In conclusione vale il teorema

Teorema. Se x^* è un punto di minimo e x^* è regolare per i vincoli allora in x^* sono soddisfatte le condizioni KKT