



UNIVERSITÀ DEGLI STUDI DI FERRARA

**FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E
NATURALI**

**Corso di Laurea Specialistica in
MATEMATICA**

**IL LEMMA DI FARKAS E LE CONDIZIONI
DI KARUSH-KUHN-TUCKER
NELL'OTTIMIZZAZIONE LINEARE E
NON LINEARE**

Relatori:

**Chiar.mo Prof.
Josef Eschgfäller**

Dott.ssa Silvia Bonettini

Laureanda:

Monica Gazzetta

Anno Accademico 2008-2009

Indice

Introduzione	3
I. IL METODO DEL SIMPLESSO	
1. Il lemma di Farkas	5
2. Programmi lineari in forma standard	13
3. Programmi lineari in forma normale	18
4. Il metodo del simplesso	26
5. Geometria elementare degli insiemi convessi	31
II. ALCUNE APPLICAZIONI	
6. Il separatore di Bennett-Mangasarian	37
7. Grafi	40
8. Flussi in un grafo diretto	44
III. OTTIMIZZAZIONE NON LINEARE	
9. Il cono tangente	54
10. Le condizioni di Karush-Kuhn-Tucker	57
11. Il caso generale	60
12. Restrizioni lineari	64
13. Restrizioni convesse	65
Bibliografia	69

Introduzione

Lo scopo di questa tesi è quello di presentare ed approfondire (anche dal punto di vista geometrico ed algebrico) alcuni concetti basilari della programmazione lineare e non lineare che sono il lemma di Farkas e le condizioni di Karush-Kuhn-Tucker (dette anche condizioni di KKT). In questo lavoro inoltre, è stato descritto un algoritmo di risoluzione per problemi di programmazione lineare, il metodo del simplesso, e sono state presentate alcune sue possibili applicazioni pratiche.

Il lemma di Farkas è l'argomento del primo capitolo ed è forse il più importante teorema di alternativa per sistemi lineari e viene inoltre utilizzato in numerosi contesti come la programmazione lineare e non lineare, la teoria dei sistemi di disequazioni lineari, i modelli economici multisettoriali e la teoria dei giochi.

Questo lemma afferma che il sistema $Ax = b$ (dove A è una matrice $m \times n$ e x e b sono vettori rispettivamente appartenenti ad \mathbb{R}^m ed \mathbb{R}^n) con la condizione $x \geq 0$ ammette soluzione se e solo se non la ammette il sistema costituito dalle due condizioni $yA \geq 0$ e $yb < 0$ con $y \in \mathbb{R}^m$.

Grazie al lemma di Farkas, si può dimostrare, sotto opportune ipotesi, che se un punto è soluzione di un problema di programmazione in generale non lineare, allora in tale punto sono soddisfatte le condizioni di KKT.

Le condizioni di Karush-Kuhn-Tucker sono descritte nel capitolo 10 e costituiscono la parte fondamentale della terza sezione della tesi. Esse sono una generalizzazione del metodo dei moltiplicatori di Lagrange applicato a problemi in cui sono presenti anche vincoli di disuguaglianza e rappresentano lo scheletro della programmazione non lineare costituendo condizioni necessarie per la soluzione di un problema di programmazione non lineare in cui i vincoli soddisfano opportune condizioni di regolarità dette condizioni di qualificazione dei vincoli. Esempi di queste qualificazioni che vengono trattate nella tesi sono il requisito di indipendenza lineare dei vincoli e il requisito di Mangasarian-Fromovitz. Per problemi di programmazione lineare invece, le condizioni di KKT sono sia necessarie che sufficienti. Queste questioni vengono descritte nei capitoli 11, 12 e 13 dove sono presenti rispettivamente vincoli generici, vincoli lineari e vincoli convessi. Nel capitolo 9 diamo alcune definizioni che caratterizzano le condizioni di KKT dal punto di vista geometrico.

La prima parte della tesi riguarda la teoria e l'implementazione del metodo del simplesso. Questo algoritmo è stato inventato dal matematico e statistico statunitense George B. Dantzig nel 1947 e rappresenta ancora oggi uno dei metodi più famosi e più efficaci nella risoluzione di problemi di programmazione lineare, problemi in cui si cerca di massimizzare o minimizzare una funzione lineare sotto opportuni vincoli di uguaglianza e disuguaglianza anch'essi lineari.

L'insieme di definizione dei vincoli è detto insieme ammissibile ed è un poliedro convesso. Se il problema di programmazione lineare am-

mette una soluzione ottimale, questa è anche un vertice del poliedro (teorema fondamentale della programmazione lineare, cap. 3).

Queste considerazioni geometriche sono alla base del metodo inventato da Dantzig, la cui idea principale può quindi essere così descritta: si determina un vertice iniziale del poliedro e si decide se tale punto è una soluzione ottima; se non lo è si determina in modo adeguato un nuovo vertice del poliedro e se ne testa l'ottimalità e così via.

L'algoritmo del simplesso si applica in una prima formulazione ai programmi lineari nella forma $fx = \max$ (risp. \min), $x \geq 0$, $Ax = b$ (forma normale) oggetto del capitolo 3. Questo tuttavia non è limitativo perchè, come viene descritto nel cap. 2, ogni altro programma lineare può essere ricondotto alla forma normale utilizzando opportune variabili di slack o di surplus. Gli aspetti teorici del metodo del simplesso vengono approfonditi nel capitolo 3 mentre l'algoritmo tradotto in Matlab (o Octave) è presentato nel quarto capitolo. Nel capitolo 5 vengono invece descritte alcune proprietà degli insiemi convessi.

Nella seconda parte della tesi esponiamo alcuni possibili campi di applicazione del metodo del simplesso: la teoria dei grafi e dei flussi, risp. capitoli 7 e 8, e il separatore di Bennett-Mangasarian nel cap. 6. In quest'ultimo in particolare si cerca di far vedere come il problema di trovare un iperpiano che separa due insiemi di punti distinti possa essere ricondotto ad un problema di programmazione lineare. È un risultato che viene ad esempio utilizzato nella ricerca sperimentale riguardante la diagnosi e la prognosi del carcinoma mammario.

È opportuno ricordare che oltre al metodo del simplesso esistono altri metodi altrettanto efficaci per la risoluzione di problemi di programmazione lineare e non lineare: i metodi a punto interno. Questi dal punto di vista geometrico si differenziano notevolmente dal metodo del simplesso in quanto la ricerca della soluzione ottimale non viene effettuata esaminando i vertici del poliedro ammissibile ma rimanendo all'interno dell'insieme di definizione dei vincoli. I metodi a punto interno discendono dalle condizioni di KKT opportunamente modificate (in modo da rimanere sempre all'interno della regione ammissibile) a cui viene applicato il metodo di Newton.

I. IL METODO DEL SIMPLESSO

1. Il lemma di Farkas

Situazione 1.1. $n, m \in \mathbb{N} + 1$.

Definizione 1.2. Usiamo le seguenti notazioni:

$\mathbb{R}_m^n :=$ insieme delle matrici reali con n righe ed m colonne

$\mathbb{R}_m := \mathbb{R}_m^1 =$ insieme dei vettori riga reali con m coefficienti

$\mathbb{R}_+^n := \{x \in \mathbb{R}^n \mid x \geq 0\}$

$\mathbb{R}_m^+ := \{f \in \mathbb{R}_m \mid f \geq 0\}$

$\mathbb{R}^+ := [0, \infty)$

Definizione 1.3. Siano $A \in \mathbb{R}_n^m$, $X \subset \mathbb{R}^n$, $U \subset \mathbb{R}_m$, $b \in \mathbb{R}^m$, $f \in \mathbb{R}_n$. Definiamo allora:

$AX := \{Ax \mid x \in X\}$

$UA := \{uA \mid u \in U\}$

$(AX = b) := \{x \in X \mid Ax = b\}$

$(UA = f) := \{u \in U \mid uA = f\}$

$(AX \leq b) := \{x \in X \mid Ax \leq b\}$

$(UA \geq f) := \{u \in U \mid uA \geq f\}$

In modo analogo sono definiti $(AX \geq b)$, $(UA \leq f)$, ecc.

Avremo quindi ad esempio $(A\mathbb{R}_+^n = b) = \{x \in \mathbb{R}_+^n \mid Ax = b\}$.

Nota 1.4. Useremo le seguenti notazioni e formule ben note del calcolo matriciale:

(1) Per $A \in \mathbb{R}_n^m$ denotiamo con A^i la i -esima riga, con A_j la j -esima colonna di A .

(2) Per $A \in \mathbb{R}_n^m$, $B \in \mathbb{R}_s^n$ abbiamo allora

$$(AB)_j^i = A^i B_j$$

$$(AB)^i = A^i B$$

$$(AB)_j = AB_j$$

$$AB = \sum_{\alpha=1}^n A_\alpha B^\alpha$$

Abbiamo quindi in particolare, per $x \in \mathbb{R}^n$ ed $f \in \mathbb{R}_m$,

$$Ax = \sum_{\alpha=1}^n A_{\alpha} x^{\alpha}$$

$$fA = \sum_{\alpha=1}^m f_{\alpha} A^{\alpha}$$

- (3) Con $\|x, y\| = x^t y$ risp. $\|f, g\| = f g^t$ denoteremo il prodotto scalare di $x, y \in \mathbb{R}^n$ risp. $f, g \in \mathbb{R}^m$.
- (4) Per vettori v_1, \dots, v_k in uno spazio vettoriale sia $SV(v_1, \dots, v_k)$ il sottospazio vettoriale generato da questi vettori. Similmente per un sottoinsieme X denoteremo con $SV(X)$ il sottospazio vettoriale generato da X .
- (5) Per $X \subset \mathbb{R}^n$ sia $X^{\perp} := \{v \in \mathbb{R}^n \mid \|x, v\| = 0 \text{ per ogni } x \in X\}$. X^{\perp} è un sottospazio vettoriale di \mathbb{R}^n , anche quando X stesso non è un sottospazio vettoriale.
- (6) W sia un sottospazio vettoriale di \mathbb{R}^n . Allora:
 $\dim W + \dim W^{\perp} = n$
 $W^{\perp\perp} = W$
 $\mathbb{R}^n = W \oplus W^{\perp}$
- (7) Per vettori v_1, \dots, v_k in uno spazio vettoriale reale sia $POS(v_1, \dots, v_k) := \{\lambda_1 v_1 + \dots + \lambda_k v_k \mid \lambda_1, \dots, \lambda_k \in \mathbb{R}^+\}$. Similmente per un sottoinsieme X sia $POS(X) := \{\lambda_1 x_1 + \dots + \lambda_k x_k \mid k \geq 1, \lambda_i \in \mathbb{R}^+, x_i \in X \text{ per } i = 1, \dots, k\}$.

Osservazione 1.5. Sia $A \in \mathbb{R}_n^m$. Allora:

- (1) $A\mathbb{R}^n = SV(A_1, \dots, A_n)$.
(2) $A\mathbb{R}_+^n = POS(A_1, \dots, A_n)$.

Dimostrazione. (1) Sia $x \in \mathbb{R}^n$. Allora $Ax = \sum_{k=1}^n A_k x^k \in SV(A_1, \dots, A_n)$ e viceversa. Il punto (2) si dimostra in modo analogo considerando $x \in \mathbb{R}_+^n$.

Lemma 1.6. Siano $A \in \mathbb{R}_n^m$ e $b \in \mathbb{R}^m$. Allora sono equivalenti:

- (1) $b \in A\mathbb{R}^n$.
(2) $(\mathbb{R}_m A = 0)b = 0$.

Dimostrazione. (1) \implies (2): Sia $f \in \mathbb{R}_m$ con $fA = 0$. Poiché per ipotesi $b \in A\mathbb{R}^n$, esiste $x \in \mathbb{R}^n$ tale che $b = Ax$. Allora $fb = fAx = 0$.

(2) \implies (1): Sia $b \notin A\mathbb{R}^n = SV(A_1, \dots, A_n) =: W$. Per la nota 1.4 $b = w + y$ con $w \in W$ e $y \in W^{\perp}$. Necessariamente $y \neq 0$ perché $b \notin W$. Abbiamo

$$\|b, y\| = \|w + y, y\| = \|w, y\| + \|y, y\| = \|y, y\| \neq 0$$

Inoltre $\|A_j, y\| = 0$ per ogni j perché $y \in W^\perp$. Con $f := y^t$ abbiamo allora $fA = 0$ ed $fb \neq 0$.

Definizione 1.7. V sia uno spazio vettoriale reale ed $x, y \in V$.

Poniamo allora

$$[x, y] := x + [0, 1](y - x) = \{x + t(y - x) \mid t \in [0, 1]\}$$

Un sottoinsieme $X \subset V$ si dice *convesso* se per ogni $x, y \in X$ si ha $[x, y] \subset X$.

Per $t \in \mathbb{R}$ sia $[x, y]_t := x + t(y - x)$.

Osservazione 1.8. V e W siano spazi vettoriali reali ed $x, y \in V$.

$\varphi : V \rightarrow W$ sia un'applicazione affine. Allora:

- (1) $\varphi([x, y]_t) = [\varphi(x), \varphi(y)]_t$ per ogni $t \in \mathbb{R}$.
- (2) $\varphi([x, y]) = [\varphi(x), \varphi(y)]$.
- (3) Se X è un sottoinsieme convesso di V , allora $\varphi(X)$ è un sottoinsieme convesso di W .

Dimostrazione. (1) Per ipotesi esistono un'applicazione lineare $\varphi_0 : V \rightarrow W$ e un vettore $b \in W$ tali che $\varphi = \varphi_0 + b$. Abbiamo allora $\varphi(x + t(y - x)) = \varphi_0(x + t(y - x)) + b = \varphi_0(x) + t\varphi_0(y) - t\varphi_0(x) + b = \varphi_0(x) + b + t(\varphi_0(y) + b - \varphi_0(x) - b) = \varphi(x) + t(\varphi(y) - \varphi(x)) = [\varphi(x), \varphi(y)]_t$.

(2) Dal punto (1) abbiamo direttamente $\varphi([x, y]) \subset [\varphi(x), \varphi(y)]$.

Sia $z \in [\varphi(x), \varphi(y)]$, ad esempio $z = [\varphi(x), \varphi(y)]_t$ con $t \in [0, 1]$. Allora $z = \varphi([x, y]_t) \in \varphi([x, y])$.

(3) Siano $u = \varphi(x)$ e $v = \varphi(y)$ con $x, y \in X$. Allora $[u, v] = [\varphi(x), \varphi(y)] = \varphi([x, y]) \subset \varphi(X)$. Nell'ultima inclusione abbiamo sfruttato il fatto che $[x, y] \subset X$ perché X è convesso.

Lemma 1.9. v_1, \dots, v_k siano vettori linearmente indipendenti di \mathbb{R}^m . Allora l'insieme $\text{POS}(v_1, \dots, v_k)$ è chiuso.

Dimostrazione. Possiamo trovare v_{k+1}, \dots, v_m tali che $v_1, \dots, v_k, v_{k+1}, \dots, v_m$ sia una base di \mathbb{R}^m . Per ogni $x \in \mathbb{R}^m$ esistono $\lambda^1(x), \dots, \lambda^m(x) \in \mathbb{R}$ tali che $x = \lambda^1(x)v_1 + \dots + \lambda^m(x)v_m$. Le funzioni $\lambda^i : \mathbb{R}^m \rightarrow \mathbb{R}$ che così otteniamo sono continue. Perciò l'insieme $\text{POS}(v_1, \dots, v_k) = (\lambda^1 \geq 0, \dots, \lambda^k \geq 0, \lambda^{k+1} = 0, \dots, \lambda^m = 0)$ è chiuso.

Proposizione 1.10 (lemma di Carathéodory). Sia $X \subset \mathbb{R}^m$ e $\text{SV}(X) \neq 0$. Sia $d := \dim \text{SV}(X)$. Allora:

- (1) Ogni punto di $\text{POS}(X)$ è combinazione lineare a coefficienti non negativi di al massimo d punti linearmente indipendenti di X .
- (2) Sia \mathcal{D} la famiglia di tutti gli insiemi che consistono di esattamente d vettori linearmente indipendenti di X .

$$\text{Allora } \text{POS}(X) = \bigcup_{D \in \mathcal{D}} \text{POS}(D).$$

Dimostrazione. (1) Sia $v \in \text{POS}(X)$. Esistono quindi $\lambda_1, \dots, \lambda_k \geq 0$ ed $x_1, \dots, x_k \in X$ tali che $v = \lambda_1 x_1 + \dots + \lambda_k x_k$. Scegliamo k minimale (ciò implica $\lambda_j > 0$ per ogni j) e supponiamo per assurdo che $k > d$.

Siccome per ipotesi $\dim \text{SV}(X) = d$, i vettori x_1, \dots, x_k devono essere linearmente dipendenti, perciò esistono $\alpha_1, \dots, \alpha_k \in \mathbb{R}$ non tutti nulli tali che $\alpha_1 x_1 + \dots + \alpha_k x_k = 0$.

Possiamo assumere che $\alpha_1 > 0$ e che $\frac{\alpha_1}{\lambda_1} \geq \frac{\alpha_j}{\lambda_j}$ per ogni j . Ciò implica $\lambda_j \geq \frac{\alpha_j}{\alpha_1} \lambda_1$ per ogni j . Abbiamo però $\sum_{j=1}^k (\lambda_j - \frac{\alpha_j}{\alpha_1} \lambda_1) x_j = v$.

In questa rappresentazione tutti i coefficienti sono ≥ 0 e il primo è nullo, in contraddizione alla minimalità di k .

(2) Ciò implica anche che $\text{POS}(X) \subset \bigcup_{D \in \mathcal{D}} \text{POS}(D)$ perché per il punto (1) per ogni elemento $x \in \text{POS}(X)$ esistono elementi $x_1, \dots, x_k \in X$ linearmente indipendenti con $x \in \text{POS}(x_1, \dots, x_k)$ e $k \leq d$; se $k < d$, allora possiamo trovare x_{k+1}, \dots, x_d tali che $\{x_1, \dots, x_k, x_{k+1}, \dots, x_d\} =: D \in \mathcal{D}$ con naturalmente ancora $x \in \text{POS}(D)$. Viceversa $\text{POS}(D) \subset \text{POS}(X)$ per ogni $D \in \mathcal{D}$.

Proposizione 1.11. Siano $v_1, \dots, v_k \in \mathbb{R}^m$. Allora l'insieme $\text{POS}(v_1, \dots, v_k)$ è chiuso.

Dimostrazione. Sia $X = \{v_1, \dots, v_k\}$. \mathcal{D} sia definito come nella prop. 1.10. Allora $\text{POS}(X) = \bigcup_{D \in \mathcal{D}} \text{POS}(D)$. Per il lemma 1.9 per ogni $D \in \mathcal{D}$ l'insieme $\text{POS}(D)$ è chiuso. Nel nostro caso \mathcal{D} è un insieme finito e quindi anche $\text{POS}(X)$ è chiuso.

Corollario 1.12. Sia $A \in \mathbb{R}_+^n$. Allora l'insieme $A\mathbb{R}_+^n$ è chiuso in \mathbb{R}^m .

Dimostrazione. Per l'oss. 1.5 abbiamo $A\mathbb{R}_+^n = \text{POS}(A_1, \dots, A_n)$. L'enunciato segue dalla proposizione 1.11.

Osservazione 1.13. L'enunciato del cor. 1.12 non è banale, perché in genere un'applicazione lineare $\mathbb{R}^n \rightarrow \mathbb{R}^m$ non trasforma chiusi in chiusi. Un esempio è la proiezione sulla prima coordinata $\mathbb{R}^2 \rightarrow \mathbb{R}$ che trasforma l'iperbole $\{x \mid x^1 x^2 = 1\}$, un insieme chiuso, in $\mathbb{R} \setminus 0$.

Osservazione 1.14. X sia un sottoinsieme chiuso e non vuoto di \mathbb{R}^m ed $y \in \mathbb{R}^m$. Allora esiste un punto $p \in X$ che possiede distanza minimale da y .

Dimostrazione. Siccome $X \neq \emptyset$, possiamo scegliere un punto (arbitrario) $z \in X$. Allora l'insieme $A =: \{a \in X \mid |a - y| \leq |z - y|\}$ è compatto e non vuoto, per cui la funzione continua $\bigcirc_a |a - y| : A \rightarrow \mathbb{R}$ assume un minimo in qualche punto p .

Lemma 1.15 (legge del parallelogramma). Siano $x, y \in \mathbb{R}^m$. Allora

$$|x + y|^2 + |x - y|^2 = 2|x|^2 + 2|y|^2$$

Dimostrazione. $|x + y|^2 + |x - y|^2 = |x|^2 + |y|^2 + 2\|x, y\| + |x|^2 + |y|^2 - 2\|x, y\| = 2|x|^2 + 2|y|^2$.

Teorema 1.16. *X sia un sottoinsieme chiuso, convesso e non vuoto di \mathbb{R}^m ed $y \in \mathbb{R}^m$. Allora esiste un punto $p \in X$ che possiede distanza minimale da y . p è univocamente determinato.*

Dimostrazione. (1) L'esistenza è stata dimostrata nell'oss.1.14.

(2) Dimostriamo l'unicità. Sia $q \in X$ tale che $|p - y| = |q - y|$. Allora $r := \frac{p+q}{2} \in X$ perchè X è convesso. Inoltre

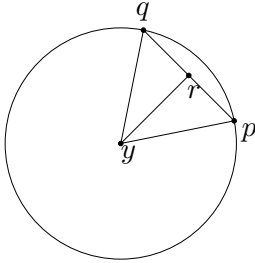
$$|r - y|^2 = |p - y|^2 - \frac{1}{4}|p - q|^2 \quad (*)$$

Infatti per il lemma 1.15 applicato a $x = p - y$ e $y = q - y$ risulta

$$\begin{aligned} 4|r - y|^2 &= 4\left|\frac{p+q}{2} - y\right|^2 \\ &= |(p - y) + (q - y)|^2 \\ &= 2|p - y|^2 + 2|q - y|^2 - |p - q|^2 \\ &= 4|p - y|^2 - |p - q|^2 \end{aligned}$$

dove abbiamo utilizzato l'ipotesi $|p - y| = |q - y|$. Ciò mostra l'uguaglianza (*). Se fosse $p \neq q$, risulterebbe $|r - y| < |p - y|$ e p non avrebbe più distanza minimale da y .

L'uguaglianza (*) e la relazione $|r - y|^2 < |p - y|^2$ per $p \neq q$ sono anche evidenti dalla figura.



Definizione 1.17. Nella situazione del teorema 1.16 il punto p si chiama la *proiezione* di y su X . È immediato che $y = p$ se e solo se $y \in X$.

Definizione 1.18. Per $p, v \in \mathbb{R}^m$ poniamo:

$$H(p, v) := \{x \in \mathbb{R}^m \mid \|x - p, v\| = 0\}$$

$$O(p, v) := \{x \in \mathbb{R}^m \mid \|x - p, v\| \leq 0\}$$

$$SO(p, v) := \{x \in \mathbb{R}^m \mid \|x - p, v\| < 0\}$$

Se $v \neq 0$, allora $H(p, v)$ è l'iperpiano ortogonale a v passante per p , $O(p, v)$ l'insieme dei punti che, riferendoci a $p + v$, si trovano *oltre* o su questo iperpiano, $O(p, -v)$ l'insieme dei punti che non si trovano oltre questo iperpiano, cioè dalla stessa parte di $p + v$. SO significa *strettamente oltre*.

Proposizione 1.19. *X sia un sottoinsieme chiuso, convesso e non vuoto di \mathbb{R}^m ed $y \in \mathbb{R}^m$. Per un punto $p \in X$ sono allora equivalenti:*

- (1) p è la proiezione di y su X .
- (2) $X \subset O(p, y - p)$.

Dimostrazione. (1) \implies (2): Siano p la proiezione di y su X ed $x \in X$. Allora per ogni $t \in [0, 1]$ il punto $p + t(x - p) \in X$ perchè X è convesso. Per la definizione di p abbiamo

$$\begin{aligned} \|y - p\|^2 &\leq \|y - (p + t(x - p))\|^2 = \\ &= \|(y - p) - t(x - p)\|^2 = \\ &= \|y - p\|^2 + t^2\|x - p\|^2 - 2t\|y - p, x - p\| \end{aligned}$$

ovvero

$$2t\|y - p, x - p\| \leq t^2\|x - p\|^2$$

Per ogni $t \in (0, 1]$ abbiamo perciò $2\|y - p, x - p\| \leq t\|x - p\|^2$.

Ciò è possibile solo se $\|y - p, x - p\| \leq 0$.

(2) \implies (1): Siano $X \subset O(p, y - p)$ ed $x \in X$. Allora $\|y - p, p - x\| \geq 0$ e quindi

$$\|y - x\|^2 = \|y - p + p - x\|^2 = \|y - p\|^2 + \|p - x\|^2 + 2\|y - p, p - x\| \geq \|y - p\|^2$$

Lemma 1.20. *X sia un sottoinsieme chiuso, convesso e non vuoto di \mathbb{R}^m ed $y \in \mathbb{R}^m$. p sia la proiezione di y su X . Allora:*

- (1) $X \subset O(p + t(y - p), y - p)$ per ogni $t \geq 0$.
- (2) In particolare $X \subset O(y, y - p)$.
- (3) Se $y \notin X$, allora $X \subset SO(y, y - p)$.

Dimostrazione. Sia $x \in X$. Allora per $t \geq 0$ si ha

$$\|(x - p) - t(y - p), y - p\| = \|x - p, y - p\| - t\|y - p\|^2 \leq 0$$

essendo per la prop 1.19 $\|x - p, y - p\| \leq 0$.

Sia ora $y \notin X$. Allora $y \neq p$ e quindi per $t = 1$ otteniamo

$$\|x - y, y - p\| = \|x - p, y - p\| - \|y - p\|^2 < 0.$$

Lemma 1.21. *X sia un sottoinsieme chiuso, convesso e non vuoto di \mathbb{R}^m ed $y \in \mathbb{R}^m \setminus X$. p sia la proiezione di y su X ed $y^* := \frac{p+y}{2}$. Allora:*

$$y \in SO(y^*, p - y)$$

$$X \subset SO(y^*, y - p)$$

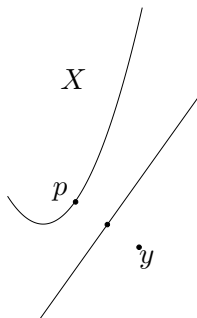
Dimostrazione.

$$(1) 2\|y - y^*, p - y\| = 2\|y - \frac{p+y}{2}, p - y\| = -\|y - p, y - p\| < 0$$

(2) Sia $x \in X$. Allora

$$\begin{aligned} 2\|x - y^*, y - p\| &= 2\|x - \frac{p+y}{2}, y - p\| \\ &= \|x - p, y - p\| + \|x - y, y - p\| < 0 \end{aligned}$$

Infatti per la prop. 1.19 abbiamo $\|x - p, y - p\| \leq 0$, mentre dal lemma 1.20 segue che $\|x - y, y - p\| < 0$.



Corollario 1.22. X sia un sottoinsieme chiuso, convesso e non vuoto di \mathbb{R}^m ed $y \in \mathbb{R}^m \setminus X$. p sia la proiezione di y su X . Sia ancora $y^* := \frac{p+y}{2}$. Allora

$$\|y, y - p\| > \|y^*, y - p\| > \|x, y - p\|$$

per ogni $x \in X$.

Dimostrazione. Sia $x \in X$. Per il lemma 1.21 abbiamo

$$\|y - y^*, y - p\| > 0 > \|x - y^*, y - p\|$$

ovvero

$$\|y, y - p\| - \|y^*, y - p\| > 0 > \|x, y - p\| - \|y^*, y - p\|$$

e quindi

$$\|y, y - p\| > \|y^*, y - p\| > \|x, y - p\|$$

Corollario 1.23. X sia un sottoinsieme chiuso e convesso di \mathbb{R}^m ed $y \in \mathbb{R}^m \setminus X$. Allora esistono $\alpha \in \mathbb{R}$ ed $f \in \mathbb{R}^m$ tali che $f \cdot y > \alpha > f \cdot x$ per ogni $x \in X$. Sostituendo f con $-f$ ed α con $-\alpha$ possiamo naturalmente anche ottenere $f \cdot y < \alpha < f \cdot x$ per ogni $x \in X$.

Dimostrazione. L'enunciato è banale per $X = \emptyset$ e segue altrimenti dal corollario 1.22, se poniamo $f := (y - p)^t$ ed $\alpha := \|y^*, y - p\|$

Osservazione 1.24. Sia $X \subset \mathbb{R}^m$. Allora $\text{POS}(X)$ è convesso.

Dimostrazione. Siano $u, v \in \text{POS}(X)$. Allora esistono $\lambda_1, \dots, \lambda_k, \mu_1, \dots, \mu_l \in \mathbb{R}^+$ e $x_1, \dots, x_k, y_1, \dots, y_l \in X$ tali che $u = \lambda_1 x_1 + \dots + \lambda_k x_k$ e $v = \mu_1 y_1 + \dots + \mu_l y_l$. Sia $t \in [0, 1]$. Allora

$$\begin{aligned} u + t(v - u) &= \lambda_1 x_1 + \dots + \lambda_k x_k + t\mu_1 y_1 + \dots + t\mu_l y_l - t\lambda_1 x_1 - \dots - t\lambda_k x_k \\ &= (1 - t)\lambda_1 x_1 + \dots + (1 - t)\lambda_k x_k + t\mu_1 y_1 + \dots + t\mu_l y_l \end{aligned}$$

I coefficienti $(1 - t)\lambda_i$ e $t\mu_j$ sono tutti non negativi, per cui vediamo che $u + t(v - u) \in \text{POS}(X)$.

Corollario 1.25. Sia $A \in \mathbb{R}_n^m$. Allora $A\mathbb{R}_+^n$ è convesso.

Dimostrazione. Ciò segue dall'oss. 1.24 perché $A\mathbb{R}_+^n = \text{POS}(A_1, \dots, A_n)$ per l'oss. 1.5.

Teorema 1.26 (lemma di Farkas). Siano $A \in \mathbb{R}_n^m$ e $b \in \mathbb{R}^m$. Allora sono equivalenti:

- (1) $b \in A\mathbb{R}_+^n$.
- (2) $(\mathbb{R}_m A \geq 0)b \geq 0$.

Dimostrazione. (1) \implies (2): Siano $x \in \mathbb{R}_+^n$ ed $f \in \mathbb{R}_m$ tali che $b = Ax$ ed $fA \geq 0$. Allora $fb = fAx \geq 0$.

(2) \implies (1): Sia $b \notin A\mathbb{R}_+^n$. Dai corollari 1.12 e 1.25 sappiamo che $A\mathbb{R}_+^n$ è un sottoinsieme convesso, chiuso e non vuoto di \mathbb{R}^m ; quindi per il corollario 1.23 esistono $\alpha \in \mathbb{R}$ e $f \in \mathbb{R}_m$ tali che $fb < \alpha < fAx$ per ogni $x \in \mathbb{R}_+^n$. Ciò deve valere in particolare per $x = 0$ e questo implica $\alpha < 0$ e quindi anche $fb < 0$.

D'altra parte però $fA \geq 0$ e ciò è in contraddizione con l'ipotesi (2). Infatti sia ad esempio $fA_j < 0$. Siccome $x := tA_j = At\delta_j \in A\mathbb{R}_+^n$, abbiamo $\alpha < ftA_j = tfA_j$ per ogni $t \geq 0$, ma ciò non è possibile perché $\lim_{t \rightarrow \infty} tfA_j = -\infty$.

Osservazione 1.27. Siano $A \in \mathbb{R}_n^m$ e $b \in \mathbb{R}^m$. Denotiamo con δ la matrice identica in \mathbb{R}_n^n . Allora:

- (1) $(A \ \delta)\mathbb{R}_+^n = \{Ax + u \mid x \in \mathbb{R}_+^n, u \in \mathbb{R}_+^m\}$.
- (2) $(A\mathbb{R}_+^n \leq b) \neq \emptyset \iff b \in (A \ \delta)\mathbb{R}_+^n$.

Proposizione 1.28. Siano $A \in \mathbb{R}_n^m$ e $b \in \mathbb{R}^m$. Allora sono equivalenti:

- (1) $(A\mathbb{R}_+^n \leq b) \neq \emptyset$.
- (2) $(\mathbb{R}_m^+ A \geq 0)b \geq 0$.

Dimostrazione. (1) \implies (2): Siano $x \in \mathbb{R}_+^n$ ed $f \in \mathbb{R}_m^+$ tali che $Ax \leq b$ ed $fA \geq 0$. Allora $fb \geq fAx \geq 0$.

(2) \implies (1): Supponiamo che $(A\mathbb{R}_+^n \leq b) = \emptyset$. Per l'oss. 1.27 ciò significa che $b \notin (A \ \delta)\mathbb{R}_+^n$. Per il teorema 1.26 possiamo quindi trovare un $f \in \mathbb{R}_m$ tale che $f(A \ \delta) \geq 0$ ed $fb < 0$. Ma $f(A \ \delta) = (fA \ f) \geq 0$ è equivalente alle due condizioni $fA \geq 0$ ed $f \geq 0$, cioè ad $f \in (\mathbb{R}_m^+ A \geq 0)$. Adesso però $fb < 0$ è in contrasto con la nostra ipotesi.

2. Programmi lineari in forma standard

Situazione 2.1. $n, m \in \mathbb{N} + 1$ ed $A \in \mathbb{R}_n^m$, $f \in \mathbb{R}_n$, $b \in \mathbb{R}^m$, quando non indicato diversamente.

Definizione 2.2. Un *programma lineare di massimizzazione in forma standard* è un problema di ottimizzazione della forma

$$\begin{aligned}fx &= \max \\x &\geq 0 \\Ax &\leq b\end{aligned}$$

con A, f, b come nella situazione 2.1.

Si cercano quindi vettori non negativi $x \in \mathbb{R}^n$ per i quali fx è massimale sotto il vincolo $Ax \leq b$.

Denotiamo con $(f(A\mathbb{R}_+^n \leq b) = \max)$ l'insieme degli $x \in \mathbb{R}_+^n$ (detti *soluzioni* del problema) con $Ax \leq b$ in cui fx assume un massimo su $(A\mathbb{R}_+^n \leq b)$, mentre indichiamo con $f(A\mathbb{R}_+^n \leq b) = \max$ (senza parentesi esterne) il problema stesso.

Gli elementi di $(A\mathbb{R}_+^n \leq b)$ si chiamano *punti ammissibili* del problema.

Definizione 2.3. Un *programma lineare di minimizzazione in forma standard* è un problema di ottimizzazione della forma

$$\begin{aligned}ub &= \min \\u &\geq 0 \\uA &\geq f\end{aligned}$$

con A, f, b come nella situazione 2.1.

Si cercano quindi vettori riga non negativi $u \in \mathbb{R}_m$ per i quali ub è minimale sotto il vincolo $uA \geq f$.

Denotiamo con $((\mathbb{R}_m^+ A \geq f)b = \min)$ l'insieme degli $u \in \mathbb{R}_m^+$ (detti *soluzioni* del problema) con $uA \geq f$ in cui ub assume un minimo su $(\mathbb{R}_m^+ A \geq f)$, mentre indichiamo con $(\mathbb{R}_m^+ A \geq f)b = \min$ (senza parentesi esterne) il problema stesso.

Gli elementi di $(\mathbb{R}_m^+ A \geq f)$ si chiamano *punti ammissibili* del problema duale.

Osservazione 2.4. I problemi $f(A\mathbb{R}_+^n \leq b) = \max$ e $(\mathbb{R}_m^+ A \geq f)b = \min$ si determinano a vicenda e si dicono uno il *duale* dell'altro.

Lemma 2.5. Siano x una soluzione del problema $f(A\mathbb{R}_+^n \leq b) = \max$ ed u una soluzione del problema $(\mathbb{R}_m^+ A \geq f)b = \min$. Allora $fx \leq uAx \leq ub$.

Dimostrazione. Per ipotesi $Ax \leq b$ e $f \leq uA$ con $x \geq 0$ e $f \geq 0$, per cui $fx \leq uAx \leq ub$.

Proposizione 2.6. Siano $x \in (A\mathbb{R}_+^n \leq b)$ ed $u \in (\mathbb{R}_m^+ A \geq f)$ tali che $fx \geq ub$.

Allora x è una soluzione di $f(A\mathbb{R}_+^n \leq b) = \max$ ed u una soluzione di $(\mathbb{R}_m^+ A \geq f)b = \min$.

Dimostrazione. Siano $y \in (A\mathbb{R}_+^n \leq b)$ e $v \in (\mathbb{R}_m^+ A \geq f)$. Per il lemma 2.5 abbiamo $fy \leq vb \leq fx \leq vb$.

Proposizione 2.7. (1) Sia $(A\mathbb{R}_+^n \leq b) = \emptyset$. Allora o $(\mathbb{R}_m^+ A \geq f) = \emptyset$ oppure l'insieme $(\mathbb{R}_m^+ A \geq f)b$ non è limitato inferiormente. Il problema $(\mathbb{R}_m^+ A \geq f)b = \min$ quindi non possiede soluzione.

(2) Sia $(\mathbb{R}_m^+ A \geq f) = \emptyset$. Allora o $(A\mathbb{R}_+^n \leq b) = \emptyset$ oppure l'insieme $f(A\mathbb{R}_+^n \leq b)$ non è limitato superiormente. Il problema $f(A\mathbb{R}_+^n \leq b) = \max$ quindi non possiede soluzione.

Dimostrazione. Dimostriamo solo il primo enunciato essendo il secondo il suo duale. Sia $(A\mathbb{R}_+^n \leq b) = \emptyset$. Per la prop. 1.28 esiste quindi un elemento $g \in \mathbb{R}_m^+$ tale che $gA \geq 0$ e $gb < 0$.

Se $(\mathbb{R}_m^+ A \geq f) = \emptyset$ è chiaro che il problema $(\mathbb{R}_m^+ A \geq f)b = \min$ non possiede soluzione.

Assumiamo quindi che $(\mathbb{R}_m^+ A \geq f) \neq \emptyset$. Sia ad esempio $u \in \mathbb{R}_m^+$ tale che $uA \geq f$. Allora per ogni $\lambda \geq 0$ si ha che $u + \lambda g \geq 0$ e $(u + \lambda g)A = uA + \lambda gA \geq f$, per cui $u + \lambda g \in (\mathbb{R}_m^+ A \geq f)$. Siccome $gb < 0$, si ha che $\lim_{\lambda \rightarrow \infty} (u + \lambda g)b = \lim_{\lambda \rightarrow \infty} (ub + \lambda gb) = -\infty$.

Osservazione 2.8. Se il problema $f(A\mathbb{R}_+^n \leq b) = \max$ possiede una soluzione x , allora il valore $\max f(A\mathbb{R}_+^n \leq b) = fx$ è ben definito.

Similmente, se il problema $(\mathbb{R}_m^+ A \geq f)b = \min$ possiede una soluzione u , allora il valore $\min(\mathbb{R}_m^+ A \geq f)b = ub$ è ben definito.

Teorema 2.9. Gli insiemi $(A\mathbb{R}_+^n \leq b)$ e $(\mathbb{R}_m^+ A \geq f)$ siano entrambi non vuoti. Allora i problemi $f(A\mathbb{R}_+^n \leq b) = \max$ e $(\mathbb{R}_m^+ A \geq f)b = \min$ sono entrambi risolvibili e si ha

$$\max f(A\mathbb{R}_+^n \leq b) = \min(\mathbb{R}_m^+ A \geq f)b$$

Dimostrazione. (1) È sufficiente dimostrare che esistono $x \in (A\mathbb{R}_+^n \leq b)$ ed $u \in (\mathbb{R}_m^+ A \geq f)$ tali che $fx \geq ub$. L'enunciato allora si ottiene applicando la prop. 2.6.

(2) Il nostro scopo quindi è quello di trovare un elemento $x \in \mathbb{R}_+^n$ ed un elemento $u \in \mathbb{R}_m^+$ tali che $Ax \leq b$, $uA \geq f$ (cioè $-A^t u^t \leq -f^t$) ed $fx \geq ub$. L'ultima disuguaglianza è equivalente a $-fx + b^t u \leq 0$.

(3) Scrivendo il problema in forma matriciale, ponendo

$$B := \begin{pmatrix} A & 0 \\ 0 & -A^t \\ -f & b^t \end{pmatrix} \in \mathbb{R}_{m+n}^{m+n+1} \quad \text{e} \quad c := \begin{pmatrix} b \\ -f^t \\ 0 \end{pmatrix} \in \mathbb{R}^{m+n+1}$$

dobbiamo dimostrare che $(B\mathbb{R}_+^{m+n} \leq c) \neq \emptyset$.

(4) Ma per la prop.1.28 in caso contrario esiste $p \in \mathbb{R}_{m+n+1}^+$ tale che $pB \geq 0$ e $pc < 0$.

p è quindi della forma $p = (g \ h \ \alpha)$ con $g \in \mathbb{R}_m^+$, $h \in \mathbb{R}_n^+$ ed $\alpha \in \mathbb{R}^+$. Inoltre

$$pB = (g \ h \ \alpha) \begin{pmatrix} A & 0 \\ 0 & -A^t \\ -f & b^t \end{pmatrix} = (gA - \alpha f \quad -hA^t + \alpha b^t) \geq 0$$

e

$$pc = (g \ h \ \alpha) \begin{pmatrix} b \\ -f^t \\ 0 \end{pmatrix} = gb - hf^t < 0$$

cosicché otteniamo

$$(I) \quad gA \geq \alpha f$$

$$(II) \quad \alpha b^t \geq hA^t$$

$$(III) \quad gb < hf^t$$

Essendo $\alpha \in \mathbb{R}^+$, abbiamo $\alpha = 0$ oppure $\alpha > 0$. Dimostriamo che entrambi i casi implicano una contraddizione.

(5) Sia $\alpha = 0$. Allora $gA \geq 0 \geq hA^t$. Per ipotesi gli insiemi $(A\mathbb{R}_+^n \leq b)$ e $(\mathbb{R}_m^+ A \geq f)$ sono entrambi non vuoti. Perciò esistono $x \in \mathbb{R}_+^n$ ed $u \in \mathbb{R}_m^+$ tali che $Ax \leq b$ ed $uA \geq f$, quindi anche $A^t u^t \geq f^t$. Ma allora

$$gb \geq gAx \geq 0 \geq hA^t u^t \geq hf^t$$

e questo contraddice la disuguaglianza (III).

(6) Rimane il caso $\alpha > 0$. Le disuguaglianze (I) e (II) significano però che $\frac{g}{\alpha} \in (\mathbb{R}_m^+ A \geq f)$ e $\frac{h^t}{\alpha} \in (A\mathbb{R}_+^n \leq b)$.

Dal lemma 2.5 segue $f \frac{h^t}{\alpha} \leq \frac{g}{\alpha} b$ ovvero $hf^t = fh^t \leq gb$, ancora in contrasto con (III).

Corollario 2.10. Siano $x \in (A\mathbb{R}_+^n \leq b)$ ed $u \in (\mathbb{R}_m^+ A \geq f)$. Allora sono equivalenti:

(1) x è una soluzione di $f(A\mathbb{R}_+^n \leq b) = \max$ ed u una soluzione di $(\mathbb{R}_m^+ A \geq f)b = \min$.

(2) $fx = ub$.

(3) $fx = uAx = ub$.

(4) $fx \geq ub$.

Dimostrazione. (1) \implies (2): Teorema 2.9.

(2) \implies (3): Per il lemma 2.5 si ha $fx \leq uAx \leq ub$, ma essendo per ipotesi $fx = ub$ segue che $fx = uAx = ub$.

(3) \implies (4): Chiaro.

(4) \implies (1): Proposizione 2.6.

Nota 2.11. Come nella dimostrazione del teorema 2.9 siano

$$B := \begin{pmatrix} A & 0 \\ 0 & -A^t \\ -f & b^t \end{pmatrix} \in \mathbb{R}_{m+n}^{m+n+1} \quad \text{e} \quad c := \begin{pmatrix} b \\ -f^t \\ 0 \end{pmatrix} \in \mathbb{R}^{m+n+1}.$$

(1) Sia $z \in (B\mathbb{R}_+^{m+n} \leq c)$ con $z = \begin{pmatrix} x \\ y \end{pmatrix}$, $x \in \mathbb{R}_+^n$ ed $y \in \mathbb{R}_+^m$.

Se poniamo $u := y^t$, allora x è una soluzione di $f(A\mathbb{R}_+^n \leq b) = \max$ ed u è una soluzione di $(\mathbb{R}_m^+ A \geq f)b = \min$.

(2) Siano viceversa x una soluzione di $f(A\mathbb{R}_+^n \leq b) = \max$ ed u una soluzione di $(\mathbb{R}_m^+ A \geq f)b = \min$. Allora con $y := u^t$ abbiamo

$$z := \begin{pmatrix} x \\ y \end{pmatrix} \in (B\mathbb{R}_+^{m+n} \leq c).$$

Dimostrazione. Siano z, x, y, u come nell'enunciato. Abbiamo quindi le seguenti disuguaglianze:

$$\begin{aligned} x &\geq 0, y \geq 0 \\ Ax &\leq b \\ -A^t u^t &\leq -f^t \\ -fx + b^t u^t &\leq 0 \end{aligned}$$

La terza disuguaglianza è equivalente ad $uA \geq f$, la quarta, essendo $b^t u^t = ub$, è equivalente a $fx \geq ub$.

Il punto (1) segue adesso dalla prop. 2.6, il punto (2) combinando il teorema 2.9 con la prop. 2.6.

Osservazione 2.12. Il teorema 2.9 e la nota 2.11 dimostrano l'importanza e l'utilità del principio di dualità nella programmazione lineare. Vediamo in particolare che la sola ricerca di un punto ammissibile (senza un compito di massimizzazione o minimizzazione) non è più facile della soluzione del problema di ottimizzazione che, come visto nella nota 2.11, può essere ricondotta alla ricerca di un punto ammissibile.

Esempio 2.13. Consideriamo il seguente problema di massimo:

$$\begin{aligned} x^1 + 2x^2 &= \max \\ 7x^1 + 4x^2 &\leq 28 \\ 4x^1 + 5x^2 &\leq 20 \\ 3x^1 + 10x^2 &\leq 30 \end{aligned}$$

Per la nota 2.11 questo problema di ottimizzazione è equivalente a quello della ricerca di un punto

$$z = \begin{pmatrix} x^1 \\ x^2 \\ y^1 \\ y^2 \\ y^3 \end{pmatrix}$$

che soddisfa la disuguaglianza

$$\begin{pmatrix} 7 & 4 & 0 & 0 & 0 \\ 4 & 5 & 0 & 0 & 0 \\ 3 & 10 & 0 & 0 & 0 \\ 0 & 0 & -7 & -4 & -3 \\ 0 & 0 & -4 & -5 & -3 \\ -1 & -2 & 28 & 20 & 30 \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \\ y^1 \\ y^2 \\ y^3 \end{pmatrix} \leq \begin{pmatrix} 28 \\ 20 \\ 30 \\ -1 \\ -2 \\ 0 \end{pmatrix}$$

Corollario 2.14. (1) x sia una soluzione di $f(A\mathbb{R}_+^n \leq b) = \max$ e $1 \leq k \leq m$. Se esiste una soluzione u di $(\mathbb{R}_m^+ A \geq f)b = \min$ tale che $u_k \neq 0$, allora $A^k x = b^k$.

(2) u sia una soluzione di $(\mathbb{R}_m^+ A \geq f)b = \min$ e $1 \leq k \leq n$. Se esiste una soluzione x di $f(A\mathbb{R}_+^n \leq b) = \max$ tale che $x^k \neq 0$, allora $uA_k = f_k$.

Dimostrazione. Dimostriamo solo il primo enunciato, essendo il secondo il suo duale.

Per il corollario 2.10 $uAx = ub$, e quindi $u(b - Ax) = 0$, cioè $\sum_{j=1}^m u_j(b^j - A^j x) = 0$. Per ipotesi però $u \geq 0$ e $b - Ax \geq 0$, quindi $u_j(b^j - A^j x) \geq 0$ per ogni $j = 1, \dots, m$. Ma allora necessariamente $u_j(b^j - A^j x) = 0$ per ogni $j = 1, \dots, m$, per cui l'ipotesi $u_k \neq 0$ implica $b^k - A^k x = 0$.

3. Programmi lineari in forma normale

Situazione 3.1. $n, m \in \mathbb{N} + 1$ ed $A \in \mathbb{R}_n^m$, $f \in \mathbb{R}_n$, $b \in \mathbb{R}^m$, quando non indicato diversamente.

Definizione 3.2. Un *programma lineare di massimizzazione* (risp. *minimizzazione*) in forma normale è un problema di ottimizzazione della forma

$$fx = \max \text{ (risp. min)}$$

$$x \geq 0$$

$$Ax = b$$

con A, f, b come nella situazione 3.1. Gli elementi di $(A\mathbb{R}_+^n = b)$ sono detti *punti ammissibili*.

Nota 3.3. (1) Sia dato un problema $f(A\mathbb{R}_+^n \leq b) = \max$ in forma standard. Introducendo un vettore variabile ausiliario p otteniamo un problema in forma normale

$$(f \ 0) \begin{pmatrix} x \\ p \end{pmatrix} = \max$$

$$\begin{pmatrix} x \\ p \end{pmatrix} \geq 0$$

$$(A \ \delta) \begin{pmatrix} x \\ p \end{pmatrix} = b$$

equivalente al primo, nel senso che da ogni soluzione del problema in forma normale otteniamo una soluzione del problema in forma standard.

(2) Viceversa, dato un problema $f(A\mathbb{R}_+^n = b) = \max$ in forma normale, considerando

$$fx = \max$$

$$x \geq 0$$

$$Ax \leq b$$

$$-Ax \leq -b$$

otteniamo un problema in forma standard equivalente al primo, nel senso che da ogni soluzione del problema in forma standard otteniamo una soluzione del problema in forma normale.

(3) In modo analogo (usando le matrici trasposte) ogni problema di minimizzazione in forma standard (def. 2.3) è equivalente ad un problema in forma normale e viceversa.

Osservazione 3.4. V e W siano spazi vettoriali reali, $\varphi : V \rightarrow W$ un'applicazione affine ed Y sia un sottoinsieme convesso di W .

Allora $\varphi^{-1}(Y)$ è convesso.

Dimostrazione. Siano $a, b \in \varphi^{-1}(Y)$, cioè $\varphi(a), \varphi(b) \in Y$. Per l'oss. 1.8 però $\varphi([a, b]) = [\varphi(a), \varphi(b)] \subset Y$ e quindi, utilizzando l'ipotesi che Y sia convesso, si ha che $[a, b] \subset \varphi^{-1}(Y)$.

Corollario 3.5. *L'insieme $(A\mathbb{R}_+^n = b)$ è chiuso e convesso.*

Dimostrazione. Identificando A con l'applicazione $\bigcirc_x Ax$ possiamo scrivere $(A\mathbb{R}_+^n = b) = A^{-1}(b) \cap \mathbb{R}_+^n$. Poiché A è continua, l'insieme $A^{-1}(b)$ è chiuso. Inoltre, essendo A un'applicazione affine e b un convesso, per l'oss. 3.4 anche $A^{-1}(b)$ è convesso. Infine \mathbb{R}_+^n è un chiuso convesso e, siccome l'intersezione di insiemi chiusi e convessi è ancora chiuso e convesso, otteniamo l'enunciato.

Definizione 3.6. V sia uno spazio vettoriale reale ed X un suo sottoinsieme convesso. Un *vertice* (o *punto estrema*) di X è un punto $x \in X$ che soddisfa la seguente condizione:

Se $u, v \in X$ ed $x \in [u, v]$, allora $x \in \{u, v\}$

Denotiamo con $\text{vertici}(X)$ l'insieme dei vertici di X .

Esempio 3.7.

- (1) X sia un poligono convesso in \mathbb{R}^2 . I vertici di X nel senso della geometria elementare coincidono con i vertici nel senso della definizione 3.6.
- (2) $\text{vertici}(\mathbb{R}^n) = \emptyset$.
- (3) $\text{vertici}(\mathbb{R}_+^n) = \{0\}$.
- (4) $X := \{z \in \mathbb{R}^n \mid |z| \leq 1\}$ sia la palla unitaria chiusa in \mathbb{R}^n . Allora $\text{vertici}(X) = \{z \in \mathbb{R}^n \mid |z| = 1\}$.
- (5) $X := \{z \in \mathbb{R}^n \mid |z| < 1\}$ sia la palla unitaria aperta in \mathbb{R}^n . Allora $\text{vertici}(X) = \emptyset$.

Lemma 3.8. V sia uno spazio vettoriale reale, X un sottoinsieme convesso di V ed $x \in X$. Allora sono equivalenti:

- (1) $x \in \text{vertici}(X)$.
- (2) $X \setminus \{x\}$ è convesso.
- (3) $x = \frac{u+v}{2}$ con $u, v \in X$ implica $u = v$.

Dimostrazione. (1) \implies (2): Supponiamo per assurdo che esistano $u, v \in X \setminus \{x\}$ tali che $[u, v] \not\subset X \setminus \{x\}$. Allora $x \in [u, v]$, ma per ipotesi ciò implica che $x \in \{u, v\}$. Quindi ad esempio $u = x$, una contraddizione.

(2) \implies (1): $X \setminus \{x\}$ sia convesso ed $u, v \in X$ con $x \in [u, v]$. Dobbiamo dimostrare che $x \in \{u, v\}$. Se così non fosse avremmo $u, v \in X \setminus \{x\}$ e quindi per ipotesi $[u, v] \subset X \setminus \{x\}$, una contraddizione.

(1) \implies (3): Sia $2x = u + v$ con $u, v \in X$. Per ipotesi $x \in \{u, v\}$, ad esempio $x = u$. Allora $x = v$ e quindi $u = v$.

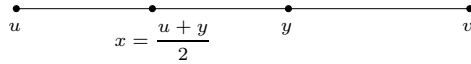
(3) \implies (1): Siano $u, v \in X$ ed $x = u + t(v - u)$ con $t \in [0, 1]$.

Per $t = 0$ e $t = 1$ si ha rispettivamente $x = u$ e $x = v$.

Per $t = \frac{1}{2}$ l'ipotesi implica $u = v$ e quindi $x = u = v$.

Rimangono i casi $0 < t < \frac{1}{2}$ e $\frac{1}{2} < t < 1$. Per simmetria è sufficiente trattare il caso $0 < t < \frac{1}{2}$.

Poniamo $y := u + 2t(v - u)$. Allora $y \in X$ perché $0 < 2t < 1$. Inoltre $y = u + 2(x - u) = 2x - u$, cosicché $x = \frac{u + y}{2}$. L'ipotesi implica $u = y$ e quindi $x = u$.



Definizione 3.9. Sia $\alpha \subset \{1, \dots, n\}$ un sottoinsieme di cardinalità $|\alpha| = s$. Poniamo allora:

- (1) $A_\alpha :=$ matrice in \mathbb{R}_s^m che si ottiene da A cancellando tutte le colonne A_j per cui $j \notin \alpha$.
- (2) $f_\alpha :=$ vettore in \mathbb{R}_s che si ottiene da f cancellando gli elementi f_j per cui $j \notin \alpha$.
- (3) Per $x \in \mathbb{R}^n$ sia x^α il vettore in \mathbb{R}^s che si ottiene da x cancellando gli elementi x^j per cui $j \notin \alpha$.

Poniamo inoltre $1 - \alpha := \{1, \dots, n\} \setminus \alpha$.

Per $s = 0$ abbiamo $\mathbb{R}_0^m = \mathbb{R}^0 = \mathbb{R}_0 = 0$, per cui $A_\emptyset = f_\emptyset = x^\emptyset = \emptyset$.

Con queste convenzioni possiamo scrivere

$$Ax = A_\alpha x^\alpha + A_{1-\alpha} x^{1-\alpha}$$

$$fx = f_\alpha x^\alpha + f_{1-\alpha} x^{1-\alpha}$$

Esempio 3.10. Siano $A \in \mathbb{R}_5^3$ ed $x \in \mathbb{R}^5$. Con $\alpha = \{1, 4\}$ si ha $1 - \alpha = \{2, 3, 5\}$. Perciò

$$\begin{aligned} A_\alpha x^\alpha + A_{1-\alpha} x^{1-\alpha} &= \begin{pmatrix} A_1^1 & A_4^1 \\ A_1^2 & A_4^2 \\ A_1^3 & A_4^3 \end{pmatrix} \begin{pmatrix} x^1 \\ x^4 \end{pmatrix} + \begin{pmatrix} A_2^1 & A_3^1 & A_5^1 \\ A_2^2 & A_3^2 & A_5^2 \\ A_2^3 & A_3^3 & A_5^3 \end{pmatrix} \begin{pmatrix} x^2 \\ x^3 \\ x^5 \end{pmatrix} \\ &= \begin{pmatrix} A_1^1 x^1 + A_4^1 x^4 \\ A_1^2 x^1 + A_4^2 x^4 \\ A_1^3 x^1 + A_4^3 x^4 \end{pmatrix} + \begin{pmatrix} A_2^1 x^2 + A_3^1 x^3 + A_5^1 x^5 \\ A_2^2 x^2 + A_3^2 x^3 + A_5^2 x^5 \\ A_2^3 x^2 + A_3^3 x^3 + A_5^3 x^5 \end{pmatrix} \\ &= Ax \end{aligned}$$

Definizione 3.11. Per $x \in \mathbb{R}^n$ sia $[x] := \{i \in \{1, \dots, n\} \mid x^i \neq 0\}$.

Osservazione 3.12. $0 \in \text{vertici}(A\mathbb{R}_+^n = 0)$.

Dimostrazione. Usiamo il lemma 3.8. Siano $u, v \in (A\mathbb{R}_+^n = 0)$ e tali che $0 = \frac{u+v}{2}$. È chiaro che ciò implica $u = v = 0$.

Teorema 3.13. Per $x \in (\mathbb{A}\mathbb{R}_+^n = b)$ sono equivalenti:

- (1) $x \in \text{vertici}(\mathbb{A}\mathbb{R}_+^n = b)$.
- (2) Le colonne di $A_{[x]}$ sono linearmente indipendenti.

Dimostrazione. Se $x = 0$, allora $[x] = \emptyset$ e l'enunciato segue dall'oss. 3.9, perché l'insieme vuoto è linearmente indipendente. Possiamo quindi assumere che $x \neq 0$ e che $[x] = \{1, \dots, s\}$ con $1 \leq s \leq n$.

(1) \implies (2): Sia $x \in \text{vertici}(\mathbb{A}\mathbb{R}_+^n = b)$ e sia $\lambda_1 A_1 + \dots + \lambda_s A_s = 0$ una combinazione lineare delle colonne A_i con $i \in [x]$ e $\lambda_i \in \mathbb{R}$ per $i = 1, \dots, s$. Supponiamo per assurdo che i λ_i non siano tutti nulli, ad esempio possiamo supporre $\lambda_1 \neq 0$. Siccome $x^i > 0$ per ogni $i = 1, \dots, s$, possiamo trovare un elemento $\varepsilon > 0$ tale che $x^i \pm \varepsilon \lambda_i \geq 0$ per ogni $i = 1, \dots, s$. Definiamo i vettori

$$u := \begin{pmatrix} x^1 + \varepsilon \lambda_1 \\ \vdots \\ x^s + \varepsilon \lambda_s \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \text{e} \quad v := \begin{pmatrix} x^1 - \varepsilon \lambda_1 \\ \vdots \\ x^s - \varepsilon \lambda_s \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Allora $u, v \in \mathbb{R}_+^n$ ed $x = \frac{u+v}{2}$.

Inoltre $Au = \sum_{k=1}^s A_k(x^k + \varepsilon \lambda_k) = Ax + \varepsilon \sum_{k=1}^s \lambda_k A_k = Ax = b$, e similmente $Av = b$. Perciò $u, v \in (\mathbb{A}\mathbb{R}_+^n = b)$. Per ipotesi $x \in \text{vertici}(\mathbb{A}\mathbb{R}_+^n = b)$ e quindi per il punto (3) della prop. 3.8 risulta $u = v$. Ma ciò non è possibile perchè $\lambda_1 \neq 0$.

(2) \implies (1): Le colonne di $A_{[x]}$ siano linearmente indipendenti. Siano $u, v \in (\mathbb{A}\mathbb{R}_+^n = b)$ tali che $x = \frac{u+v}{2}$. Ciò implica $u^i + v^i = 0$ per ogni $i = 1, \dots, s$, e quindi $u^i = v^i = 0$, essendo $u, v \geq 0$. D'altra parte abbiamo $Au = Av = b$ e quindi $0 = A(u-v) = \sum_{i=1}^s (u^i - v^i)A_i$. Per ipotesi ciò implica $u^i = v^i$ per ogni $i = 1, \dots, s$, e quindi $u = v$. Per la prop. 3.8 quindi risulta che $x \in \text{vertici}(\mathbb{A}\mathbb{R}_+^n = b)$.

Definizione 3.14. Sia $\alpha \subset \{1, \dots, n\}$. A si dice α -invertibile, se $|\alpha| = m \leq n$ e la matrice quadratica A_α è invertibile.

Osservazione 3.15. Un insieme di indici α tale che A sia α -invertibile esiste se e solo se $\text{rango}(A) = m$. Si noti che $\text{rango}(A) = m$ da solo implica $m \leq n$.

Lemma 3.16. K sia un campo e V uno spazio vettoriale su K . Siano $v_1, \dots, v_n \in V$ e $W := \text{SV}(v_1, \dots, v_n)$. Siano $m := \dim W$ e $1 \leq s \leq m$ tali che i vettori v_1, \dots, v_s siano linearmente indipendenti.

Allora possiamo trovare $m - s$ indici $j_1, \dots, j_{m-s} \in \{s + 1, \dots, n\}$ in modo tale che gli m vettori $v_1, \dots, v_s, v_{j_1}, \dots, v_{j_{m-s}}$ siano linearmente indipendenti.

Dimostrazione. (1) Se $s = m$ non c'è nulla da dimostrare.

Supponiamo quindi che $s < m$ e dimostriamo che esiste un indice $j \in \{s + 1, \dots, n\}$ tale che gli $s + 1$ vettori v_1, \dots, v_s, v_j sono linearmente indipendenti.

Infatti, se così non fosse, avremmo $v_j \in \text{SV}(v_1, \dots, v_s)$ per ogni $j \in \{s + 1, \dots, n\}$ e quindi $W = \text{SV}(v_1, \dots, v_s)$ in contraddizione all'ipotesi $\dim W = m$.

(2) Se adesso $s + 1 = m$, non c'è più nulla da dimostrare. Altrimenti ripetiamo il ragionamento con $s + 1$ al posto di s .

Corollario 3.17. x sia un vertice di $(A\mathbb{R}_+^n = b)$.

Se $\text{rango}(A) = m$, allora $[x]$ è contenuto in un insieme di indici α tali che A sia α -invertibile.

Osservazione 3.18. Sia $x \in (A\mathbb{R}_+^n = b)$ ed $[x]$ contenuto in un insieme di indici α tali che A sia α -invertibile. Allora abbiamo $x^{1-\alpha} = 0$ e quindi $b = Ax = A_\alpha x^\alpha + A_{1-\alpha} x^{1-\alpha} = A_\alpha x^\alpha$, per cui $x^\alpha = A_\alpha^{-1}b$. Perciò x è univocamente determinato dalle relazioni

$$\begin{aligned} x^\alpha &= A_\alpha^{-1}b \\ x^{1-\alpha} &= 0 \end{aligned}$$

Definizione 3.19. Sia $\text{rango}(A) = m$. Poniamo

$$\begin{aligned} J(A) &:= \{\alpha \subset \{1, \dots, n\} \mid |\alpha| = m \text{ ed } A_\alpha \text{ invertibile}\} \\ J^+(A, b) &:= \{\alpha \in J(A) \mid A_\alpha^{-1}b \geq 0\} \end{aligned}$$

Proposizione 3.20. Sia $\text{rango}(A) = m$. Allora l'applicazione $\theta : J^+(A, b) \rightarrow \text{vertici}(A\mathbb{R}_+^n = b)$ definita da $\theta(\alpha) = x$ con

$$\begin{aligned} x^\alpha &= A_\alpha^{-1}b \\ x^{1-\alpha} &= 0 \end{aligned}$$

è suriettiva.

Dimostrazione. Ciò segue dall'oss. 3.18 e dal cor. 3.17.

Corollario 3.21. Sia $\text{rango}(A) = m$. Allora l'insieme dei vertici di $(A\mathbb{R}_+^n = b)$ è finito.

Dimostrazione. $J^+(A, b)$ è finito e quindi, poichè l'immagine di un insieme finito è finito, anche $\text{vertici}(A\mathbb{R}_+^n = b)$ è un insieme finito.

Proposizione 3.22. Sia $(A\mathbb{R}_+^n = b) \neq \emptyset$. Allora $\text{vertici}(A\mathbb{R}_+^n = b) \neq \emptyset$.

Dimostrazione. Scegliamo $x \in (A\mathbb{R}_+^n = b)$ in modo tale che $||[x]||$ sia minimale.

(1) Se $x = 0$, allora $b = 0$ ed $x \in \text{vertici}(A\mathbb{R}_+^n = b)$ per l'oss. 3.12.

(2) Sia quindi $x \neq 0$. Poniamo $\alpha := [x]$. Dobbiamo dimostrare che le colonne di A_α sono linearmente indipendenti. Supponiamo che non lo siano. Allora possiamo trovare un vettore $\lambda \in \mathbb{R}^s \setminus 0$ tale che $A_\alpha \lambda = 0$. Possiamo assumere che almeno uno dei coefficienti di λ sia minore di 0. Siccome $x^i > 0$ per ogni $i \in \alpha$, esiste $\varepsilon > 0$ tale che $x^i + \varepsilon \lambda^i \geq 0$ per ogni $i \in \alpha$ ed $x^h + \varepsilon \lambda^h = 0$ per almeno un $h \in \alpha$. Sia ora $y \in \mathbb{R}^{n+1}$ definito dalle relazioni

$$\begin{aligned} y^\alpha &= x^\alpha + \varepsilon \lambda \\ y^{1-\alpha} &= 0 \end{aligned}$$

Allora $y \geq 0$, inoltre

$$\begin{aligned} Ay &= A_\alpha y^\alpha + A_{1-\alpha} y^{1-\alpha} = A_\alpha y^\alpha = A_\alpha (x^\alpha + \varepsilon \lambda) \\ &= A_\alpha x^\alpha + \varepsilon A_\alpha \lambda = A_\alpha x^\alpha = Ax = b \end{aligned}$$

Quindi $y \in (A\mathbb{R}_+^n = b)$ e $\|y\| < \|x\|$, una contraddizione al fatto che abbiamo scelto x in modo tale che $\|x\|$ sia minimale.

Osservazione 3.23. Il problema $f(A\mathbb{R}_+^n = b) = \max$ abbia una soluzione. Se poniamo $\mu := \max f(A\mathbb{R}_+^n = b)$, allora con

$$B := \begin{pmatrix} A \\ f \end{pmatrix}, \quad c := \begin{pmatrix} b \\ \mu \end{pmatrix}$$

si ha $(f(A\mathbb{R}_+^n = b) = \max) = (B\mathbb{R}_+^n = c)$.

Per il cor. 3.5 l'insieme $(f(A\mathbb{R}_+^n = b) = \max)$ è perciò chiuso e convesso (ciò è banalmente vero anche quando è vuoto).

Proposizione 3.24.

$$\text{vertici}(f(A\mathbb{R}_+^n = b) = \max) = (f(A\mathbb{R}_+^n = b) = \max) \cap \text{vertici}(A\mathbb{R}_+^n = b)$$

Dimostrazione. (1) Se il problema $f(A\mathbb{R}_+^n = b) = \max$ non ha soluzione, l'enunciato è banalmente vero.

(2) Altrimenti con $\mu := \max f(A\mathbb{R}_+^n = b)$ poniamo $P := (A\mathbb{R}_+^n = b)$, $R := (f\mathbb{R}_+^n = \mu)$. Allora $(f(A\mathbb{R}_+^n = b) = \max) = P \cap R$, per cui dobbiamo dimostrare che $\text{vertici}(P \cap R) = P \cap R \cap \text{vertici}(P)$. Adesso usiamo il lemma 3.8.

Sia $x \in \text{vertici}(P \cap R)$. Dobbiamo dimostrare che $x \in \text{vertici}(P)$. Siano $u, v \in P$ tali che $x = \frac{u+v}{2}$. Però $fu \leq \mu$, $fv \leq \mu$, e se ad esempio $fu < \mu$, si avrebbe $fx = \frac{fu+fv}{2} < \mu$. Perciò $fu = fv = \mu$, cosicché $u, v \in P \cap R$, e quindi $u = v$, perché $x \in \text{vertici}(P \cap R)$.

Viceversa, sia $x \in P \cap R \cap \text{vertici}(P)$. Siano $u, v \in P \cap R$ tali che $x = \frac{u+v}{2}$. Allora $u = v$ perché $x \in \text{vertici}(P)$. Quindi $x \in \text{vertici}(P \cap R)$.

Teorema 3.25 (teorema fondamentale della programmazione lineare). Se il problema $f(A\mathbb{R}_+^n = b) = \max$ possiede una soluzione, allora esiste anche una soluzione che è un vertice di $(A\mathbb{R}_+^n = b)$.

Dimostrazione. Con le notazioni dell'oss. 3.23, dalla prop. 3.24 abbiamo $(f(A\mathbb{R}_+^n = b) = \max) \cap \text{vertici}(A\mathbb{R}_+^n = b) = \text{vertici}(B\mathbb{R}_+^n = c)$.

Siccome per ipotesi l'insieme $(B\mathbb{R}_+^n = c) = (f(A\mathbb{R}_+^n = b) = \max)$ non è vuoto, dalla prop. 3.22 segue che anche $\text{vertici}(B\mathbb{R}_+^n = c) \neq \emptyset$.

Nota 3.26. Mediante l'algoritmo di eliminazione di Gauss possiamo sempre ottenere la condizione $\text{rango}(A) = m$ richiesta nella prop. 3.20 e nel corollario 3.21. L'insieme dei vertici di $(A\mathbb{R}_+^n = b)$ allora è finito e per il teorema 3.25 una soluzione del problema di massimo si trova percorrendo tutti i vertici e calcolando $\max \{fx \mid x \in \text{vertici}(A\mathbb{R}_+^n = b)\}$. Il numero dei vertici è però molto alto già in problemi di modeste dimensioni, per cui è necessario un algoritmo migliore, il *metodo del simplesso*, che verrà presentato nel prossimo capitolo.

Esempio 3.27. Consideriamo il problema di massimo

$$3x^1 + x^2 - x^3 + 2x^4 - 2x^5 + x^6 = \max$$

$$3x^1 + 2x^2 - 5x^3 + 4x^4 - x^5 - x^6 = 18$$

$$x^1 - x^2 - x^3 + 4x^4 - 6x^5 + x^6 = 15$$

che possiamo scrivere nella solita forma

$$fx = \max$$

$$Ax = b$$

$$x \geq 0$$

con

$$f := (3 \quad 1 \quad -1 \quad 2 \quad -2 \quad 1)$$

$$A := \begin{pmatrix} 3 & 2 & -5 & 4 & -1 & -1 \\ 1 & -1 & -1 & 4 & -6 & 1 \end{pmatrix}$$

$$b := \begin{pmatrix} 18 \\ 15 \end{pmatrix}$$

Chiaramente $\text{rango}(A) = 2$. Per applicare l'idea della nota 3.26 dobbiamo considerare le $\binom{6}{2} = 15$ sottomatrici 2×2 A_α di A e calcolare $x^\alpha = A_\alpha^{-1}b$ ogni volta che A_α sia invertibile. Se $x^\alpha \geq 0$, calcoliamo $f_\alpha x^\alpha$. Il massimo dei valori così ottenuti è il massimo cercato. Alla pagina seguente riportiamo la tabella con i calcoli effettuati. Il valore massimo che si ottiene è 81.

α	A_α	A_α^{-1}	$x^\alpha = A_\alpha^{-1}b$	x	fx
{1,2}	$\begin{pmatrix} 3 & 2 \\ 1 & -1 \end{pmatrix}$	$-\frac{1}{5} \begin{pmatrix} -1 & -2 \\ -1 & 3 \end{pmatrix}$	$\begin{pmatrix} 48/5 \\ -27/5 \end{pmatrix}$	$\begin{pmatrix} 48/5 \\ -27/5 \end{pmatrix}$	
{1,3}	$\begin{pmatrix} 3 & -5 \\ 1 & -1 \end{pmatrix}$	$\frac{1}{2} \begin{pmatrix} -1 & 5 \\ -1 & 3 \end{pmatrix}$	$\begin{pmatrix} 57/2 \\ 27/2 \end{pmatrix}$	$\begin{pmatrix} 57/2 \\ 27/2 \end{pmatrix}$	72
{1,4}	$\begin{pmatrix} 3 & 4 \\ 1 & 4 \end{pmatrix}$	$\frac{1}{8} \begin{pmatrix} 4 & -4 \\ -1 & 3 \end{pmatrix}$	$\begin{pmatrix} 3/2 \\ 27/8 \end{pmatrix}$	$\begin{pmatrix} 3/2 \\ 27/8 \end{pmatrix}$	45/4
{1,5}	$\begin{pmatrix} 3 & -1 \\ 1 & -6 \end{pmatrix}$	$-\frac{1}{17} \begin{pmatrix} -6 & 1 \\ -1 & 3 \end{pmatrix}$	$\begin{pmatrix} 93/17 \\ -27/17 \end{pmatrix}$	$\begin{pmatrix} 93/17 \\ -27/17 \end{pmatrix}$	
{1,6}	$\begin{pmatrix} 3 & -1 \\ 1 & 1 \end{pmatrix}$	$\frac{1}{4} \begin{pmatrix} 1 & 1 \\ -1 & 3 \end{pmatrix}$	$\begin{pmatrix} 33/4 \\ 27/4 \end{pmatrix}$	$\begin{pmatrix} 33/4 \\ 27/4 \end{pmatrix}$	63/2
{2,3}	$\begin{pmatrix} 2 & -5 \\ -1 & -1 \end{pmatrix}$	$-\frac{1}{7} \begin{pmatrix} -1 & 5 \\ 1 & 2 \end{pmatrix}$	$\begin{pmatrix} -57/7 \\ -48/7 \end{pmatrix}$	$\begin{pmatrix} -57/7 \\ -48/7 \end{pmatrix}$	
{2,4}	$\begin{pmatrix} 2 & 4 \\ -1 & 4 \end{pmatrix}$	$\frac{1}{12} \begin{pmatrix} 4 & -4 \\ 1 & 2 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 4 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 4 \end{pmatrix}$	9
{2,5}	$\begin{pmatrix} 2 & -1 \\ -1 & -6 \end{pmatrix}$	$-\frac{1}{13} \begin{pmatrix} -6 & 1 \\ 1 & 2 \end{pmatrix}$	$\begin{pmatrix} 93/13 \\ -48/13 \end{pmatrix}$	$\begin{pmatrix} 93/13 \\ -48/13 \end{pmatrix}$	
{2,6}	$\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$	$\begin{pmatrix} 33 \\ 48 \end{pmatrix}$	$\begin{pmatrix} 33 \\ 48 \end{pmatrix}$	81
{3,4}	$\begin{pmatrix} -5 & 4 \\ -1 & 4 \end{pmatrix}$	$-\frac{1}{16} \begin{pmatrix} 4 & -4 \\ 1 & -5 \end{pmatrix}$	$\begin{pmatrix} -3/4 \\ 57/16 \end{pmatrix}$	$\begin{pmatrix} -3/4 \\ 57/16 \end{pmatrix}$	
{3,5}	$\begin{pmatrix} -5 & -1 \\ -1 & -6 \end{pmatrix}$	$\frac{1}{29} \begin{pmatrix} -6 & 1 \\ 1 & -5 \end{pmatrix}$	$\begin{pmatrix} -93/29 \\ -57/29 \end{pmatrix}$	$\begin{pmatrix} -93/29 \\ -57/29 \end{pmatrix}$	
{3,6}	$\begin{pmatrix} -5 & -1 \\ -1 & 1 \end{pmatrix}$	$-\frac{1}{6} \begin{pmatrix} 1 & 1 \\ 1 & -5 \end{pmatrix}$	$\begin{pmatrix} -11/2 \\ 57/6 \end{pmatrix}$	$\begin{pmatrix} -11/2 \\ 57/6 \end{pmatrix}$	
{4,5}	$\begin{pmatrix} 4 & -1 \\ 4 & -6 \end{pmatrix}$	$-\frac{1}{20} \begin{pmatrix} -6 & 1 \\ -4 & 4 \end{pmatrix}$	$\begin{pmatrix} 93/20 \\ 3/5 \end{pmatrix}$	$\begin{pmatrix} 93/20 \\ 3/5 \end{pmatrix}$	18
{4,6}	$\begin{pmatrix} 4 & -1 \\ 4 & 1 \end{pmatrix}$	$\frac{1}{8} \begin{pmatrix} 1 & 1 \\ -4 & 4 \end{pmatrix}$	$\begin{pmatrix} 33/8 \\ -3/2 \end{pmatrix}$	$\begin{pmatrix} 33/8 \\ -3/2 \end{pmatrix}$	
{5,6}	$\begin{pmatrix} -1 & -1 \\ -6 & 1 \end{pmatrix}$	$-\frac{1}{7} \begin{pmatrix} 1 & 1 \\ 6 & -1 \end{pmatrix}$	$\begin{pmatrix} -33/7 \\ -53/7 \end{pmatrix}$	$\begin{pmatrix} -33/7 \\ -53/7 \end{pmatrix}$	

4. Il metodo del simplesso

Situazione 4.1. $n, m \in \mathbb{N} + 1$ ed $A \in \mathbb{R}_n^m$, $f \in \mathbb{R}_n$, $b \in \mathbb{R}^m$.
Supponiamo inoltre che $\text{rango}(A) = m$.

Definizione 4.2. Per $x \in \mathbb{R}^n$ sia $J(x, A) := \{\alpha \in J(A) \mid [x] \subset \alpha\}$.

Osservazione 4.3. Sia $x \in (A\mathbb{R}_+^n = b)$. Allora

$$x \in \text{vertici}(A\mathbb{R}_+^n = b) \iff J(x, A) \neq \emptyset$$

Dimostrazione. (1) Sia $x \in \text{vertici}(A\mathbb{R}_+^n = b)$. Poichè $\text{rango}(A) = m$, dal corollario 3.17 segue che $J(x, A) \neq \emptyset$.

(2) Sia $\alpha \in J(A)$ con $[x] \subset \alpha$. Allora le colonne di A_α sono linearmente indipendenti e quindi anche le colonne di $A_{[x]}$ lo sono. Per il teorema 3.13 si ha che $x \in \text{vertici}(A\mathbb{R}_+^n = b)$.

Lemma 4.4. Siano $x, y \in (A\mathbb{R}^n = b)$ ed $\alpha \in J(x, A)$. Allora

$$fy = fx + (f - f_\alpha A_\alpha^{-1} A)_{1-\alpha} y^{1-\alpha}$$

Dimostrazione. L'ipotesi implica $x^\alpha = A_\alpha^{-1} b$ ed $x^{1-\alpha} = 0$.
Inoltre $b = Ay = A_\alpha y^\alpha + A_{1-\alpha} y^{1-\alpha}$, per cui

$$y^\alpha = A_\alpha^{-1} b - A_\alpha^{-1} A_{1-\alpha} y^{1-\alpha} = x^\alpha - A_\alpha^{-1} A_{1-\alpha} y^{1-\alpha}$$

cosicchè

$$\begin{aligned} fy &= f_\alpha y^\alpha + f_{1-\alpha} y^{1-\alpha} = f_\alpha x^\alpha - f_\alpha A_\alpha^{-1} A_{1-\alpha} y^{1-\alpha} + f_{1-\alpha} y^{1-\alpha} \\ &= fx + (f - f_\alpha A_\alpha^{-1} A)_{1-\alpha} y^{1-\alpha} \end{aligned}$$

Corollario 4.5. Siano $x \in (A\mathbb{R}_+^n = b)$ ed $\alpha \in J(x, A)$ tali che $(f - f_\alpha A_\alpha^{-1} A)_{1-\alpha} \leq 0$.

Allora x è una soluzione del problema $f(A\mathbb{R}_+^n = b) = \max$.

Dimostrazione. Infatti dal lemma 4.4 si ha che $fy \leq fx$ per ogni $y \in (A\mathbb{R}_+^n = b)$.

Definizione 4.6. Siano $x \in \mathbb{R}^n$ ed $\alpha \in J(x, A)$.

Per $j \in 1 - \alpha$ e $t \in \mathbb{R}$ definiamo $x(t, j, \alpha) \in \mathbb{R}^n$ mediante

$$x^\alpha(t, j, \alpha) := x^\alpha - t A_\alpha^{-1} A_j$$

$$x^j(t, j, \alpha) := t$$

$$x^i(t, j, \alpha) := 0 \text{ per } i \in (1 - \alpha) \setminus \{j\}$$

Osservazione 4.7. Siano $x \in (A\mathbb{R}^n = b)$, $\alpha \in J(x, A)$, $j \in 1 - \alpha$ e $t \in \mathbb{R}$. Allora:

(1) $Ax(t, j, \alpha) = b$.

(2) $fx(t, j, \alpha) = fx + t(f_j - f_\alpha A_\alpha^{-1} A_j)$.

(3) $x(0, j, \alpha) = x$.

Dimostrazione.

$$\begin{aligned}(1) \quad Ax(t, j, \alpha) &= A_\alpha x^\alpha(t, j, \alpha) + A_{1-\alpha} x^{1-\alpha}(t, j, \alpha) \\ &= A_\alpha(x^\alpha - tA_\alpha^{-1}A_j) + A_j x^j(t, j, \alpha) \\ &= A_\alpha x^\alpha - tA_j + tA_j = A_\alpha x^\alpha = b\end{aligned}$$

(2) Per il punto (1) possiamo applicare il lemma 4.4, ottenendo

$$\begin{aligned}fx(t, j, \alpha) &= fx + (f - f_\alpha A_\alpha^{-1}A)_{1-\alpha} x^{1-\alpha}(t, j, \alpha) \\ &= fx + t(f - f_\alpha A_\alpha^{-1}A)_j = fx + t(f_j - f_\alpha A_\alpha^{-1}A_j)\end{aligned}$$

(3) Chiaro.

Corollario 4.8. *Nelle ipotesi dell'oss. 4.7 sia $f_j - f_\alpha A_\alpha^{-1}A_j > 0$.*

Allora $fx(t, j, \alpha) > fx$ per ogni $t > 0$.

Proposizione 4.9. *Siano $x \in (\mathbb{A}\mathbb{R}_+^n = b)$, $\alpha \in J(x, A)$, $j \in 1 - \alpha$ ed $f_j - f_\alpha A_\alpha^{-1}A_j > 0$. Se $A_\alpha^{-1}A_j \leq 0$, allora il problema $f(\mathbb{A}\mathbb{R}_+^n = b) = \max$ non possiede soluzioni.*

Dimostrazione. Per l'oss. 4.7 $\lim_{t \rightarrow \infty} fx(t, j, \alpha) = \infty$. Dobbiamo però verificare che $x(t, j, \alpha) \geq 0$ per ogni $t \geq 0$. Ma per $t \geq 0$ l'ipotesi $A_\alpha^{-1}A_j \leq 0$ implica

$$\begin{aligned}x(t, j, \alpha) &= x^\alpha(t, j, \alpha) + x^{1-\alpha}(t, j, \alpha) \\ &= x^\alpha - tA_\alpha^{-1}A_j + x^j(t, j, \alpha) = x^\alpha - tA_\alpha^{-1}A_j + t \geq 0\end{aligned}$$

Nota 4.10. Siano $x \in \text{vertici}(\mathbb{A}\mathbb{R}_+^n = b)$, $\alpha \in J(x, A)$, $j \in 1 - \alpha$ e $w := A_\alpha^{-1}A_j \in \mathbb{R}^m$. Scriviamo α nella forma $\alpha = \{\alpha_1, \dots, \alpha_m\}$. Assumiamo che $f_j - f_\alpha w > 0$ e che esista almeno un indice $i \in \{1, \dots, m\}$ tale che $w^i > 0$. Siano $t := \min\{x^{\alpha_i}/w^i \mid i \in \{1, \dots, m\}, w^i > 0\}$, ad esempio $t = \frac{x^{\alpha_k}}{w^k}$, e $\beta := (\alpha \cup \{j\}) \setminus \{\alpha_k\}$. Allora:

- (1) $|\beta| = m$.
- (2) $[x(t, j, \alpha)] \subset \beta$.
- (3) $\beta \in J(A)$.
- (4) $x(t, j, \alpha) \in \text{vertici}(\mathbb{A}\mathbb{R}_+^n = b)$.
- (5) Se $t > 0$, allora $fx(t, j, \alpha) > fx$.

Dimostrazione. (1) Chiaro.

(2) Bisogna dimostrare che $x^{\alpha_k}(t, j, \alpha) = 0$.
Ma $x^{\alpha_k}(t, j, \alpha) = x^{\alpha_k} - tw^k = 0$.

(3) Dobbiamo dimostrare che le colonne di $A_{[\beta]}$ sono linearmente indipendenti. Possiamo assumere che $\alpha = \{1, \dots, m\}$, $k = m$ e $j > m$. In particolare allora $w^m > 0$.

Siano $\lambda_1, \dots, \lambda_{m-1}, \lambda_j \in \mathbb{R}$ tali che $\lambda_1 A_1 + \dots + \lambda_{m-1} A_{m-1} + \lambda_j A_j = 0$. Però $A_j = A_\alpha w$, per cui abbiamo

$$\begin{aligned}
0 &= \lambda_1 A_1 + \dots + \lambda_{m-1} A_{m-1} + \lambda_j A_\alpha w \\
&= \lambda_1 A_1 + \dots + \lambda_{m-1} A_{m-1} + \lambda_j (A_1 w^1 + \dots + A_m w^m) \\
&= (\lambda_1 + \lambda_j w^1) A_1 + \dots + (\lambda_{m-1} + \lambda_j w^{m-1}) A_{m-1} + \lambda_j w^m A_m
\end{aligned}$$

Siccome le colonne di α sono linearmente indipendenti, segue che $\lambda_1 + \lambda_j w^1 = \dots = \lambda_{m-1} + \lambda_j w^{m-1} = \lambda_j w^m = 0$. Ma $w^m \neq 0$ e quindi $\lambda_j = 0$ e ciò a sua volta implica $\lambda_1 = \dots = \lambda_{m-1} = 0$.

(4) Teorema 3.13 (oppure oss. 4.3).

(5) Ciò segue dal cor. 4.8.

Definizione 4.11. Un vertice di $(A\mathbb{R}_+^n = b)$ si dice *non degenero* se $|\alpha| = m$.

Osservazione 4.12. Sia $x \in (A\mathbb{R}_+^n = b)$. Allora sono equivalenti:

- (1) x è un vertice non degenero di $(A\mathbb{R}_+^n = b)$.
- (2) $J(x, A) = \{\alpha\}$.

Dimostrazione. Chiaro per l'oss. 4.3, perché per ogni $\alpha \in J(x, A)$ vale $|\alpha| = m$.

Teorema 4.13. x sia un vertice non degenero di $(A\mathbb{R}_+^n = b)$.

Sia $\alpha := [x] = \{\alpha_1, \dots, \alpha_m\}$. Allora:

- (1) Se $(f - f_\alpha A_\alpha^{-1} A)_{1-\alpha} \leq 0$, allora x è una soluzione del problema $f(A\mathbb{R}_+^n = b) = \max$.
- (2) Altrimenti esiste $j \in 1 - \alpha$ tale che $f_j - f_\alpha A_\alpha^{-1} A_j > 0$.
In questo caso poniamo $w := A_\alpha^{-1} A_j$.
Se $w \leq 0$, allora il problema $f(A\mathbb{R}_+^n = b) = \max$ non possiede soluzioni.
Altrimenti, posto $t := \min\{x^{\alpha_i}/w^i \mid i \in \{1, \dots, m\}, w^i > 0\}$, si ha $f(x(t, j, \alpha)) > f(x)$.

Dimostrazione. (1) Cor. 4.5.

(2) Il caso $w \leq 0$ segue dalla prop. 4.9. Se non vale $w \leq 0$, per la nota 4.10 è sufficiente dimostrare che $t > 0$. Ma per ipotesi $\alpha = [x]$, per cui $x^{\alpha_i} > 0$ per ogni $i \in \{1, \dots, m\}$, cosicché anche $t > 0$.

Nota 4.14. L'algoritmo del simplesso per la risoluzione del problema di massimo $f(A\mathbb{R}_+^n = b) = \max$ può così essere formulato nel modo seguente (x è un vertice non degenero di $(A\mathbb{R}_+^n = b)$):

- (1) $\alpha = [x]$ (necessariamente $|\alpha| = m$). Scriviamo $\alpha = \{\alpha_1, \dots, \alpha_m\}$ con $\alpha_1 < \dots < \alpha_m$.
- (2) $\beta = 1 - \alpha$.
- (3) $g = f_\alpha A_\alpha^{-1}$.

- (4) Determiniamo il più piccolo indice $j \in 1 - \alpha$ tale che $f_j > gA_j$, se un tale j esiste.
- (5) Se j non esiste, x è un punto di massimo, l'algoritmo restituisce x e termina.
- (6) $w = A_\alpha^{-1}A_j$; $t = \infty$. Cerchiamo l'insieme $I := \{i \in \alpha \mid w^i > 0\}$.
- (7) Per $i = 1, \dots, m$ controlliamo se $w^i > 0$; in tal caso sia $s = x^{\alpha_i}/w^i$ e se $s < t$, allora poniamo $t = s$ e $k = i$.
- (8) Se dopo il ciclo in (7) risulta ancora $t = \infty$, allora il problema non possiede soluzione e viene restituito il valore ∞ (o un altro valore che esprime la non risolubilità).
- (9) $x = x(t, j, \alpha)$.
- (10) $\alpha = (\alpha \setminus \{\alpha_k\}) \cup \{j\}$.
- (11) Torniamo al punto (2).

La scelta di j e k nei punti (4) e (7) porta il nome di *regola di Bland* e garantisce che l'algoritmo termina sempre e non incorre in cicli infiniti; una dimostrazione si trova ad esempio in Geiger/Kanzow, pagg. 108-111.

Nota 4.15. Traduciamo l'algoritmo in Octave (o in Matlab):

```
function y=simpleso(A,b,f,x,mostra=0)
n=length(x); alfa=find(x)';
while 1
if mostra X=x', Ax=(A*x)' endif
beta=complement(alfa,1:n);
f_alfa=f(alfa); A_alfa=A(:,alfa); g=f_alfa/A_alfa;
trovatoj=0;
for j=beta Aj=A(:,j);
if f(j)>g*Aj trovatoj=1;break; endif; endfor
if !trovatoj y=x; break; endif
w=A_alfa\Aj; t=Inf;
for i=1:length(alfa)
if w(i)>0 s=x(alfa(i))/w(i);
if s<t t=s; k=i; endif; endif; endfor
if t==Inf y=Inf; break; endif
x=sostituzione(x,t,j,alfa,w,n);
alfa=union(complement(alfa(k),alfa),j);
endwhile; end

function y=sostituzione(x,t,j,alfa,w,n)
y=zeros(n,1); y(alfa)=x(alfa)-t*w; y(j)=t;
end
```

Osservazione 4.16. Octave per il calcolo di programmi lineari prevede la funzione *glpk*; cfr. Eaton/Bateman/Hauberg, pagg. 339-345.

Osservazione 4.17. Quando il problema è dato nella forma standard,

$$fx = \max$$

$$x \geq 0$$

$$Ax \leq b$$

con la tecnica della nota 3.3 la possiamo trasformare in forma normale

$$(f \ 0) \begin{pmatrix} x \\ p \end{pmatrix} = \max$$

$$\begin{pmatrix} x \\ p \end{pmatrix} \geq 0$$

$$(A \ \delta) \begin{pmatrix} x \\ p \end{pmatrix} = b$$

In questo caso il vettore $\begin{pmatrix} 0 \\ b \end{pmatrix}$ è un vertice non degenere del nuovo problema e possiamo applicare l'algoritmo della nota 4.14.

Osservazione 4.18. Consideriamo adesso un problema della forma

$$fx = \max$$

$$x \geq 0$$

$$Ax = b$$

con A, f, b definiti come in precedenza. Esso evidentemente è equivalente al problema

$$fx - \infty u^1 - \dots - \infty u^m = \max$$

$$x \geq 0, u \geq 0$$

$$Ax + u = b$$

dove con ∞ calcoliamo in modo naturale. ∞ in Matlab (o Octave) viene rappresentato da *Inf* e i calcoli vengono eseguiti correttamente.

Ciò ci permette di applicare di nuovo l'algoritmo del simplesso. Una dimostrazione dettagliata si trova in Geiger/Kanzow, pagg. 111-115.

5. Geometria elementare degli insiemi convessi

Nota 5.1. V sia uno spazio vettoriale reale. Continuiamo la discussione delle proprietà geometriche elementari degli insiemi convessi iniziata nel primo capitolo.

Definizione 5.2. Usiamo le notazioni introdotte in precedenza e poniamo inoltre:

$$\mathbb{R}_{++}^n := \{x \in \mathbb{R}^n \mid x > 0\}$$

$$\mathbb{R}_m^{++} := \{f \in \mathbb{R}_m \mid f > 0\}$$

$$\mathbb{R}_{\text{stoc}}^n := \{\lambda = (\lambda^1, \dots, \lambda^n) \in \mathbb{R}_+^n \mid \sum_{h=1}^n \lambda^h = 1\}$$

La condizione $x > 0$ significa che $x^h > 0$ per ogni h ; la condizione $f > 0$ è definita similmente.

$\mathbb{R}_{\text{stoc}}^n$ è l'insieme dei vettori stocastici in \mathbb{R}^n .

Per un sottoinsieme A di uno spazio topologico sia $\text{int } A$ l'interno di A .

Lemma 5.3. Siano $a, x \in \mathbb{R}^m$ e $t, \rho \in \mathbb{R}^{++}$. Allora

$$a + t(|\mathbb{R}^m - x| < \rho) = (|\mathbb{R}^m - (a + tx)| < t\rho)$$

Dimostrazione. Per $y \in \mathbb{R}^m$ sono equivalenti:

$$(1) |y - (a + tx)| < t\rho$$

$$(2) \left| \frac{y-a}{t} - x \right| < \rho$$

$$(3) \frac{y-a}{t} \in (|\mathbb{R}^m - x| < \rho)$$

$$(4) y \in a + t(|\mathbb{R}^m - x| < \rho)$$

Proposizione 5.4. X sia un sottoinsieme convesso di \mathbb{R}^m .

Allora anche $\text{int } X$ e \overline{X} sono convessi.

Dimostrazione. (1) Siano $x, y \in \text{int } X$ e $t \in (0, 1)$. Allora esiste $\rho > 0$ tale che $(|\mathbb{R}^m - x| < \rho) \subset X$. Sia $z := tx + (1-t)y$. Dobbiamo dimostrare che $z \in \text{int } X$. Usando la convessità di X e il lemma 5.3 con $a = (1-t)y$ si ha che

$$\begin{aligned} X \supset (1-t)y + t(|\mathbb{R}^m - x| < \rho) &\stackrel{5.3}{=} (|\mathbb{R}^m - ((1-t)y + tx)| < t\rho) = \\ &(|\mathbb{R}^m - z| < t\rho) \end{aligned}$$

Abbiamo quindi trovato un intorno aperto di z contenuto in X per cui $z \in \text{int } X$.

(2) Siano $u, v \in \overline{X}$ e $t \in [0, 1]$. Allora esistono due successioni \tilde{x}, \tilde{y} in X tali che $\tilde{x} \rightarrow u, \tilde{y} \rightarrow v$. Ciò implica $t\tilde{x} + (1-t)\tilde{y} \rightarrow tu + (1-t)v =: z$. Per la convessità di X però gli elementi di $t\tilde{x} + (1-t)\tilde{y}$ appartengono tutti ad X e quindi $z \in \overline{X}$.

Definizione 5.5. Per sottoinsiemi X, Y di V e $\lambda \in \mathbb{R}$ poniamo:

$$X + Y := \{x + y \mid x \in X, y \in Y\}$$

$$\lambda X := \{\lambda x \mid x \in X\}$$

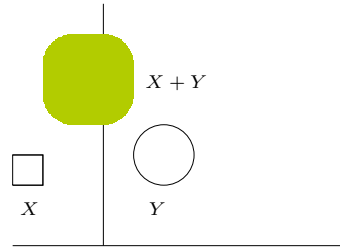
Osservazione 5.6. X ed Y siano sottoinsiemi convessi di V e sia $\lambda \in \mathbb{R}$. Allora anche gli insiemi $X + Y$ e λX sono convessi.

Dimostrazione. (1) Ciò è evidente per λX .

(2) Siano $u_1, u_2 \in X + Y$. Allora esistono $x_1, x_2 \in X$ e $y_1, y_2 \in Y$ tali che $u_1 = x_1 + y_1$ e $u_2 = x_2 + y_2$. Per $t \in [0, 1]$ allora

$tx_1 + (1 - t)x_2 \in X$ e $ty_1 + (1 - t)y_2 \in Y$, per cui

$$\begin{aligned} tu_1 + (1 - t)u_2 &= t(x_1 + y_1) + (1 - t)(x_2 + y_2) \\ &= (tx_1 + (1 - t)x_2) + (ty_1 + (1 - t)y_2) \in X + Y \end{aligned}$$



Definizione 5.7. Un cono di V è un sottoinsieme non vuoto X di V tale che $\lambda X \subset X$ per ogni $\lambda > 0$.

Esempio 5.8. Gli insiemi \mathbb{R}_+^m e \mathbb{R}_{++}^m sono coni convessi di \mathbb{R}^m .

L'insieme $(\mathbb{R}^+ \times 0) \cup (0 \times \mathbb{R}^+)$ in \mathbb{R}^2 (l'unione dei due semiassi delle coordinate) è un cono, ma non è convesso.

Osservazione 5.9. X sia un cono di V .

Allora X è convesso se e solo se $X + X \subset X$.

Dimostrazione. Siano $x, y \in X$.

(1) X sia convesso. Allora $\frac{x+y}{2} \in X$ perché X è convesso, e quindi $x + y = 2 \frac{x+y}{2} \in X$ perché X è un cono.

(2) Sia $X + X \subset X$ e siano $t \in [0, 1]$ e $z := tx + (1 - t)y$. Per $t = 0$ risp. $t = 1$ si ha rispettivamente $z = y \in X$ e $z = x \in X$.

Per $t \in (0, 1)$ invece $tx \in X$ e $(1 - t)y \in X$ perché X è un cono. Per ipotesi quindi $z \in X$.

Osservazione 5.10. L'intersezione di una famiglia arbitraria di sottoinsiemi convessi di V è ancora convessa.

Abbiamo utilizzato questa osservazione già nel cor. 3.5.

Definizione 5.11. Generalizzando la notazione della def. 1.7 per sottoinsiemi non vuoti $X_1, \dots, X_n \subset V$ poniamo

$$[X_1, \dots, X_n] := \left\{ \sum_{k=1}^n \lambda^k x_k \mid n \in \mathbb{N} + 1, x_1 \in X_1, \dots, x_n \in X_n, \lambda \in \mathbb{R}_{\text{stoc}}^n \right\}$$

Si noti che $[X_1] = X_1$.

Definizione 5.12. Per $X \subset V$ sia $\text{conv } X$ l'intersezione di tutti i sottoinsiemi convessi di V che contengono X .

Per l'oss. 5.10 questo insieme, la *chiusura convessa* di X , è il più piccolo insieme convesso in V che contiene X .

Per $X = \{x_1, \dots, x_n\}$ scriviamo $\text{conv}(x_1, \dots, x_n)$ invece di $\text{conv } X$.

Osservazione 5.13. Sia X un sottoinsieme non vuoto di V . Allora

$$\text{conv } X = \left\{ \sum_{k=1}^n \lambda^k x_k \mid n \in \mathbb{N} + 1, x_1, \dots, x_n \in X, \lambda \in \mathbb{R}_{\text{stoc}}^n \right\} = \bigcup_{n=1}^{\infty} \underbrace{[X, \dots, X]}_n$$

Dimostrazione. È immediato che l'insieme alla destra è convesso e contenuto in ogni insieme convesso che contiene X .

Osservazione 5.14. Siano X_1, \dots, X_n sottoinsiemi non vuoti di V .

Allora $[X_1, \dots, X_n] \subset \text{conv}(X_1 \cup \dots \cup X_n)$

Teorema 5.15. X_1, \dots, X_n siano sottoinsiemi convessi e non vuoti di V . Allora l'insieme $[X_1, \dots, X_n]$ è convesso e

$$[X_1, \dots, X_n] = \text{conv}(X_1 \cup \dots \cup X_n).$$

Dimostrazione. Per l'oss. 5.14 è sufficiente dimostrare che $[X_1, \dots, X_n]$ è convesso. Siano $u, v \in [X_1, \dots, X_n]$ e $t \in [0, 1]$. Allora esistono $x_1, y_1 \in X_1, \dots, x_n, y_n \in X_n$ e $\lambda, \mu \in \mathbb{R}_{\text{stoc}}^n$ tali che $u = \lambda^1 x_1 + \dots + \lambda^n x_n$ e $v = \mu^1 y_1 + \dots + \mu^n y_n$. Ponendo $s := 1 - t$ abbiamo

$$su + tv = \sum_{k=1}^n (s\lambda^k x_k + t\mu^k y_k) = \sum_{k=1}^n \rho^k z_k$$

dove per ogni k abbiamo posto $\rho^k := s\lambda^k + t\mu^k$ e

$$z_k := \begin{cases} \frac{s\lambda^k}{\rho^k} x_k + \frac{t\mu^k}{\rho^k} y_k & \text{per } \rho^k \neq 0 \\ x_k & \text{altrimenti} \end{cases}$$

Dalla convessità di X_k segue che $z_k \in X_k$ per ogni k . Inoltre

$$\sum_{k=1}^n \rho^k = \sum_{k=1}^n (s\lambda^k + t\mu^k) = s \sum_{k=1}^n \lambda^k + t \sum_{k=1}^n \mu^k = s + t = 1$$

Ciò mostra che $su + tv \in [X_1, \dots, X_n]$.

Corollario 5.16. Siano $x_1, \dots, x_n \in V$. Allora

$$\text{conv}(x_1, \dots, x_n) = [x_1, \dots, x_n].$$

Osservazione 5.17. Siano X_1, \dots, X_n sottoinsiemi non vuoti di V . Allora

$$[X_1, \dots, X_n] = \bigcup_{(x_1, \dots, x_n) \in X_1 \times \dots \times X_n} [x_1, \dots, x_n] = \bigcup_{(x_1, \dots, x_n) \in X_1 \times \dots \times X_n} \text{conv}(x_1, \dots, x_n)$$

Corollario 5.18. X_1, \dots, X_n siano sottoinsiemi convessi e non vuoti di V . Allora

$$\text{conv}(X_1 \cup \dots \cup X_n) = \bigcup_{(x_1, \dots, x_n) \in X_1 \times \dots \times X_n} \text{conv}(x_1, \dots, x_n)$$

Proposizione 5.19. X sia un sottoinsieme limitato di \mathbb{R}^m .

Allora $\text{conv } X$ è limitato.

Dimostrazione. Siccome X è limitato, esiste una palla K tale che $X \subset K$. Allora $\text{conv } X \subset \text{conv } K = K$.

Proposizione 5.20. X sia un aperto di \mathbb{R}^m .

Allora anche $\text{conv } X$ è un aperto.

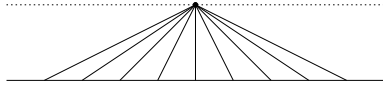
Dimostrazione. Dobbiamo dimostrare che $\text{conv } X \subset \text{int conv } X$. Per ipotesi però $\text{int } X = X$, per cui

$$\text{conv } X = \text{conv int } X \subset \text{conv int conv } X \stackrel{5.4}{=} \text{int conv } X$$

Osservazione 5.21. La chiusura convessa di un chiuso di \mathbb{R}^m invece in generale non è più un chiuso.

Infatti sia $X := (\mathbb{R} \times 0) \cup \{i\}$ in $\mathbb{R}^2 = \mathbb{C}$ l'insieme costituito dalla retta reale e dal numero immaginario i . Per il cor. 5.18 allora

$\text{conv } X = \bigcup_{a \in \mathbb{R}} [a, i] = (\mathbb{R} \times [0, 1]) \cup \{i\}$. Ma questo insieme non è chiuso.



Definizione 5.22. Per $X \subset \mathbb{R}^m$ ed $Y \subset \mathbb{R}^k$ usiamo l'abbreviazione

$$X \square Y := \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \mid x \in X, y \in Y \right\} \subset \mathbb{R}^{m+k}$$

Per $k = 1$ ed $Y = \{1\}$ abbiamo in particolare $X \square 1 \subset \mathbb{R}^{m+1}$.

Similmente per $x \in \mathbb{R}^m$ abbiamo $x \square 1 = \begin{pmatrix} x \\ 1 \end{pmatrix} \in \mathbb{R}^{m+1}$.

Lemma 5.23. Siano $X \subset \mathbb{R}^m$ e $v \in \mathbb{R}^m$. Allora

$$v \in \text{conv } X \iff v \square 1 \in \text{POS}(X \square 1)$$

Dimostrazione. Per $X = \emptyset$ l'enunciato è banalmente vero. Sia quindi $X \neq \emptyset$.

(1) Sia $v \in \text{conv } X$. Allora esistono $x_1, \dots, x_n \in X$ e $\lambda \in \mathbb{R}_{\text{stoc}}^n$ tali che $v = \lambda^1 x_1 + \dots + \lambda^n x_n$. Ciò implica

$$\begin{aligned} v \square 1 &= (\lambda^1 x_1 + \dots + \lambda^n x_n) \square 1 \\ &= \lambda^1 (x_1 \square 1) + \dots + \lambda^n (x_n \square 1) \in \text{POS}(x \square 1) \end{aligned}$$

(2) Sia viceversa $v \square 1 \in \text{POS}(X \square 1)$. Allora esistono $x_1, \dots, x_n \in X$ e $\lambda^1, \dots, \lambda^n \geq 0$ tali che $v \square 1 = \lambda^1 (x_1 \square 1) + \dots + \lambda^n (x_n \square 1)$. Ma ciò è equivalente alle relazioni $v = \lambda^1 x_1 + \dots + \lambda^n x_n$ e $\lambda^1 + \dots + \lambda^n = 1$.

Definizione 5.24. I punti $x_1, \dots, x_s \in V$ si dicono *affinemente indipendenti*, se una coppia di relazioni

$$\lambda^1 x_1 + \dots + \lambda^s x_s = 0$$

$$\lambda^1 + \dots + \lambda^s = 0$$

con $\lambda^1, \dots, \lambda^s \in \mathbb{R}$ implica $\lambda^1 = \dots = \lambda^s = 0$.

Definizione 5.25. X sia un sottoinsieme non vuoto di V . La dimensione $\dim X$ di X è definita come la dimensione dello spazio affine $\text{aff } X$ generato da X .

È chiaro che $\dim X$ è uguale alla dimensione usuale di X , se X è un sottospazio affine di X .

Si dimostra facilmente che $\dim X$ coincide con il massimo $d \in \mathbb{N}$ tale che esistono $d + 1$ punti affinemente indipendenti in X .

Osservazione 5.26. Siano $x_1, \dots, x_s \in \mathbb{R}^m$ e $\lambda^1, \dots, \lambda^s \in \mathbb{R}$. Allora sono equivalenti:

$$(1) \lambda^1 x_1 + \dots + \lambda^s x_s = 0 \quad \text{e} \quad \lambda^1 + \dots + \lambda^s = 0.$$

$$(2) \lambda^1(x_1 \square 1) + \dots + \lambda^s(x_s \square 1) = 0.$$

Corollario 5.27. X sia un sottoinsieme non vuoto di \mathbb{R}^m . Allora $\dim X = \dim \text{SV}(X \square 1) - 1$

Corollario 5.28. Siano $x_1, \dots, x_s \in \mathbb{R}^m$. Allora sono equivalenti:

(1) I punti x_1, \dots, x_s sono affinemente indipendenti.

(2) $x_1 \square 1, \dots, x_s \square 1$ sono linearmente indipendenti.

Teorema 5.29 (teorema di Carathéodory). X sia un sottoinsieme non vuoto di \mathbb{R}^m con $\dim X = d$. Allora $\text{conv } X = \underbrace{[X, \dots, X]}_{d+1}$.

Dimostrazione. Sia $v \in \text{conv } X$. Per il lemma 5.23 $v \square 1 \in \text{POS}(X \square 1)$. Per il lemma di Carathéodory (prop. 1.10) e il cor. 5.27 esistono $x_1, \dots, x_{d+1} \in X$ e $\lambda^1, \dots, \lambda^{d+1} \in \mathbb{R}^+$ tali che

$$v \square 1 = \lambda^1(x_1 \square 1) + \dots + \lambda^{d+1}(x_{d+1} \square 1)$$

Ciò significa che $v = \lambda^1 x_1 + \dots + \lambda^{d+1} x_{d+1}$ e $\lambda^1 + \dots + \lambda^{d+1} = 1$ e vediamo che $v \in [x_1, \dots, x_{d+1}]$.

Lemma 5.30. X_1, \dots, X_n siano sottoinsiemi compatti e non vuoti di \mathbb{R}^m . Allora $[X_1, \dots, X_n]$ è compatto.

Dimostrazione. L'applicazione

$$f : X_1 \times \dots \times X_n \times \mathbb{R}_{\text{stoc}}^n \longrightarrow \mathbb{R}^m$$

$$(x_1, \dots, x_n, \lambda) \longrightarrow \lambda^1 x_1 + \dots + \lambda^n x_n$$

è continua e si ha $[X_1, \dots, X_n] = f(X_1 \times \dots \times X_n \times \mathbb{R}_{\text{stoc}}^n)$.

Anche il simpleso $\mathbb{R}_{\text{stoc}}^n$ è compatto e ciò implica l'enunciato.

Proposizione 5.31. *X sia un compatto di \mathbb{R}^m . Allora anche $\text{conv } X$ è compatto, quindi chiuso e limitato.*

Dimostrazione. Segue dal teorema 5.29 e dal lemma 5.30.

Corollario 5.32. *Siano $x_1, \dots, x_n \in \mathbb{R}^m$. Allora $\text{conv}(x_1, \dots, x_n)$ è compatto.*

Proposizione 5.33. *Siano $x_1, \dots, x_n \in \mathbb{R}^m$ ed $y \in \mathbb{R}^m$. Per un punto $p \in X$ sono allora equivalenti:*

(1) *p è la proiezione di y su $\text{conv}(x_1, \dots, x_n)$.*

(2) *$|x_k - p, y - p| \leq 0$ per ogni $k = 1, \dots, n$.*

Dimostrazione. Sia $z = \sum_{k=1}^n \lambda^k x_k$ con $\lambda \in \mathbb{R}_{\text{stoc}}^n$. Per la prop. 1.19 è sufficiente dimostrare che la condizione (2) implica $\|z - p, y - p\| \leq 0$.

Osserviamo che $z - p = \sum_{k=1}^n \lambda^k x_k - p = \sum_{k=1}^n \lambda^k (x_k - p)$ perchè $\sum_{k=1}^n \lambda^k = 1$.

Perciò

$$\|z - p, y - p\| = \left\| \sum_{k=1}^n \lambda^k (x_k - p), y - p \right\| \stackrel{(2)}{\leq} 0$$

dove abbiamo anche usato che i coefficienti λ^k sono tutti non negativi.

Lemma 5.34. *X sia un sottoinsieme chiuso e convesso di \mathbb{R}^m tale che $0 \notin X$. Allora esistono $\alpha \in \mathbb{R}$ ed $f \in \mathbb{R}_m$ tali che $f x > \alpha > 0$ per ogni $x \in X$.*

Dimostrazione. Ciò segue dal cor. 1.23 per $y = 0$.

Corollario 5.35. *Siano $x_1, \dots, x_n \in \mathbb{R}^m$ tali che $0 \notin \text{conv}(x_1, \dots, x_n)$. Allora esistono $\alpha \in \mathbb{R}$ ed $f \in \mathbb{R}_m$ tali che $f x_k > \alpha > 0$ per ogni $k = 1, \dots, n$.*

Osservazione 5.36. *Siano $x_1, \dots, x_n \in \mathbb{R}^m$ ed $\alpha \in \mathbb{R}$ tali che $f x_k > \alpha$ risp. $f x_k \geq \alpha$ per ogni $k = 1, \dots, n$.*

Allora $f z > \alpha$ risp. $f z \geq \alpha$ per ogni $z \in \text{conv}(x_1, \dots, x_n)$.

Dimostrazione. Supponiamo che $f x_k > \alpha$ per ogni $k = 1, \dots, n$.

Sia $z = \sum_{k=1}^n \lambda^k x_k$ con $\lambda \in \mathbb{R}_{\text{stoc}}^n$. Allora

$$f z = \sum_{k=1}^n \lambda^k f x_k > \alpha \sum_{k=1}^n \lambda^k = \alpha$$

Qui abbiamo usato che $\sum_{k=1}^n \lambda^k = 1$ e che i coefficienti λ^k , i quali sono non negativi, non possono essere tutti nulli.

Nello stesso modo si dimostra la seconda parte dell'enunciato.

II. ALCUNE APPLICAZIONI

6. Il separatore di Bennett-Mangasarian

Situazione 6.1. Siano $X \in \mathbb{R}^m$, $Y \in \mathbb{R}^k$.

Osservazione 6.2. Consideriamo i sottoinsiemi $\tilde{X} := \{X^1, \dots, X^m\}$ ed $\tilde{Y} := \{Y^1, \dots, Y^k\}$ di \mathbb{R}^n . Seguendo il lavoro di Bennett/Mangasarian, faremo vedere come il compito di separare in modo approssimato (ma ottimale) i punti di \tilde{X} ed \tilde{Y} tramite un iperpiano possa essere ricondotto ad un compito di programmazione lineare.

Definizione 6.3. Per $l \in \mathbb{N} + 1$ siano $1^{(l)} := \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^l$,

$1_{(l)} := (1, \dots, 1) \in \mathbb{R}_l$.

Definizione 6.4. Per $\alpha \in \mathbb{R}$ sia $\alpha_+ := \max(\alpha, 0)$.

Per $v = \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix} \in \mathbb{R}^n$ siano $v_+ := \begin{pmatrix} v_+^1 \\ \vdots \\ v_+^n \end{pmatrix}$ e $\|v\|_1 := \sum_{i=1}^n |v^i|$.

Si osservi che per $v \geq 0$ si ha $\|v\|_1 = 1_{(n)}v$.

Definizione 6.5. Gli insiemi \tilde{X} e \tilde{Y} si dicono *linearmente separabili*, se esiste $v \in \mathbb{R}^n$ tale che

$$\min_{1 \leq i \leq m} X^i v > \max_{1 \leq j \leq k} Y^j v$$

Lemma 6.6. Sono equivalenti:

(1) \tilde{X} e \tilde{Y} sono linearmente separabili.

(2) Esistono $w \in \mathbb{R}^n$ ed $\alpha \in \mathbb{R}$ tali che

$$Xw \geq (\alpha + 1)1^{(m)}$$

$$Yw \leq (\alpha - 1)1^{(k)}$$

Dimostrazione. (1) \implies (2): Ponendo $\text{MIN} := \min_{1 \leq i \leq m} X^i v$, $\text{MAX} := \max_{1 \leq j \leq k} Y^j v$, $\alpha := \frac{\text{MIN} + \text{MAX}}{\text{MIN} - \text{MAX}}$ e $w := \frac{2v}{\text{MIN} - \text{MAX}}$, abbiamo

$$X^i w \geq \frac{2 \text{MIN}}{\text{MIN} - \text{MAX}} = \frac{\text{MIN} - \text{MAX} + \text{MIN} + \text{MAX}}{\text{MIN} - \text{MAX}} = 1 + \alpha \quad \text{per ogni } i \quad \text{e}$$

$$Y^j w \leq \frac{2 \text{MAX}}{\text{MIN} - \text{MAX}} = \frac{\text{MIN} + \text{MAX} - (\text{MIN} - \text{MAX})}{\text{MIN} - \text{MAX}} = \alpha - 1 \quad \text{per ogni } j$$

(2) \implies (1): Chiaro.

Lemma 6.7. *Sono equivalenti:*

(1) \tilde{X} e \tilde{Y} sono linearmente separabili.

(2) $\min\{\frac{1}{m}\|(-Xw + (\alpha + 1)1^{(m)})_+\|_1 + \frac{1}{k}\|(Yw - (\alpha - 1)1^{(k)})_+\|_1$

$$| w \in \mathbb{R}^n, \alpha \in \mathbb{R}\} = 0.$$

Dimostrazione. È chiaro che (2) è equivalente alla condizione che esistono $w \in \mathbb{R}^n$ ed $\alpha \in \mathbb{R}$ tali che $-Xw + (\alpha + 1)1^{(n)} \leq 0$ e $Yw - (\alpha - 1)1^{(k)} \leq 0$ e quindi, per il lemma 6.6, alla lineare separabilità di \tilde{X} e \tilde{Y} .

Osservazione 6.8. $\min\{(1 + \alpha)_+ + (1 - \alpha)_+ | \alpha \in \mathbb{R}\} = 2.$

Dimostrazione. Per $\alpha \in \mathbb{R}$ sia $f(\alpha) := (1 + \alpha)_+ + (1 - \alpha)_+.$
Allora $f(0) = 2.$ Consideriamo 3 casi:

(1) Sia $\alpha \leq -1.$ Allora $(1 + \alpha)_+ = 0,$ $(1 - \alpha)_+ = 1 + |\alpha|$ e quindi $f(\alpha) = 0 + 1 + |\alpha| \geq 2.$

(2) Sia $-1 \leq \alpha \leq 1.$ Allora $(1 + \alpha)_+ = 1 + \alpha,$ $(1 - \alpha)_+ = 1 - \alpha$ e quindi $f(\alpha) = 1 + \alpha + 1 - \alpha = 2.$

(3) Sia $\alpha \geq 1.$ Allora $(1 + \alpha)_+ = 1 + \alpha,$ $(1 - \alpha)_+ = 0$ e quindi $f(\alpha) = 1 + \alpha + 0 \geq 2.$

Osservazione 6.9. \tilde{X} e \tilde{Y} siano linearmente separabili.
Allora nel punto (2) del lemma 6.7 $w \neq 0.$

Dimostrazione. Altrimenti avremmo la contraddizione

$$0 = \min\{\frac{1}{m}\|((\alpha + 1)1^{(m)})_+\|_1 + \frac{1}{k}\|((1 - \alpha)1^{(k)})_+\|_1 | \alpha \in \mathbb{R}\}$$

$$= \min\{(1 + \alpha)_+ + (1 - \alpha)_+ | \alpha \in \mathbb{R}\} \stackrel{6.9}{=} 2$$

Lemma 6.10. *Siano $g : \mathbb{R}^n \rightarrow \mathbb{R}^m,$ $h : \mathbb{R}^n \rightarrow \mathbb{R}^k$ ed $E \subset \mathbb{R}^n.$*

Sia $x_0 \in E.$ Allora sono equivalenti:

(1) $\|g(x_0)_+\|_1 + \|h(x_0)_+\|_1 = \min\{\|g(x)_+\|_1 + \|h(x)_+\|_1 | x \in E\}.$

(2) *Esistono $y_0 \in \mathbb{R}_+^m$ e $z_0 \in \mathbb{R}_+^k$ tali che $y_0 \geq g(x_0),$ $z_0 \geq h(x_0)$ e $1_{(m)}y_0 + 1_{(k)}z_0 = \min\{1_{(m)}y + 1_{(k)}z | y \in \mathbb{R}_+^m, z \in \mathbb{R}_+^k, x \in E, y \geq g(x), z \geq h(x)\}.$*

Dimostrazione. Nelle soluzioni in (2) necessariamente $y = g(x)_+$ e $z = h(x)_+.$ Per $y \geq 0$ però $1_{(m)}y = \|y\|_1$ e similmente per $z \geq 0.$

Proposizione 6.11. *Il compito di minimizzazione*

$$\min\{\frac{1}{m}\|(-Xw + (\alpha + 1)1^{(m)})_+\|_1 + \frac{1}{k}\|(Yw - (\alpha - 1)1^{(k)})_+\|_1 | w \in \mathbb{R}^n, \alpha \in \mathbb{R}\}$$

è equivalente al compito di ottimizzazione lineare

$$\min\left\{\frac{1}{m}1^{(m)}y + \frac{1}{k}1^{(k)}z \mid y \in \mathbb{R}_+^m, z \in \mathbb{R}_+^k, w \in \mathbb{R}^n, \right. \\ \left. Xw - \alpha 1^{(m)} + y \geq 1^{(m)}, -Yw + \alpha 1^{(k)} + z \geq 1^{(k)}\right\}.$$

Dimostrazione. La dimostrazione discende dal lemma 6.10 con $g(x) = -Xw + (\alpha + 1)1^{(m)}$ e $h(x) = Yw - (\alpha - 1)1^{(k)}$.

Nota 6.12. Nei lavori di Bennett/Mangasarian e Mangasarian/Street/Wolberg questi risultati vengono applicati a compiti di diagnosi e prognosi per il carcinoma mammario.

7. Grafi

Definizione 7.1. Un *grafo diretto* è una coppia (P, S) in cui P è un insieme finito ed $S \subset P \times P$ una relazione binaria su P .

Gli elementi di P si dicono *posizioni* (o *vertici*) del grafo diretto, gli elementi di S *archi* (o *spigoli*).

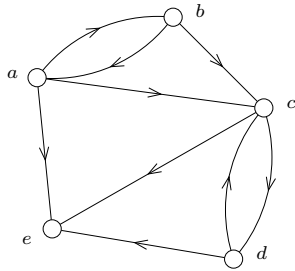
Per un arco $s = (a, b) \in S$ la posizione a si chiama *l'inizio* (o la *provenienza*) di s , la posizione b la *fine* (o il *punto d'arrivo*) di s .

Poniamo allora $s^1 := a$, $s^2 := b$ e, per $A, B \subset P$,

$$S(A, B) := \{s \in S \mid s^1 \in A, s^2 \in B\} = S \cap (A \times B)$$

Nota 7.2. Un grafo diretto viene spesso rappresentato disegnando le posizioni come punti nel piano e unendo per ogni arco (a, b) le posizioni a e b con una freccia diretta da a verso b .

Esempio 7.3.



La figura corrisponde al grafo diretto (P, S) con:

$$P = \{a, b, c, d, e\}$$

$$S = \{(a, b), (a, c), (a, e), (b, a), (b, c), (c, d), (c, e), (d, c), (d, e)\}$$

Definizione 7.4. Un *cappio* in un grafo diretto è un arco della forma (a, a) , cioè è un arco in cui la fine coincide con l'inizio.

Nota 7.5. Fissando una numerazione a_1, \dots, a_n dell'insieme delle posizioni di un grafo diretto (P, S) con $|P| = n$, gli possiamo associare una *matrice di adiacenza* $A \in \{0, 1\}_n^n$ ponendo

$$A_j^i = \begin{cases} 1 & \text{se } (a_i, a_j) \in S \\ 0 & \text{se } (a_i, a_j) \notin S \end{cases}$$

Viceversa ogni matrice $A \in \{0, 1\}_n^n$ determina, a meno di isomorfia, un grafo diretto che risulta senza cappi se e solo se la matrice contiene solo zeri nella diagonale principale.

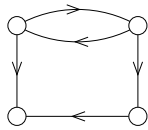
Esempio 7.6. Per il grafo dell'esempio 7.3 con le posizioni elencate nell'ordine a, b, c, d, e , la matrice di adiacenza diventa

$$\begin{pmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Viceversa la matrice

$$\begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

corrisponde ad un grafo diretto della forma

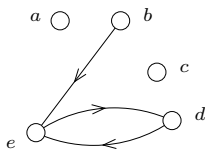


Definizione 7.7. Una posizione a in un grafo diretto (P, S) si dice *isolata*, se sono soddisfatte le seguenti condizioni:

- (1) $(a, y) \notin S$ per ogni $y \in P$.
- (2) $(x, a) \notin S$ per ogni $x \in P$.

Ciò significa che non ci sono archi che hanno a come inizio o fine.

Nella figura le posizioni a e c sono isolate.



Definizione 7.8. Un grafo diretto (P, S) si dice *bipartito*, se non possiede posizioni isolate e se esistono sottoinsiemi $A, B \subset P$ tali che:

- (1) $A \cap B = \emptyset$.
- (2) $S \subset (A \times B) \cup (B \times A)$.

Si noti che la condizione che (P, S) non abbia posizioni isolate implica $P = A \cup B$. Inoltre è chiaro che un grafo diretto bipartito non possiede cappi.

Gli insiemi A e B non sono univocamente determinati, nemmeno quando si trascura l'ordine in cui vengono presi, come mostra il grafo diretto bipartito (P, S) dato da $P := \{1, 2, 3, 4\}$, $S := \{(1, 2), (3, 4)\}$, in cui possiamo prendere ad esempio $A = \{1, 3\}$, $B = \{2, 4\}$ oppure $A = \{1, 4\}$, $B = \{2, 3\}$.

Definizione 7.9. Un grafo diretto (P, S) si dice *unidirezionalmente bipartito*, se non possiede posizioni isolate e se esistono sottoinsiemi $A, B \subset P$ tali che:

- (1) $A \cap B = \emptyset$.
- (2) $S \subset (A \times B)$.

È chiaro che (P, S) è allora anche bipartito, per cui non possiede cappi e $P = A \cup B$. In questo caso però gli insiemi A e B sono univocamente determinati, infatti

$$A = \{a \in P \mid \text{esiste } b \in P \text{ con } (a, b) \in S\}$$

$$B = \{b \in P \mid \text{esiste } a \in P \text{ con } (a, b) \in S\}$$

Le posizioni in A si chiamano *sorgenti*, le posizioni in B *destinazioni*.

Definizione 7.10. Nella teoria dei grafi si considerano anche *multigrafi diretti*. Questi possono essere definiti come coppie (P, G) in cui P è un insieme finito e $G : P \times P \rightarrow \mathbb{N}$ è un'applicazione. Anche in questo caso gli elementi di P si chiamano posizioni o vertici.

Nota 7.11. Un multigrafo diretto (P, G) viene spesso rappresentato disegnando le posizioni nel piano e unendo, per $(a, b) \in P \times P$, le posizioni a e b con tante frecce (dirette da a verso b) quante sono indicate dalla molteplicità $G(a, b)$.

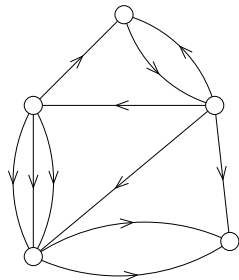
Fissando una numerazione a_1, \dots, a_n dell'insieme delle posizioni di un multigrafo diretto (P, G) con $|P| = n$, gli possiamo associare una matrice di adiacenza $A \in \mathbb{N}_n^n$ ponendo $A_j^i = n$ con $n = G(a_i, a_j)$.

Viceversa ogni matrice $A \in \mathbb{N}_n^n$ determina, a meno di isomorfia, un multigrafo diretto con n posizioni.

Così ad esempio la matrice

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \end{pmatrix}$$

corrisponde ad un multigrafo diretto della forma



Nella numerazione usata la prima posizione si trova in alto a sinistra, le altre si ottengono procedendo in senso orario.

Nota 7.12. Ogni grafo diretto (P, S) può essere considerato come un multigrafo diretto (P, G) se poniamo

$$G(a, b) := \begin{cases} 1 & \text{se } (a, b) \in S \\ 0 & \text{altrimenti} \end{cases}$$

Viceversa ogni multigrafo diretto (P, G) con $G \leq 1$ è essenzialmente la stessa cosa come un grafo diretto.

8. Flussi in un grafo diretto

Definizione 8.1. Una *rete* è una quintupla $R = (P, S, \alpha, \omega, c)$ in cui:

- (1) (P, S) è un grafo diretto senza cappi.
- (2) $\alpha, \omega \in P$ sono due posizioni con $\alpha \neq \omega$.
 α si chiama la *sorgente* della rete, ω la *destinazione*.
- (3) $c : S \rightarrow \mathbb{R}^+$ è un'applicazione.
 c si chiama la *capacità* della rete.

Situazione 8.2. $R = (P, S, \alpha, \omega, c)$ sia una rete.

Per $A, B \subset P$ usiamo la notazione

$$S(A, B) := \{s \in S \mid s^1 \in A, s^2 \in B\}$$

introdotta nella def. 7.1.

Definizione 8.3. Un *flusso* in R è una funzione $f : S \rightarrow \mathbb{R}^+$.

Definizione 8.4. Per un flusso f in R ed $A, B \subset P$ sia

$$f(A, B) := \sum_{s \in S(A, B)} f(s)$$

Per $s \in S$ abbiamo in particolare $f(s) = f(s^1, s^2)$.

Definizione 8.5. Per un flusso $f : S \rightarrow \mathbb{R}^+$ ed $a \in P$ definiamo

$$\text{ingresso}(a, f) := f(P, a)$$

$$\text{uscita}(a, f) := f(a, P)$$

$$\partial f(a) := \text{ingresso}(a, f) - \text{uscita}(a, f) = f(P, a) - f(a, P)$$

$\partial f(a)$ si chiama il *flusso locale in a indotto da f* .

Otteniamo così un'applicazione $\partial f : P \rightarrow \mathbb{R}$.

Osservazione 8.6. f sia un flusso in R ed $a \in P$. Allora

$$f(P, a) = f(P \setminus a, a)$$

$$f(a, P) = f(a, P \setminus a)$$

Dimostrazione. Chiaro perché, per definizione di rete, il grafo diretto è senza cappi, per cui $S(P, a) = S(P \setminus a, a)$ e $S(a, P) = S(a, P \setminus a)$.

Osservazione 8.7. Per $A, B, C \subset P$ con $A \cap B = \emptyset$ abbiamo

$$S(A \cup B, C) = S(A, C) \dot{\cup} S(B, C)$$

$$S(C, A \cup B) = S(C, A) \dot{\cup} S(C, B)$$

Per un flusso f abbiamo quindi

$$f(A \cup B, C) = f(A, C) + f(B, C)$$

$$f(C, A \cup B) = f(C, A) + f(C, B)$$

Lemma 8.8. *f* sia un flusso in R ed $A \subset P$. Allora

$$\sum_{a \in A} \partial f(a) = f(P \setminus A, A) - f(A, P \setminus A)$$

Dimostrazione. Usando l'oss. 8.7 abbiamo

$$\begin{aligned} \sum_{a \in A} \partial f(a) &= \sum_{a \in A} (f(P, a) - f(a, P)) = f(P, A) - f(A, P) \\ &= f(A, A) + f(P \setminus A, A) - f(A, A) - f(A, P \setminus A) \\ &= f(P \setminus A, A) - f(A, P \setminus A) \end{aligned}$$

Corollario 8.9. Sia f un flusso in R . Allora $\sum_{a \in P} \partial f(a) = 0$.

Dimostrazione. Per il lemma 8.8 abbiamo

$$\sum_{a \in P} \partial f(a) = f(\emptyset, P) - f(P, \emptyset) = 0$$

Definizione 8.10. Un flusso $f : S \rightarrow \mathbb{R}^+$ si dice *ammissibile*, se sono soddisfatte le seguenti condizioni:

- (1) $f \leq c$.
- (2) $\partial f(a) = 0$ per ogni $a \in P \setminus \{\alpha, \omega\}$.

La prima condizione significa che il flusso non supera la capacità disponibile, la seconda che per ogni $a \in P \setminus \{\alpha, \omega\}$ si ha

$$\text{uscita}(a, f) = \text{ingresso}(a, f)$$

Osservazione 8.11. Esiste sempre un flusso ammissibile in R . Infatti il *flusso nullo* $\bigcirc_s 0$ è ammissibile.

Proposizione 8.12. f sia un flusso ammissibile in R .

$$\text{Allora } \partial f(\omega) = -\partial f(\alpha).$$

Dimostrazione. Dal cor. 8.9 e dalla condizione (2) nella def. 8.10 abbiamo

$$0 = \sum_{a \in P} \partial f(a) = \sum_{a \in P \setminus \{\alpha, \omega\}} \partial f(a) + \partial f(\alpha) + \partial f(\omega) = \partial f(\alpha) + \partial f(\omega)$$

Definizione 8.13. f sia un flusso in R . Allora $\text{val}(f) := \partial f(\omega)$ si chiama il *valore del flusso*.

$\text{val}(f)$ è quindi il flusso locale nella destinazione.

Se f è ammissibile, allora $\text{val}(f) = -\partial f(\alpha)$.

Osservazione 8.14. Spesso la sorgente α e la destinazione ω sono scelte in modo tale che $S(P, \alpha) = \emptyset$ e $S(\omega, P) = \emptyset$.

In questo caso $\text{val}(f) = \text{ingresso}(\omega, f)$ e, se f è ammissibile, si ha anche $\text{val}(f) = \text{uscita}(\alpha, f)$.

Definizione 8.15. Un *flusso massimale* in R è un flusso ammissibile di valore massimale.

Definizione 8.16. Un *taglio* (o una *sezione*, in inglese *cut*) di R è un sottoinsieme $X \subset P$ tale che $\alpha \in X, \omega \notin X$.

$$\text{cap}(X) := c(X, P \setminus X) = \sum_{s \in S(X, P \setminus X)} c(s)$$

si chiama allora la *capacità* di X .

Denotiamo con $\mathcal{T}(R)$ l'insieme dei tagli di R . Siccome per ipotesi $\alpha \neq \omega$, si ha sempre $\mathcal{T}(R) \neq \emptyset$. Ad esempio $\{\alpha\} \in \mathcal{T}(R)$ e $P \setminus \omega \in \mathcal{T}(R)$.

Osservazione 8.17. f sia un flusso ammissibile in R ed $A, B \subset P$. Allora $f(A, B) \leq c(A, B)$.

Lemma 8.18. f sia un flusso ammissibile in R ed X un taglio di R . Allora

$$\text{val}(f) = f(X, P \setminus X) - f(P \setminus X, X) \leq \text{cap}(X)$$

Dimostrazione. Siccome $\alpha \notin P \setminus X$ si ha $\partial f(a) = 0$ per ogni elemento $a \in P \setminus X$ diverso da ω . Perciò

$$\begin{aligned} \text{val}(f) &= \partial f(\omega) = \sum_{a \in P \setminus X} \partial f(a) \stackrel{8.8}{=} f(X, P \setminus X) - f(P \setminus X, X) \\ &\leq f(X, P \setminus X) \stackrel{8.17}{\leq} c(X, P \setminus X) = \text{cap}(X) \end{aligned}$$

Corollario 8.19. f sia un flusso ammissibile in R . Allora

$$\begin{aligned} \text{val}(f) &\leq c(\alpha, P \setminus \alpha) \\ \text{val}(f) &\leq c(P \setminus \omega, \omega) \end{aligned}$$

Corollario 8.20. f sia un flusso ammissibile in R ed X un taglio di R tale che $\text{val}(f) = \text{cap}(X)$. Allora f è un flusso massimale ed X un taglio di capacità minimale.

Corollario 8.21. f sia un flusso ammissibile in R ed X un taglio di R . Assumiamo che

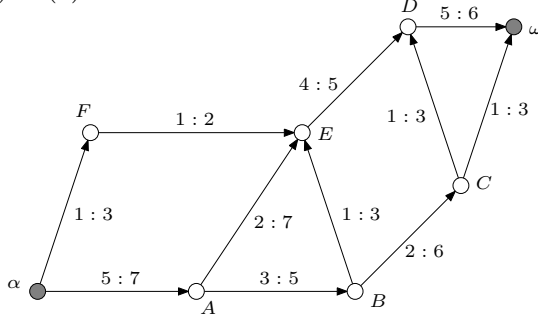
$$f(s) = \begin{cases} c(s) & \text{per } s \in S(X, P \setminus X) \\ 0 & \text{per } s \in S(P \setminus X, X) \end{cases}$$

Allora $\text{val}(f) = \text{cap}(X)$, per cui f è un flusso massimale ed X un taglio di capacità minimale.

Dimostrazione. L'ipotesi implica

$$\begin{aligned} \text{val}(f) &\stackrel{8.18}{=} f(X, P \setminus X) - f(P \setminus X, X) = f(X, P \setminus X) = c(X, P \setminus X) \\ &= \text{cap}(X) \end{aligned}$$

Esempio 8.22. Nella figura è rappresentato un flusso ammissibile con valore 6. Per ogni arco s abbiamo indicato $f(s)$ e $c(s)$ nella forma $f(s) : c(s)$.



In questo esempio $\text{cap}(\alpha) = c(\alpha, P \setminus \alpha) = 10$, $\text{cap}(P \setminus \omega) = c(P \setminus \omega, \omega) = 9$, quindi ogni flusso ammissibile in questa rete possiede un valore ≤ 9 .

Definizione 8.23. Un *cammino* in R è una successione $\gamma = (a_0, a_1, \dots, a_n)$ di punto distinti con $n \geq 1$ tale che $(a_{i-1}, a_i) \in S$ per ogni $i = 1, \dots, n$.

γ si chiama allora un cammino da a_0 ad a_n .

Per $a, b \in P$ sia $\mathcal{C}(a, b)$ l'insieme dei cammini da a a b .

Proposizione 8.24. X sia un taglio di R e $\gamma = (a_0, \dots, a_n) \in \mathcal{C}(\alpha, \omega)$. Allora esiste un $i \in \{1, \dots, n\}$ tale che $(a_{i-1}, a_i) \in S(X, P \setminus X)$.

Dimostrazione. Siccome $a_0 = \alpha \in X$ ed $a_n = \omega \in P \setminus X$, deve esistere un più piccolo $i \geq 1$ tale che $a_i \in P \setminus X$. Allora $(a_{i-1}, a_i) \in S(X, P \setminus X)$.

Osservazione 8.25. f sia un flusso ammissibile in R e $\gamma = (a_0, \dots, a_n) \in \mathcal{C}(\alpha, \omega)$. Per l'ipotesi (1) della def. 8.10 allora $c(a_{i-1}, a_i) - f(a_{i-1}, a_i) \geq 0$ per ogni $i = 1, \dots, n$.

Sia $\Delta := \Delta(\gamma, f) := \min_{1 \leq i \leq n} \{c(a_{i-1}, a_i) - f(a_{i-1}, a_i)\}$.

Se allora definiamo un flusso $g : S \rightarrow \mathbb{R}^+$ ponendo

$$g(s) := \begin{cases} f(s) + \Delta & \text{se } s = (a_{i-1}, a_i) \text{ per un } i \in \{1, \dots, n\} \\ f(s) & \text{altrimenti} \end{cases} \quad (*)$$

allora g è un flusso ammissibile in R con $\text{val}(g) = \text{val}(f) + \Delta$.

Dimostrazione. (1) È chiaro che $g \leq c$.

(2) Sia $b \in P \setminus \{\alpha, \omega\}$. Se $b \notin \{a_0, \dots, a_n\}$, allora $g(P, b) = f(P, b)$ e $g(b, P) = f(b, P)$, per cui $\partial g(b) = \partial f(b) = 0$.

Sia $b \in \{a_0, \dots, a_n\}$. Siccome $b \neq \alpha, \omega$, deve esistere un $i \in \{1, \dots, n-1\}$ tale che $b = a_i$. Siccome i punti che appaiono in γ sono tutti distinti, i è univocamente determinato e gli unici archi in $S(P, b)$ risp. $S(b, P)$ in cui avviene l'aumento in (*) sono (a_{i-1}, a_i) e (a_i, a_{i+1}) . Ciò implica

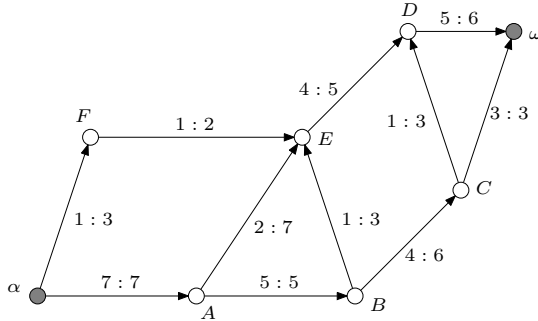
$$\text{ingresso}(b, g) = \text{ingresso}(b, f) + \Delta$$

$$\text{uscita}(b, g) = \text{uscita}(b, f) + \Delta$$

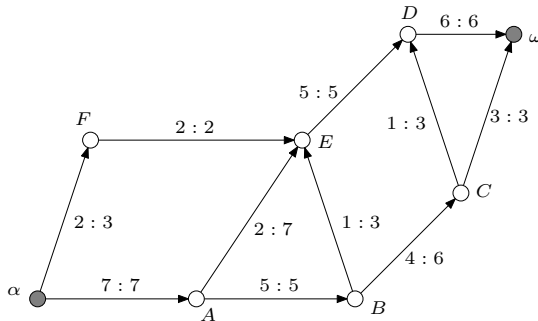
per cui $\partial g(b) = \partial f(b) = 0$.

(3) Nello stesso modo si vede che $\text{ingresso}(\omega, g) = \text{ingresso}(\omega, f) + \Delta$ mentre $\text{uscita}(\omega, g) = \text{uscita}(\omega, f)$, perché ω non può apparire all'interno di γ . Ciò implica $\text{val}(g) = \partial g(\omega) = \partial f(\omega) + \Delta = \text{val}(f) + \Delta$.

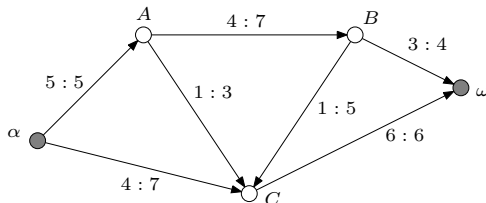
Esempio 8.26. Nella rete dell'esempio 8.22 possiamo prima considerare il cammino $(\alpha, A, B, C, \omega)$ in cui $\Delta = 2$, ottenendo il flusso



di valore 8, e successivamente il cammino $(\alpha, F, E, D, \omega)$ in cui $\Delta = 1$, ottenendo così un flusso di valore 9 il quale, come abbiamo visto nell'esempio 8.22, è già un flusso massimale:



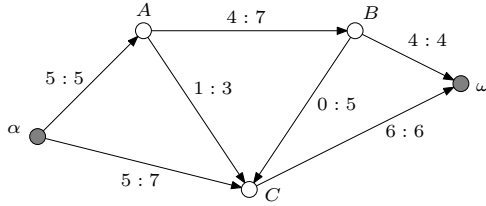
Nota 8.27. Può accadere però che la tecnica dell'oss. 8.25 non sia applicabile, cioè può accadere che $\Delta(\gamma, f) = 0$ per ogni $\gamma \in \mathcal{C}(\alpha, \omega)$, pur non essendo f massimale. Consideriamo ad esempio il flusso f descritto dalla figura



Calcoliamo $\Delta(\gamma, f)$ per ogni $\gamma \in \mathcal{C}(\alpha, \omega)$:

γ	$\Delta(\gamma, f)$
(α, A, B, w)	0
(α, A, B, C, w)	0
(α, A, C, w)	0
(α, C, w)	0

Quindi non possiamo aumentare il valore del flusso con il metodo dell'oss. 8.25. Se però diminuiamo di 1 il carico sull'arco (B, C) , possiamo aumentare di 1 il carico sugli archi (B, ω) e (α, C) , ottenendo il flusso ammissibile



Per il cor. 8.19 questo è un flusso ammissibile di valore massimale.

Definizione 8.28. Un *semicammino* in R è una successione $\gamma = (a_0, \varepsilon_1, a_1, \varepsilon_2, \dots, a_{n-1}, \varepsilon_n, a_n)$ tale che:

- (1) $n \geq 1$.
- (2) a_0, \dots, a_n sono punti distinti di P .
- (3) $\varepsilon_1, \dots, \varepsilon_n \in \{1, -1\}$.
- (4) Per ogni $i = 1, \dots, n$ vale:
 $\varepsilon_i = 1 \implies (a_{i-1}, a_i) \in S$
 $\varepsilon_i = -1 \implies (a_i, a_{i-1}) \in S$

γ si chiama allora un *semicammino* da a_0 ad a_n .

Per $a, b \in P$ sia $S(a, b)$ l'insieme dei *semicammini* da a a b .

Poniamo $\text{vertici}(\gamma) := \{a_0, \dots, a_n\}$.

Definizione 8.29. f sia un flusso ammissibile in R e

$\gamma = (a_0, \varepsilon_1, a_1, \dots, a_{n-1}, \varepsilon_n, a_n)$ un *semicammino* in R . Diciamo che γ è *f-aumentante* se per ogni $i = 1, \dots, n$ si ha

$$\varepsilon_i = 1 \implies f(a_{i-1}, a_i) < c(a_{i-1}, a_i)$$

$$\varepsilon_i = -1 \implies f(a_i, a_{i-1}) > 0$$

Per $a, b \in P$ denotiamo con $S_f(a, b)$ l'insieme dei *semicammini f-aumentanti* da a a b .

Esempio 8.30. Nella prima figura dell'oss. 8.27 il *semicammino* $(\alpha, 1, C, -1, B, 1, \omega)$ da α a ω è *f-aumentante*. Infatti per $\varepsilon_1 = 1$ si ha $f(\alpha, C) < c(\alpha, C)$ perché $4 < 7$; per $\varepsilon_2 = -1$ si ha $f(B, C) = 1 > 0$ ed infine per $\varepsilon_3 = 1$ si ha $f(B, \omega) < c(B, \omega)$ perché $3 < 4$.

Definizione 8.31. $\gamma = (a_0, \varepsilon_1, a_1, \dots, a_{n-1}, \varepsilon_n, a_n)$ sia un *semicammino* in R . Poniamo allora

$$\begin{aligned}
S_\gamma^+ &:= \{(a_{i-1}, a_i) \mid i = 1, \dots, n \text{ ed } \varepsilon_i = 1\} \\
S_\gamma^- &:= \{(a_i, a_{i-1}) \mid i = 1, \dots, n \text{ ed } \varepsilon_i = -1\} \\
S_\gamma &:= S_\gamma^+ \cup S_\gamma^-
\end{aligned}$$

È chiaro che $S_\gamma^+ \cap S_\gamma^- = \emptyset$.

Infatti sia $(a_{i-1}, a_i) = (a_j, a_{j-1})$, ovvero $a_{i-1} = a_j$ e $a_i = a_{j-1}$. Siccome i punti a_0, \dots, a_n sono distinti, ciò implica $i - 1 = j$ e $i = j - 1$, e ciò è impossibile.

Definizione 8.32. f sia un flusso ammissibile in R e

$\gamma = (a_0, \varepsilon_1, a_1, \dots, a_{n-1}, \varepsilon_n, a_n)$ un semicammino in R .

Per $s \in S_\gamma$ poniamo

$$\begin{aligned}
\Delta(\gamma, f, s) &:= c(a_{i-1}, a_i) - f(a_{i-1}, a_i) \quad \text{se } (a_{i-1}, a_i) \in S_\gamma^+ \\
\Delta(\gamma, f, s) &:= f(a_i, a_{i-1}) \quad \text{se } (a_i, a_{i-1}) \in S_\gamma^-
\end{aligned}$$

Definiamo adesso

$$\Delta(\gamma, f) := \min_{s \in S_\gamma} \Delta(\gamma, f, s)$$

Esempio 8.33. Nella prima figura della nota 8.27 consideriamo ancora il semicammino $\gamma = (\alpha, 1, C, -1, B, 1, \omega)$. Allora

$$\Delta(\gamma, f, (\alpha, C)) = c(\alpha, C) - f(\alpha, C) = 7 - 4 = 3$$

$$\Delta(\gamma, f, (B, C)) = f(B, C) = 1$$

$$\Delta(\gamma, f, (B, \omega)) = c(B, \omega) - f(B, \omega) = 4 - 3 = 1$$

per cui $\Delta(\gamma, f) = 1$.

Osservazione 8.34. Nella situazione della def. 8.29

γ è f -aumentante se e solo se $\Delta(\gamma, f) > 0$.

Proposizione 8.35. f sia un flusso ammissibile in R e $\gamma \in \mathcal{S}(\alpha, \omega)$.

Sia $\Delta := \Delta(\gamma, f)$. Se allora definiamo un flusso $g : S \rightarrow \mathbb{R}^+$ ponendo

$$g(s) := \begin{cases} f(s) + \Delta & \text{per } s \in S_\gamma^+ \\ f(s) - \Delta & \text{per } s \in S_\gamma^- \\ f(s) & \text{per } s \in S \setminus S_\gamma \end{cases} \quad (*)$$

allora g è un flusso ammissibile in R e $\text{val}(g) = \text{val}(f) + \Delta$.

Dimostrazione. (1) È chiaro che $0 \leq g \leq c$.

(2) Sia $\gamma = (a_0, \varepsilon_1, a_1, \dots, a_{n-1}, \varepsilon_n, a_n)$. Sia $b \in P \setminus \{\alpha, \omega\}$.

Se $b \notin \{a_0, \dots, a_n\}$, allora $g(b, P) = f(b, P)$ e $g(P, b) = f(P, b)$ per cui, essendo f un flusso ammissibile, si ha $\partial g(b) = \partial f(b) = 0$.

Supponiamo ora che $b \in \{a_0, \dots, a_n\}$. Siccome $b \neq \alpha, \omega$, deve esistere un $i \in \{1, \dots, n-1\}$ tale che $b = a_i$. Siccome i punti che appaiono in γ sono tutti distinti, i è univocamente determinato e gli aumenti in (*) avvengono solo su semicammini di una delle quattro forme seguenti per le quali abbiamo calcolato i cambiamenti:

$$\begin{array}{ll}
(a_{i-1}, -1, b, -1, a_{i+1}) & \dots \text{ ingresso}(b, g) = \text{ingresso}(b, f) - \Delta \\
& \dots \text{ uscita}(b, g) = \text{uscita}(b, f) - \Delta \\
(a_{i-1}, -1, b, 1, a_{i+1}) & \dots \text{ ingresso}(b, g) = \text{ingresso}(b, f) \\
& \dots \text{ uscita}(b, g) = \text{uscita}(b, f) - \Delta + \Delta \\
(a_{i-1}, 1, b, -1, a_{i+1}) & \dots \text{ ingresso}(b, g) = \text{ingresso}(b, f) + \Delta - \Delta \\
& \dots \text{ uscita}(b, g) = \text{uscita}(b, f) \\
(a_{i-1}, 1, b, 1, a_{i+1}) & \dots \text{ ingresso}(b, g) = \text{ingresso}(b, f) + \Delta \\
& \dots \text{ uscita}(b, g) = \text{uscita}(b, f) + \Delta
\end{array}$$

In tutti questi casi risulta $\partial g(b) = \partial f(b) = 0$.

(3) Per calcolare $\text{val}(g)$ consideriamo i due casi possibili $\varepsilon_n = 1$ e $\varepsilon_n = -1$:

$$\begin{array}{ll}
\varepsilon_n = 1 & \dots \text{ ingresso}(\omega, g) = \text{ingresso}(\omega, f) + \Delta \\
& \dots \text{ uscita}(\omega, g) = \text{uscita}(\omega, f) \\
\varepsilon_n = -1 & \dots \text{ ingresso}(\omega, g) = \text{ingresso}(\omega, f) \\
& \dots \text{ uscita}(\omega, g) = \text{uscita}(\omega, f) - \Delta
\end{array}$$

perché ω non può apparire all'interno di γ . Ciò implica $\text{val}(g) = \partial g(\omega) = \partial f(\omega) + \Delta = \text{val}(f) + \Delta$.

Corollario 8.36. *f* sia un flusso ammissibile in R . Se esiste un semicammino f -aumentante da α ad ω , allora f non è un flusso massimale.

Osservazione 8.37. *f* sia un flusso ammissibile in R tale che $\mathcal{S}_f(\alpha, \omega) = \emptyset$. Siano

$$\begin{aligned}
\Gamma &:= \{\gamma \in \mathcal{S}_f(\alpha, a) \mid a \in P\} \\
X &:= \{\alpha\} \cup \bigcup_{\gamma \in \Gamma} \text{vertici}(\gamma)
\end{aligned}$$

L'ipotesi implica che $\omega \notin X$, perciò X è un taglio di R .

Se $\Gamma \neq \emptyset$, si ha naturalmente $X = \bigcup_{\gamma \in \Gamma} \text{vertici}(\gamma)$.

Teorema 8.38. *f* sia un flusso ammissibile in R . Allora sono equivalenti:

- (1) *f* è massimale.
- (2) Non esiste un semicammino f -aumentante da α a ω .
- (3) Esiste un taglio X di R di capacità minimale tale che $\text{val}(f) = \text{cap}(X)$.
- (4) Esiste un taglio X di R tale che $\text{val}(f) = \text{cap}(X)$.

Dimostrazione. (1) \implies (2): Corollario 8.36.

(2) \implies (3): Sia $\mathcal{S}_f(\alpha, \omega) = \emptyset$. Definiamo il taglio X come nell'oss. 8.37. Per costruzione allora

$$s \in S(X, P \setminus X) \implies f(s) = c(s)$$

$$s \in (P \setminus X, X) \implies f(s) = 0$$

Per il cor. 8.21 $\text{val}(f) = \text{cap}(X)$ ed X è un taglio di capacità minimale.

(3) \implies (4): Chiaro.

(4) \implies (1): Corollario 8.20.

Teorema 8.39. *f* sia un flusso massimale in R ed Y un taglio di capacità minimale di R . Allora $\text{val}(f) = \text{cap}(Y)$.

Dimostrazione. (1) Dal lemma 8.18 sappiamo che $\text{val}(f) \leq \text{cap}(Y)$.

(2) Per il teorema 8.38 esiste un taglio X di R di capacità minimale tale che $\text{val}(f) = \text{cap}(X)$. Siccome anche Y è di capacità minimale, abbiamo $\text{cap}(X) = \text{cap}(Y)$.

Proposizione 8.40. Per ogni $s \in S$ sia $c(s) \in \mathbb{N}$. Allora:

(1) Esiste sempre un taglio di capacità minimale di R .

(2) Esiste sempre un flusso massimale f in R tale che $f(s) \in \mathbb{N}$ per ogni $s \in S$.

Dimostrazione. (1) Chiaro, perché P è un insieme finito.

(2) Iniziando con il flusso nullo $f_0 := \bigcirc_s 0$, per ogni $i = 0, 1, \dots$ procediamo nel modo seguente:

Se $\mathcal{S}_{f_i}(\alpha, \omega) = \emptyset$, ci fermiamo.

Altrimenti sia $\gamma \in \mathcal{S}_{f_i}(\alpha, \omega)$. È chiaro che $\Delta(\gamma, f_i) \in \mathbb{N} + 1$. Per la prop. 8.35 troviamo un flusso ammissibile f_{i+1} con

$$\text{val}(f_{i+1}) = \text{val}(f_i) + \Delta(\gamma, f_i) \geq \text{val}(f_i) + 1$$

Per costruzione inoltre $f_{i+1}(s) \in \mathbb{N}$ per ogni $s \in S$. Siccome $\text{val}(f_{i+1}) \leq c(P \setminus \omega, \omega)$ per il cor. 8.19, l'algoritmo deve terminare.

Nota 8.41. Seguendo Aigner, pagg. 305-307, facciamo vedere adesso come il problema dei flussi massimali possa essere risolto nell'ambito della programmazione lineare.

Definiamo una matrice $M \in \mathbb{R}_S^P$ ponendo, per $p \in P$ ed $s \in S$,

$$M_s^p := \begin{cases} 1 & \text{se } p = s^2 \\ -1 & \text{se } p = s^1 \\ 0 & \text{altrimenti} \end{cases}$$

Senza perdere di generalità possiamo assumere che $(\omega, \alpha) \in S$ e che $c(\omega, \alpha) = \infty$ (oppure che $c(\omega, \alpha) > \sum_{s \in S \setminus \{\omega, \alpha\}} c(s)$).

Numerando $P \in S$, possiamo identificare P con $\{1, \dots, m\}$ ed S con $\{1, \dots, n\}$. Allora $M \in \mathbb{R}_n^m$.

Per la rete della nota 8.27, aggiungendo l'arco (ω, α) , M è data dalla tabella

	αA	αC	AB	AC	BC	$B\omega$	$C\omega$	$\omega\alpha$
α	-1	-1	0	0	0	0	0	1
A	1	0	-1	-1	0	0	0	0
B	0	0	1	0	-1	-1	0	0
C	0	1	0	1	1	0	-1	0
ω	0	0	0	0	0	1	1	-1

La capacità c può essere identificata con un vettore $c \in \mathbb{R}_+^n$, e similmente un flusso ammissibile f con un vettore $x \in \mathbb{R}_+^n$ tale che $x \leq c$ con $Mx = 0$. Quest'ultima equazione esprime proprio la condizione (2) della def. 8.10. Il valore del flusso è uguale alla componente $x^{(\omega, \alpha)}$.

Il problema del flusso massimale è quindi equivalente al problema di ottimizzazione lineare

$$Mx = 0$$

$$0 \leq x \leq c \quad (*)$$

$$(0 \dots 01)x = \max$$

Usando il comando `glpk` di Matlab per il nostro esempio troviamo così una soluzione con le istruzioni:

```
M=[-1 -1 0 0 0 0 0 1; 1 0 -1 -1 0 0 0 0;
0 0 1 0 -1 -1 0 0; 0 1 0 1 1 0 -1 0; 0 0 0 0 0 1 1 -1];
A=[M; eye(8)];
b=[0 0 0 0 0 5 7 7 3 5 4 6 Inf]';
f=[0 0 0 0 0 0 0 0 1];
[x,m]=glpk(f',M,b,[],[],'SSSSUUUUUUU','CCCCCCC',-1)
```

Usando il teorema 2.9 con questo ragionamento possiamo dimostrare l'esistenza di un flusso massimale anche senza la condizione che $c(s) \in \mathbb{N}$ per ogni $s \in S$ (cfr. prop. 8.40). Infatti con

$$A := \begin{pmatrix} M \\ -M \\ \delta \end{pmatrix}, \quad b := \begin{pmatrix} 0 \\ 0 \\ c \end{pmatrix}, \quad f := (0 \dots 01)$$

il problema (*) può essere scritto nella forma

$$fx = \max$$

$$x \geq 0$$

$$Ax \leq b$$

Siccome $c \geq 0$ e quindi anche $b \geq 0$, abbiamo $X := (A\mathbb{R}_+^n \leq b) \neq \emptyset$ perché $0 \in X$ e $Y := (\mathbb{R}_m^+ A \geq f) \neq \emptyset$ perché $(0 \quad f) \in Y$. Dal teorema 2.9 segue che il nostro problema possiede una soluzione.

III. OTTIMIZZAZIONE NON LINEARE

9. Il cono tangente

Situazione 9.1. X sia un sottoinsieme di \mathbb{R}^n , Ω un aperto di \mathbb{R}^n tale che $X \subset \Omega$ ed $f : \Omega \rightarrow \mathbb{R}$ una funzione di classe C^1 .

In questa parte della tesi seguiamo soprattutto Geiger/Kanzow, pagg. 41-76.

Nota 9.2. Usiamo le seguenti abbreviazioni:

$$(1) [\mathbb{R}^{++} \downarrow 0] := \{\bigcirc_k t_k \in \mathbb{R}^{\mathbb{N}} \mid t_0 > t_1 > t_2 > \dots > 0 \text{ e } \lim_{k \rightarrow \infty} t_k = 0\}$$

(2) Per $y \in \mathbb{R}^n$ sia

$$[X \rightarrow y] := \{\bigcirc_k x_k \in X^{\mathbb{N}} \mid \lim_{k \rightarrow \infty} x_k = y\}$$

Ponendo $X = \mathbb{R}^n$ e $X = \mathbb{R}^{++}$ abbiamo in particolare definito gli insiemi $[\mathbb{R}^n \rightarrow y]$ e $[\mathbb{R}^{++} \rightarrow 0]$.

Osservazione 9.3. Sia $\bigcirc_k t_k \in [\mathbb{R}^{++} \rightarrow 0]$. Ponendo $n_0 := 0$ troviamo un indice $n_1 > n_0$ tale che $t_{n_1} < t_{n_0}$, poi un indice $n_2 > n_1$ tale che $t_{n_2} < t_{n_1}$ ecc. In questo modo otteniamo una successione

$$\bigcirc_j t_{n_j} \in [\mathbb{R}^{++} \downarrow 0].$$

Definizione 9.4. Sia $x \in X$. Un vettore $v \in \mathbb{R}^n$ si dice *tangente* ad X in x , se esistono due successioni $\bigcirc_k x_k \in [X \rightarrow x]$ e $\bigcirc_k t_k \in [\mathbb{R}^{++} \rightarrow 0]$

tali che $\bigcirc_k \frac{x_k - x}{t_k} \in [\mathbb{R}^n \rightarrow v]$, cioè tali che $\lim_{k \rightarrow \infty} \frac{x_k - x}{t_k} = v$.

L'insieme $T(X, x)$ di tutti i vettori tangenti ad X in x si chiama il *cono tangente* di X in x .

Osservazione 9.5. Siano $x \in X$ e $v \in \mathbb{R}^n$. Usando l'oss. 9.3 vediamo che $v \in T(X, x)$ se e solo se esistono due successioni $\bigcirc_k x_k \in [X \rightarrow x]$

e $\bigcirc_k t_k \in [\mathbb{R}^{++} \downarrow 0]$ tali che $\bigcirc_k \frac{x_k - x}{t_k} \in [\mathbb{R}^n \rightarrow v]$.

Osservazione 9.6. Sia $x \in X$. Allora $T(X, x)$ è un cono di \mathbb{R}^n (nel senso della def. 5.7).

Dimostrazione. È chiaro che $0 \in T(X, x)$, per cui $T(X, x) \neq \emptyset$.

Se $v \in T(X, x)$ e $\lambda > 0$, e se $\bigcirc_k x_k \in [X \rightarrow x]$ e $\bigcirc_k t_k \in [\mathbb{R}^{++} \rightarrow 0]$ sono tali che $\lim_{k \rightarrow \infty} \frac{x_k - x}{t_k} = v$, allora evidentemente

$\bigcirc_k t_k/\lambda \in [\mathbb{R}^{++} \rightarrow 0]$ e $\lim_{k \rightarrow \infty} \frac{x_k - x}{t_k/\lambda} = \lambda v$, per cui $\lambda v \in T(X, x)$.

Osservazione 9.7. x sia un punto interno di X (rispetto alla topologia di \mathbb{R}^n). Allora $T(X, x) = \mathbb{R}^n$.

Dimostrazione. Sia $v \in \mathbb{R}^n$. Dobbiamo far vedere che v è tangente ad X in x . Siccome x è un punto interno di X , esiste un $m \in \mathbb{N} + 1$ tale che $x_k = x + \frac{1}{m+k}v \in X$ per ogni $k \in \mathbb{N}$. Poniamo $t_k := \frac{1}{m+k}$. Allora $\bigcirc_k t_k \in [\mathbb{R}^{++} \rightarrow 0]$, $\bigcirc_k x_k \in [X \rightarrow x]$, mentre $\frac{x_k - x}{t_k} = v$ per ogni $k \in \mathbb{N}$, per cui sicuramente $\bigcirc_k \frac{x_k - x}{t_k} \in [\mathbb{R}^n \rightarrow v]$.

Lemma 9.8. Sia $x \in X$. Allora $T(X, x)$ è un sottoinsieme chiuso di \mathbb{R}^n .

Dimostrazione. Siano $v \in T(X, x)$ e $\bigcirc_j v_j \in [T(X, x) \rightarrow v]$. Allora per ogni $j \in \mathbb{N}$ esistono due successioni $\bigcirc_k x_k^{(j)} \in [X \rightarrow x]$ e $\bigcirc_k t_k^{(j)} \in [\mathbb{R}^{++} \downarrow 0]$ tali che $\lim_{k \rightarrow \infty} \frac{x_k^{(j)} - x}{t_k^{(j)}} = v_j$.

Ciò implica che possiamo trovare una successione $\bigcirc_j m_j \in \mathbb{N}^{\mathbb{N}}$ tale che per ogni $j \in \mathbb{N} + 1$ si abbia $|x_{m_j}^{(j)} - x| \leq \frac{1}{j}$, $0 < t_{m_j}^{(j)} \leq \frac{1}{j}$ e

$\left| \frac{x_{m_j}^{(j)} - x}{t_{m_j}^{(j)}} - v_j \right| \leq \frac{1}{j}$. L'ultima disuguaglianza implica

$$\left| \frac{x_{m_j}^{(j)} - x}{t_{m_j}^{(j)}} - v \right| \leq \left| \frac{x_{m_j}^{(j)} - x}{t_{m_j}^{(j)}} - v_j \right| + |v_j - v|$$

e vediamo che $\bigcirc_j x_{m_j}^{(j)} \in [X \rightarrow x]$, $\bigcirc_j t_{m_j}^{(j)} \in [\mathbb{R}^{++} \rightarrow 0]$ e

$$\lim_{j \rightarrow \infty} \frac{x_{m_j}^{(j)} - x}{t_{m_j}^{(j)}} = v, \text{ cosicché } v \in T(X, x).$$

Definizione 9.9. $x \in X$ è un punto di minimo locale di f in X , se esiste un intorno $U \in \mathcal{U}(x, \text{in } X)$ tale che

$$f(x) \leq f(u) \quad \text{per ogni } u \in U$$

Per la definizione di topologia in \mathbb{R}^n (e quindi su X) ciò è equivalente alla condizione che esista $r > 0$ tale che $f(x) \leq f(u)$ ogni $u \in X$ per cui $|x - u| < r$.

Definizione 9.10. Per $x \in \Omega$ e $v \in \mathbb{R}^n$ sia

$$f'(x; v) := \lim_{t \rightarrow 0} \frac{f(x + tv) - f(x)}{t}$$

la derivata in direzione v di f in x .

$$\text{È noto dall'analisi che } f'(x; v) = \|\nabla f(x), v\|.$$

Osservazione 9.11. x sia un punto interno di X .

Allora sono equivalenti:

- (1) $f'(x; v) \geq 0$ per ogni $v \in T(X, x)$.
- (2) $f'(x; v) = 0$ per ogni $v \in T(X, x)$.
- (3) $\nabla f(x) = 0$.

Dimostrazione. Le implicazioni (3) \implies (2) \implies (1) sono banali.

L'implicazione (1) \implies (3) segue dall'oss. 9.7.

Lemma 9.12. *x sia un punto di minimo locale di f in X .*

Allora $f'(x; v) \geq 0$ per ogni $v \in T(X, x)$.

Dimostrazione. Sia $v \in T(X, x)$. Allora esistono successioni

$\bigcirc_k x_k \in [X \rightarrow x]$ e $\bigcirc_k t_k \in [\mathbb{R}^{++} \rightarrow 0]$ tali che $\bigcirc_k \frac{x_k - x}{t_k} \in [\mathbb{R}^n \rightarrow v]$.

Per il teorema del valor medio esiste una successione $\bigcirc_k \xi_k$ tale che $\xi_k \in [x, x_k]$ e $f(x_k) - f(x) = \|\nabla f(\xi_k), x_k - x\|$ per ogni k .

L'ipotesi che x sia un punto di minimo locale di f in X implica che $f(x_k) - f(x) \geq 0$ per $k \gg 0$ e quindi anche $\|\nabla f(\xi_k), \frac{x_k - x}{t_k}\| \geq 0$ per $k \gg 0$. Inoltre $\bigcirc_k \xi_k \rightarrow x$ perché $\xi_k \in [x, x_k]$ per ogni k . Ciò implica

$$f'(x; v) = \|\nabla f(x), v\| = \lim_{k \rightarrow \infty} \|\nabla f(\xi_k), \frac{x_k - x}{t_k}\| \geq 0$$

Esempio 9.13. Consideriamo il caso $n = 1$, $X = [0, 1]$ e Ω un intervallo aperto che contiene $[0, 1]$. Il cono tangente di X coincide con $[0, \infty)$, il cono tangente in 1 con $(-\infty, 0]$.

Il lemma 9.12 afferma che se 0 è un punto di minimo locale di f in X , allora $f'(0) \geq 0$, e se 1 è un punto di minimo locale in X , allora $f'(1) \leq 0$.

Osservazione 9.14. Il cono tangente $T(X, x)$ in generale è piuttosto complicato, cosicché spesso non è possibile applicare direttamente il lemma 9.12. Descriveremo adesso, in alcuni casi importanti, condizioni di ottimalità più pratiche.

10. Le condizioni di Karush-Kuhn-Tucker

Situazione 10.1. Siano date funzioni $f, g_1, \dots, g_m, h_1, \dots, h_p : \mathbb{R}^n \rightarrow \mathbb{R}$ tutte di classe C^1 . Sono ammessi i casi $m = 0$ o $p = 0$.

Sia $X := (g_1 \leq 0, \dots, g_m \leq 0, h_1 = 0, \dots, h_p = 0)$.

Poniamo anche $g := (g_1, \dots, g_m), h := (h_1, \dots, h_p)$.

Il risultato principale di questo capitolo è il teorema 10.11 che, come conseguenza del lemma di Farkas, ci fornisce una condizione necessaria e facilmente verificabile per un minimo locale di f in X .

Definizione 10.2. Per $x \in X$ sia $I(x) := \{i \in \{1, \dots, m\} \mid g_i(x) = 0\}$.

Definizione 10.3. Per $x \in X$ sia

$$\begin{aligned} \text{Tl}(x) := \{v \in \mathbb{R}^n \mid & \|\nabla g_i(x), v\| \leq 0 \quad \text{per ogni } i \in I(x), \\ & \|\nabla h_j(x), v\| = 0 \quad \text{per ogni } j \in \{1, \dots, p\}\} \end{aligned}$$

$\text{Tl}(x)$ si chiama il *cono tangente linearizzato* di X in x .

Si noti che esso nel caso generale dipende non solo da X , ma anche dalla scelta delle restrizioni g_i ed h_j che usiamo per descrivere X .

È immediato che $\text{Tl}(x)$ è un cono nel senso della def. 5.7.

Proposizione 10.4. Sia $x \in X$. Allora $\text{T}(X, x) \subset \text{Tl}(x)$.

Dimostrazione. La dimostrazione è molto simile alla dimostrazione del lemma 9.12.

Sia $v \in \text{T}(X, x)$. Allora esistono successioni $\bigcirc_k x_k \in [X \rightarrow x]$ e

$$\bigcirc_k t_k \in [\mathbb{R}^{++} \rightarrow 0] \text{ tali che } \bigcirc_k \frac{x_k - x}{t_k} \in [\mathbb{R}^n \rightarrow v].$$

(1) Sia $i \in I(x)$. Per il teorema del valor medio esiste una successione $\bigcirc_k \xi_k$ tale che $\xi_k \in [x, x_k]$ e $g_i(x_k) = g_i(x) + \|\nabla g_i(\xi_k), x_k - x\|$ per ogni k . Per ipotesi però $g_i(x) = 0$ perché $i \in I(x)$ e $g_i(x_k) \leq 0$ perché ogni $x_k \in X$. Perciò $\|\nabla g_i(\xi_k), x_k - x\| \leq 0$ e quindi anche $\|\nabla g_i(\xi_k), \frac{x_k - x}{t_k}\| \leq 0$ per ogni k . Passando al limite otteniamo $\|\nabla g_i(x), v\| \leq 0$.

(2) Nello stesso modo si dimostra che $\|\nabla h_j(x), v\| = 0$ per ogni $j = 1, \dots, p$.

Definizione 10.5. Un punto $x \in X$ si dice *regolare nel senso di Abadie*, se $\text{T}(X, x) = \text{Tl}(x)$.

Esempio 10.6. Troveremo fra poco semplici condizioni sufficienti affinché un punto sia regolare nel senso di Abadie. Non è comunque sempre così, come mostra il seguente esempio.

$$\begin{aligned} \text{Sia } X := & \{x \in \mathbb{R}^2 \mid 0 \leq x^2 \leq (x^1)^3\} \\ & = \{x \in \mathbb{R}^2 \mid -x^2 \leq 0, x^2 - (x^1)^3 \leq 0\} \end{aligned}$$

X è quindi l'insieme dei punti nel primo quadrante che si trovano al di sotto della parabola cubica data dall'equazione $x^2 = (x^1)^3$.

Sia $x_0 := \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. È chiaro che $T(X, x) = \mathbb{R}^+ \times 0$.

Con $g_1(x) := -x^2$ e $g_2(x) := x^2 - (x^1)^3$ troviamo invece $\nabla g_1(x) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$, da cui $\nabla g_1(x_0) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$, e $\nabla g_2(x) = \begin{pmatrix} -3(x^1)^2 \\ 1 \end{pmatrix}$ da cui $\nabla g_2(x_0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

Siccome $g_1(x_0) = g_2(x_0) = 0$, abbiamo $I(x_0) = \{1, 2\}$, cosicché per $v \in \mathbb{R}^2$ si ha $v \in Tl(x_0) \iff -v^2 \leq 0$ e $v^2 \leq 0 \iff v^2 = 0$, per cui $Tl(x_0) = \mathbb{R} \times 0$.

$T(X, x_0)$ non coincide perciò con $Tl(x_0)$, cosicché x_0 non è un punto regolare nel senso di Abadie.

Definizione 10.7. Definiamo un'applicazione $L : \mathbb{R}^n \times \mathbb{R}_m \times \mathbb{R}_p \longrightarrow \mathbb{R}$ tramite

$$\begin{aligned} L(x, \lambda, \mu) &:= f(x) + \sum_{i=1}^m \lambda_i g_i(x) + \sum_{j=1}^p \mu_j h_j(x) \\ &= f(x) + \|\lambda, g(x)\| + \|\mu, h(x)\| \end{aligned}$$

L'applicazione L si chiama la *funzione di Lagrange* di (X, f) (o più precisamente della situazione descritta dalle funzioni f, g, h).

Definizione 10.8. Le condizioni di Karush-Kuhn-Tucker (KKT) rispetto alla situazione 10.1 sono

$$\begin{aligned} \nabla_x L(x, \lambda, \mu) &= 0 \\ g(x) &\leq 0 \\ h(x) &= 0 \\ \lambda &\geq 0 \\ \|\lambda, g(x)\| &= 0 \end{aligned}$$

Il gradiente ∇_x rispetto alla variabile x è definito da

$$\nabla_x L(x, \lambda, \mu) := \nabla f(x) + \sum_{i=1}^m \lambda_i \nabla g_i(x) + \sum_{j=1}^p \mu_j \nabla h_j(x)$$

I parametri $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_p$ (oppure anche i vettori λ, μ) vengono detti *moltiplicatori di Lagrange*.

Un punto (x, λ, μ) , in cui le condizioni di KKT sono soddisfatte, si chiama un *punto di KKT* di (X, f) (o più precisamente della situazione descritta dalle funzioni f, g, h).

Si noti che la seconda e la terza condizione implicano che $x \in X$.

Osservazione 10.9. Se $m = p = 0$, le condizioni KKT si riducono alla condizione $\nabla f(x) = 0$.

Osservazione 10.10. (x, λ, μ) sia un punto di KKT di (X, f) . Allora $\lambda_i g_i(x) = 0$ per ogni $i = 1, \dots, m$. Quindi per ogni $i = 1, \dots, m$ almeno uno dei due numeri λ_i e $g_i(x)$ deve essere uguale a zero.

Se inoltre $\lambda_i + g_i(x) \neq 0$ per ogni $i = 1, \dots, m$ (cioè se λ_i e $g_i(x)$ non si annullano simultaneamente per nessun indice i), allora (x, λ, μ) si dice un *punto di KKT stretto* di (X, f) .

Teorema 10.11. *Sia $x \in X$ un punto regolare nel senso di Abadie. x sia un punto di minimo locale di f in X .*

Allora esistono $\lambda \in \mathbb{R}_m$ e $\mu \in \mathbb{R}_p$ tali che (x, λ, μ) sia un punto di KKT di (X, f) .

Dimostrazione. Per il lemma 9.12 abbiamo $f'(x; v) \geq 0$ per ogni $v \in T(X, x)$ e quindi, per ipotesi, per ogni $v \in Tl(x)$.

Sia $I(x) = \{i_1, \dots, i_k\}$ con $i_1 < \dots < i_k$, quindi $k = |I(x)|$ (è ammesso il caso $I(x) = \emptyset$ e quindi $k = 0$).

Consideriamo la matrice

$$A := (-\nabla g_{i_1}(x), \dots, -\nabla g_{i_k}(x), \nabla h_1(x), -\nabla h_1(x), \dots, \nabla h_p(x), -\nabla h_p(x))$$

che appartiene a \mathbb{R}_{k+2p}^n . Allora

$$v \in Tl(x) \iff v^t A \geq 0 \iff v^t \in (\mathbb{R}_n A \geq 0)$$

Siccome $f'(x; v) = v^t \nabla f(x)$, abbiamo quindi $(\mathbb{R}_n A \geq 0) \nabla f(x) \geq 0$.

Per il lemma di Farkas (teorema 1.26) ciò implica $\nabla f(x) \in A \mathbb{R}_+^{k+2p}$.

È quindi possibile trovare coefficienti $(\lambda_{i_1}, \dots, \lambda_{i_k}, \mu_{11}, \mu_{12}, \dots, \mu_{p1}, \mu_{p2}) \in \mathbb{R}^+$ tali che

$$\begin{aligned} \nabla f(x) = & -\lambda_{i_1} \nabla g_{i_1}(x) - \dots - \lambda_{i_k} \nabla g_{i_k}(x) + (\mu_{11} - \mu_{12}) \nabla h_1(x) + \dots + \\ & + (\mu_{p1} - \mu_{p2}) \nabla h_p(x) \end{aligned}$$

Ponendo $\lambda_i := 0$ per $i \notin I(x)$ e $\mu_j := \mu_{j1} - \mu_{j2}$ per $j = 1, \dots, p$ abbiamo

$$\nabla_x L(x, \lambda, \mu) := \nabla f(x) + \sum_{i=1}^m \lambda_i \nabla g_i(x) + \sum_{j=1}^p \mu_j \nabla h_j(x) = 0$$

con $g(x) \leq 0$, $h(x) = 0$, $\lambda \geq 0$, $\|\lambda, g(x)\| = 0$.

(x, λ, μ) è perciò un punto di KKT.

11. Il caso generale

Situazione 11.1. Come nel capitolo precedente siano date funzioni $f, g_1, \dots, g_m, h_1, \dots, h_p : \mathbb{R}^n \longrightarrow \mathbb{R}$ tutte di classe C^1 .

Sia ancora $X := (g_1 \leq 0, \dots, g_m \leq 0, h_1 = 0, \dots, h_p = 0)$.

Poniamo $g := (g_1, \dots, g_m)$, $h := (h_1, \dots, h_p)$.

Denotiamo con $\nabla h := (\nabla h_1, \dots, \nabla h_p)$ la trasposta della jacobiana di h .

L'insieme di indici $I(x)$ è stato definito nella def. 10.2.

Lemma 11.2. *Sia $x \in X$. I vettori $\nabla h_1(x), \dots, \nabla h_p(x)$ siano linearmente indipendenti e $v \in \mathbb{R}^n$ tale che $\|\nabla h_j(x), v\| = 0$ per ogni $j = 1, \dots, p$. Allora esistono $\varepsilon > 0$ e una curva $\varphi : (-\varepsilon, \varepsilon) \longrightarrow \mathbb{R}^p$ con le seguenti proprietà:*

(1) φ è di classe C^1 .

(2) $\varphi(0) = 0$.

(3) $\dot{\varphi}(0) = 0$.

(4) $h_j(x + tv + \nabla h(x)\varphi(t)) = 0$ per ogni $t \in (-\varepsilon, \varepsilon)$ ed ogni $j = 1, \dots, p$.

Se le funzioni h_j sono tutte di classe C^2 , anche φ è di classe C^2 .

Dimostrazione. Per ogni $j = 1, \dots, p$ definiamo un'applicazione $H_j : \mathbb{R}^{p+1} \longrightarrow \mathbb{R}$ tramite

$$H_j(y, t) := h_j(x + tv + \nabla h(x)y)$$

ottenendo così una funzione $H : \mathbb{R}^{p+1} \longrightarrow \mathbb{R}^p$ con $H(0, 0) = 0$.

La jacobiana $H'_y(0, 0)$ è data da $H'_y(0, 0) = (\nabla h(x))^t \nabla h(x)$ ed è quindi invertibile essendo per ipotesi i vettori $\nabla h_1(x), \dots, \nabla h_p(x)$, cioè le colonne della matrice $\nabla h(x)$, linearmente indipendenti.

Per il teorema delle funzioni implicite esistono $\varepsilon > 0$ ed una funzione $\varphi : (-\varepsilon, \varepsilon) \longrightarrow \mathbb{R}^p$ di classe C^1 (risp. di classe C^2 se anche le funzioni h_j sono tutte di classe C^2) tali che $\varphi(0) = 0$ e $H(\varphi(t), t) = 0$ per ogni $t \in (-\varepsilon, \varepsilon)$.

Inoltre $\dot{\varphi}(t) = -(H'_y(\varphi(t), t))^{-1} H'_t(\varphi(t), t)$, per cui

$$\dot{\varphi}(0) = -(H'_y(0, 0))^{-1} H'_t(0, 0) = -(H'_y(0, 0))^{-1} (\nabla h(x))^t v = 0$$

$H(\varphi(t), t) = 0$ significa però proprio $h_j(x + tv + \nabla h(x)\varphi(t)) = 0$ per ogni $j = 1, \dots, p$.

Corollario 11.3. *Sia $x \in X$. I vettori $\nabla h_1(x), \dots, \nabla h_p(x)$ siano linearmente indipendenti e $v \in \mathbb{R}^n$ tale che $\|\nabla h_j(x), v\| = 0$ per ogni $j = 1, \dots, p$. Allora esistono $\varepsilon > 0$ e una curva $\gamma : (-\varepsilon, \varepsilon) \longrightarrow \mathbb{R}^n$ con le seguenti proprietà:*

(1) γ è di classe C^1 .

(2) $\gamma(0) = x$.

(3) $\dot{\gamma}(0) = v$.

(4) $h_j(\gamma(t)) = 0$ per ogni $t \in (-\varepsilon, \varepsilon)$ ed ogni $j = 1, \dots, p$.

Se le funzioni h_j sono tutte di classe C^2 , anche γ è di classe C^2 .

Dimostrazione. Nel lemma 11.2 è sufficiente porre $\gamma(t) := x + tv + \nabla h(x)\varphi(t)$.

Proposizione 11.4. Sia $x \in X$. I vettori $\nabla h_1(x), \dots, \nabla h_p(x)$ siano linearmente indipendenti e $v \in \mathbb{R}^n$ tale che $\|\nabla h_j(x), v\| = 0$ per ogni $j = 1, \dots, p$, come nel lemma 11.2, ma anche $\|\nabla g_i(x), v\| < 0$ per ogni $i \in I(x)$.

Allora esistono $\varepsilon > 0$ e una curva $\gamma : (-\varepsilon, \varepsilon) \rightarrow \mathbb{R}^n$ con le seguenti proprietà:

(1) γ è di classe C^1 .

(2) $\gamma(0) = x$.

(3) $\dot{\gamma}(0) = v$.

(4) $h_j(\gamma(t)) = 0$ per ogni $t \in (-\varepsilon, \varepsilon)$ ed ogni $j = 1, \dots, p$.

(5) $g_i(\gamma(t)) < 0$ per ogni $t \in (0, \varepsilon)$ ed ogni $i = 1, \dots, m$.

Se le funzioni g_i ed h_j sono tutte di classe C^2 , anche γ è di classe C^2 .

Dimostrazione. Usiamo la curva γ del cor. 11.3, restringendo, se necessario, ε . Infatti per $i \notin I(x)$, dalla continuità di g_i segue $g_i(\gamma(t)) < 0$ per ogni t sufficientemente piccolo. Possiamo quindi assumere che $g_i(\gamma(t)) < 0$ per ogni $t \in (-\varepsilon, \varepsilon)$ ed ogni $i = 1, \dots, m$.

Sia $i \notin I(x)$ fissato. Definiamo $\alpha : (-\varepsilon, \varepsilon) \rightarrow \mathbb{R}$ tramite $\alpha(t) := g_i(\gamma(t))$. Allora $\dot{\alpha}(t) = \|\nabla g_i(\gamma(t)), \dot{\gamma}(t)\|$ e quindi $\dot{\alpha}(0) = \|\nabla g_i(\gamma(0)), \dot{\gamma}(0)\| = \|\nabla g_i(x), v\| < 0$. Ciò implica $\alpha(t) < 0$ per ogni $t > 0$ sufficientemente piccolo. Restringendo ancora ε troviamo che $g_i(\gamma(t)) < 0$ per ogni $i = 1, \dots, m$ ed ogni $t \in (0, \varepsilon)$.

Definizione 11.5. Un punto $x \in X$ si dice *regolare nel senso di Mangasarian-Fromovitz*, se sono soddisfatte le seguenti condizioni:

(1) I vettori $\nabla h_1(x), \dots, \nabla h_p(x)$ sono linearmente indipendenti.

(2) Esiste un vettore $v \in \mathbb{R}^n$ tale che $\|\nabla h_j(x), v\| = 0$ per ogni $j = 1, \dots, p$ e tale che $\|\nabla g_i(x), v\| < 0$ per ogni $i \in I(x)$.

In tal caso v si chiama un *vettore di Mangasarian-Fromovitz* per x .

Queste condizioni coincidono con le ipotesi della prop. 11.4.

Lemma 11.6. Sia $x \in X$ un punto regolare nel senso di Mangasarian-Fromovitz e sia v un vettore di Mangasarian-Fromovitz per x . Sia $u \in \text{Tl}(x)$. Allora $u + \theta v \in \text{T}(X, x)$ per ogni $\theta > 0$.

Dimostrazione. Sia $\theta > 0$ fissato.

(1) Allora anche $u + \theta v$ è un vettore di Mangasarian-Fromovitz per x . Infatti, usando l'ipotesi che $u \in \text{Tl}(x)$, abbiamo

$$\| \nabla h_j(x), u + \theta v \| = \| \nabla h_j(x), u \| + \theta \| \nabla h_j(x), v \| = 0$$

per ogni $j = 1, \dots, p$ e

$$\| \nabla g_i(x), u + \theta v \| = \| \nabla g_i(x), u \| + \theta \| \nabla g_i(x), v \| < 0$$

per ogni $i \in I(x)$.

(2) Per la prop. 11.4 esistono $\varepsilon > 0$ e una curva $\gamma : (-\varepsilon, \varepsilon) \rightarrow \mathbb{R}^n$ con le proprietà enunciate in quella proposizione, con $u + \theta v$ al posto di v . In particolare abbiamo $\gamma(0) = x$ e $\dot{\gamma}(0) = u + \theta v$. Ma allora

$$u + \theta v = \dot{\gamma}(0) = \lim_{k \rightarrow \infty} \frac{\gamma(1/k) - \gamma(0)}{1/k} = \lim_{k \rightarrow \infty} \frac{\gamma(1/k) - x}{1/k}$$

e siccome $\lim_{k \rightarrow \infty} \gamma(1/k) = \gamma(0) = x$, vediamo che $u + \theta v \in T(X, x)$.

Proposizione 11.7. *Sia $x \in X$ un punto regolare nel senso di Mangasarian-Fromovitz. Allora x è un punto regolare nel senso di Abadie.*

Dimostrazione. Per la prop. 10.4 dobbiamo dimostrare che $Tl(x) \subset T(X, x)$.

Siano $u \in Tl(x)$ e v un vettore di Mangasarian-Fromovitz per x . Per il lemma 11.6 $u + \theta v \in T(X, x)$ per ogni $\theta > 0$. Ma $\lim_{\theta \rightarrow 0} u + \theta v = u$, quindi vediamo che $u \in \overline{T(X, x)}$. Dal lemma 9.8 sappiamo però che $T(X, x)$ è chiuso e quindi $u \in T(X, x)$.

Teorema 11.8. *Sia $x \in X$ regolare nel senso di Mangasarian-Fromovitz. x sia inoltre un punto di minimo locale di f in X .*

Allora esistono $\lambda \in \mathbb{R}_m$ e $\mu \in \mathbb{R}_p$ tali che (x, λ, μ) sia un punto di KKT di (X, f) .

Dimostrazione. Ciò segue dal teorema 10.11 applicando la prop. 11.7.

Definizione 11.9. *Sia $x \in X$. Diciamo che x è un punto a vincoli linearmente indipendenti, se i gradienti $\nabla g_i(x)$ per $i \in I(x)$ e $\nabla h_1(x), \dots, \nabla h_p(x)$ sono linearmente indipendenti.*

Lemma 11.10. *x sia un punto a vincoli linearmente indipendenti. Allora x è regolare nel senso di Mangasarian-Fromovitz.*

Dimostrazione. La condizione (1) della def. 11.5 è evidentemente soddisfatta.

(2) Possiamo assumere che $I(x) = \{1, \dots, q\}$ con $0 \leq q \leq m$. Per ipotesi esistono vettori $A_{q+p+1}, \dots, A_n \in \mathbb{R}^n$ tali che la matrice $A := (\nabla g_1(x), \dots, \nabla g_q(x), \nabla h_1(x), \dots, \nabla h_p(x), A_{q+p+1}, \dots, A_n)$ sia invertibile. Sia $b := (\underbrace{-1, \dots, -1}_q, \underbrace{0, \dots, 0}_p, 1, \dots, 1)^t$. Allora l'equazione $A^t v = b$

possiede un'unica soluzione v perché la matrice A è invertibile. È immediato che v soddisfa la condizione (2) della def. 11.5.

Teorema 11.11. *x sia un punto di minimo locale di f in X e allo stesso tempo anche un punto a vincoli linearmente indipendenti.*

Allora esistono, univocamente determinati, $\lambda \in \mathbb{R}_m$ e $\mu \in \mathbb{R}_p$ tali che (x, λ, μ) sia un punto di KKT di (X, f) .

Dimostrazione. (1) Per il lemma 11.10 l'esistenza di λ e μ segue dal teorema 11.8.

(2) Dimostriamo l'unicità. Assumiamo di nuovo che $I(x) = \{1, \dots, k\}$ con $0 \leq k \leq m$. Dall'oss. 10.10 sappiamo che $\lambda_i = 0$ per $i > k$. Inoltre

$$\sum_{i=1}^k \lambda_i \nabla g_i(x) + \sum_{j=1}^p \mu_j \nabla h_j(x) = -\nabla f(x)$$

Questa rappresentazione, e quindi i numeri $\lambda_1, \dots, \lambda_k, \mu_1, \dots, \mu_p$, sono univocamente determinati perché, per ipotesi, i vettori $\nabla g_1(x), \dots, \nabla g_k(x), \nabla h_1(x), \dots, \nabla h_p(x)$ sono linearmente indipendenti.

12. Restrizioni lineari

Situazione 12.1. Siano $f : \mathbb{R}^n \rightarrow \mathbb{R}$ di classe C^1 ed $a^1, \dots, a^m, b^1, \dots, b^p \in \mathbb{R}_n$, $\alpha_1, \dots, \alpha_m, \beta_1, \dots, \beta_p \in \mathbb{R}$. Sia

$$X := \{x \in \mathbb{R}^n \mid a^1 x \leq \alpha_1, \dots, a^m x \leq \alpha_m, b^1 x = \beta_1, \dots, b^p x = \beta_p\}$$

Ponendo $g_i := \bigcirc_x a^i x - \alpha_i$ per $i = 1, \dots, m$ ed $h_j := \bigcirc_x b^j x - \beta_j$ per $j = 1, \dots, p$ abbiamo $X = (g_1 \leq 0, \dots, g_m \leq 0, h_1 = 0, \dots, h_p = 0)$ come nei capitoli precedenti.

Proposizione 12.2. Ogni punto di X è regolare nel senso di Abadie.

Dimostrazione. Sia $x \in X$. Per la prop. 10.4 dobbiamo solo dimostrare che $\text{Tl}(x) \subset \text{T}(X, x)$.

Sia $v \in \text{Tl}(x)$. Siccome $\nabla g_i(x) = (a^i)^t$ e $\nabla h_j(x) = (b^j)^t$ per ogni i, j , ciò significa che $a^i v \leq 0$ per ogni $i \in I(x)$ e $b^j v = 0$ per ogni $j = 1, \dots, p$.

Per $k \in \mathbb{N} + 1$ siano $t_k := \frac{1}{k}$ ed $x_k := x + t_k v$. Per k sufficientemente grande abbiamo allora

$$a^i x_k = a^i(x + t_k v) = \alpha_i + t_k a^i v \leq \alpha_i \text{ per } i \in I(x)$$

$$a^i x_k = a^i(x + t_k v) = a^i x + t_k a^i v < \alpha_i \text{ per } i \notin I(x)$$

$$b^j x_k = b^j(x + t_k v) = \beta_j + t_k b^j v = \beta_j \text{ per } j = 1, \dots, p$$

Ciò mostra che $x_k \in X$ per $k \gg 0$. Inoltre da un lato $\bigcirc_k x_k \rightarrow x$,

dall'altro $\bigcirc_k \frac{x_k - x}{t_k} = \bigcirc_k v \rightarrow v$, per cui $v \in \text{T}(X, x)$.

Teorema 12.3. x sia un punto di minimo locale di f . Allora esistono $\lambda \in \mathbb{R}_m$ e $\mu \in \mathbb{R}_p$ tali che sono soddisfatte le seguenti condizioni:

$$(\nabla f(x))^t + \sum_{i=1}^m \lambda_i a^i + \sum_{j=1}^p \mu_j b^j = 0$$

$$a^i x \leq \alpha_i \text{ per } i = 1, \dots, m$$

$$b^j x = \beta_j \text{ per } j = 1, \dots, p$$

$$\lambda \geq 0$$

$$\lambda_i (a^i x - \alpha_i) = 0 \text{ per } i = 1, \dots, m$$

Dimostrazione. Per la prop. 12.2 x è regolare nel senso di Abadie, per cui possiamo applicare il teorema 10.11. Le condizioni nell'enunciato sono esattamente le condizioni di KKT, tenendo conto dell'oss. 10.10.

13. Restrizioni convesse

Osservazione 13.1. X sia un sottoinsieme convesso di \mathbb{R}^n ed $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una funzione convessa. Se x è un punto di minimo locale di f in X , allora x è anche un punto di minimo globale, cioè $f(x) \leq f(y)$ per ogni $y \in X$. Possiamo quindi tralasciare le specificazioni "locale" risp. "globale".

Dimostrazione. Altrimenti esiste un $y \in X$ tale che $f(y) < f(x)$. Siccome x è un punto di minimo locale, possiamo trovare $t \in (0, 1]$ con $f(x + t(y - x)) \geq f(x)$. La funzione f è però convessa, per cui

$f(x + t(y - x)) \leq f(x) + tf(y) - tf(x) < f(x) + tf(x) - tf(x) = f(x)$ una contraddizione. Nell'ultima disuguaglianza abbiamo utilizzato la condizione che $t > 0$.

Lemma 13.2. La funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}$ sia di classe C^1 . Allora sono equivalenti:

- (1) f è convessa.
- (2) Per $x, y \in \mathbb{R}^n$ vale $\|\nabla f(x), y - x\| \leq f(y) - f(x)$.

Dimostrazione. Corsi di Analisi.

Situazione 13.3. Siano $f, g_1, \dots, g_m : \mathbb{R}^n \rightarrow \mathbb{R}$ funzioni convesse di classe C^1 e $b^1, \dots, b^p \in \mathbb{R}^n, \beta_1, \dots, \beta_p \in \mathbb{R}$. Siano

$$X := \{x \in \mathbb{R}^n \mid g_1(x) \leq 0, \dots, g_m(x) \leq 0, b^1x = \beta_1, \dots, b^px = \beta_p\}$$

$$\tilde{X} := \{x \in \mathbb{R}^n \mid g_1(x) < 0, \dots, g_m(x) < 0, b^1x = \beta_1, \dots, b^px = \beta_p\}$$

È immediato che gli insiemi X e \tilde{X} sono convessi.

Osservazione 13.4. Sia $(x, \lambda, \mu) \in \mathbb{R}^n \times \mathbb{R}_m \times \mathbb{R}_p$. Allora sono equivalenti:

- (1) (x, λ, μ) è un punto di KKT di (X, f) .
- (2) $\nabla f(x) + \sum_{i \in I(x)} \lambda_i \nabla g_i(x) + \sum_{j=1}^p \mu_j (b^j)^t = 0$
 $g_i(x) \leq 0$ per ogni $i = 1, \dots, m$
 $b^jx = \beta_j$ per ogni $j = 1, \dots, p$
 $\lambda \geq 0$
 $\|\lambda, g(x)\| = 0$

Dimostrazione. È sufficiente osservare che $\lambda_i = 0$ per $i \notin I(x)$ per l'oss. 10.10.

Proposizione 13.5. $(x, \lambda, \mu) \in \mathbb{R}^n \times \mathbb{R}_m \times \mathbb{R}_p$ sia un punto di KKT di (X, f) . Allora x è un punto di minimo di f in X .

Dimostrazione. Per definizione abbiamo $x \in X$. Sia $y \in X$. Dalle condizioni di KKT e dal lemma 13.2 segue

$$\begin{aligned} f(y) &\geq f(x) + \|\nabla f(x), y - x\| \\ &= f(x) - \sum_{i \in I(x)} \lambda_i \|\nabla g_i(x), y - x\| - \sum_{j=1}^p \mu_j b^j(y - x) \\ &= f(x) - \sum_{i \in I(x)} \lambda_i \|\nabla g_i(x), y - x\| \end{aligned}$$

dove abbiamo tenuto conto dell'oss. 13.4 e del fatto che anche $y \in X$.

Però anche le funzioni g_i sono convesse, perciò dal lemma 13.2 otteniamo

$$\begin{aligned} \sum_{i \in I(x)} \lambda_i \|\nabla g_i(x), y - x\| &\leq \sum_{i \in I(x)} \lambda_i (g_i(y) - g_i(x)) \\ &= \sum_{i \in I(x)} \lambda_i g_i(y) \leq 0 \end{aligned}$$

e ciò mostra $f(x) \leq f(y)$.

Definizione 13.6. Per $x \in X$ sia

$$\begin{aligned} \text{Tls}(x) &:= \{v \in \mathbb{R}^n \mid \|\nabla g_i(x), v\| < 0 \text{ per ogni } i \in I(x), \\ &\quad b^j v = 0 \text{ per } j = 1, \dots, p\} \end{aligned}$$

$\text{Tls}(x)$ si chiama il *cono tangente linearizzato stretto* di X in x .

Lemma 13.7. Sia $x \in X$. Allora $\text{Tls}(x) \subset T(X, x)$.

Dimostrazione. Sia $v \in \text{Tls}(x)$. Per $k \in \mathbb{N} + 1$ poniamo $t_k := \frac{1}{k}$ e $x_k := x + t_k v$. Allora $\bigcirc_k x_k \rightarrow x$, $\bigcirc_k t_k \in [\mathbb{R}^{++} \rightarrow 0]$ e

$\bigcirc_k \frac{x_k - x}{t_k} = \bigcirc_k v \in [\mathbb{R}^n \rightarrow v]$. Pertanto è sufficiente dimostrare che $x_k \in X$ per $k \gg 0$.

(1) In primo luogo per ogni $j = 1, \dots, p$ ed ogni $k \in \mathbb{N} + 1$ abbiamo

$$b^j x_k = b^j(x + t_k v) = b^j x + t_k b^j v = \beta_j$$

perché $x \in X$ e $v \in \text{Tls}(x)$.

(2) Sia $i \in \{1, \dots, m\}$ ed $i \notin I(x)$. Allora $g_i(x) < 0$ e quindi $g_i(x_k) < 0$ per $k \gg 0$.

(3) Sia $i \in I(x)$ e quindi $g_i(x) = 0$. Per ipotesi

$$0 > \|\nabla g_i(x), v\| = \lim_{t \rightarrow 0} \frac{g_i(x + tv) - g_i(x)}{t} = \lim_{t \rightarrow 0} \frac{g_i(x + tv)}{t} = \lim_{k \rightarrow \infty} \frac{g_i(x_k)}{t_k}$$

Ciò implica $g_i(x_k) < 0$ per $k \gg 0$.

Lemma 13.8. Sia $x \in X$. Se $\tilde{X} \neq \emptyset$, allora

$$\text{Tls}(x) = T(X, x) = \text{TI}(x).$$

Dimostrazione. (1) Siccome per il lemma 9.8 $T(X, x)$ è chiuso, la prop. 10.4 e il lemma 13.7 implicano che $\overline{\text{Tls}(x)} \subset T(X, x) \subset \text{TI}(x)$.

(2) È sufficiente dimostrare quindi che $\text{Tl}(x) \subset \overline{\text{Tls}(x)}$.
 Sia $v \in \text{Tl}(x)$ e quindi $\|\nabla g_i(x), v\| \leq 0$ per ogni $i \in I(x)$ e $b^j v = 0$ per ogni $j = 1, \dots, p$.

Per ipotesi $\tilde{X} \neq \emptyset$, perciò possiamo trovare $y \in \mathbb{R}^n$ tale che $g_i(y) < 0$ per ogni $i = 1, \dots, m$ e $b^j y = \beta_j$ per ogni $j = 1, \dots, p$.

Dal lemma 13.2 abbiamo $\|\nabla g_i(x), y - x\| \leq g_i(y) - g_i(x) = g_i(y) < 0$ per ogni $i \in I(x)$, mentre $b^j(y - x) = b^j y - b^j x = \beta_j - \beta_j = 0$ per ogni $j = 1, \dots, p$. Se per $k \in \mathbb{N} + 1$ poniamo $t_k := \frac{1}{k}$ e $w_k := v + t_k(y - x)$, abbiamo quindi

$\|\nabla g_i(x), w_k\| = \|\nabla g_i(x), v\| + t_k \|\nabla g_i(x), y - x\| < 0$
 per ogni $i \in I(x)$ e $b^j w_k = b^j v = 0$, cosicché $w_k \in \text{Tls}(x)$ per ogni $k \in \mathbb{N} + 1$. Però $\lim_{k \rightarrow \infty} w_k = v$ e vediamo che $v \in \overline{\text{Tls}(x)}$.

Teorema 13.9. Sia $\tilde{X} \neq \emptyset$. Allora per $x \in X$ sono equivalenti:

- (1) x è un punto di minimo per f in X .
- (2) Esiste $(\lambda, \mu) \in \mathbb{R}_m \times \mathbb{R}_p$ tale che (x, λ, μ) è un punto di KKT di (X, f) .

Dimostrazione. (1) \implies (2): Lemma 13.8 e teorema 10.11

(2) \implies (1): Prop. 13.5.

Teorema 13.10. Le funzioni g_i per $i = 1, \dots, m$ siano della forma $g_i(x) = a^i x - \alpha_i$ con $a^i \in \mathbb{R}^n$ ed $\alpha_i \in \mathbb{R}$.

Allora per $x \in X$ sono equivalenti:

- (1) x è un punto di minimo per f in X .
- (2) Esiste $(\lambda, \mu) \in \mathbb{R}_m \times \mathbb{R}_p$ tale che (x, λ, μ) è un punto di KKT di (X, f) .

Dimostrazione. Se qui $\tilde{X} = \emptyset$, allora possiamo aggiungere le funzioni $\bigcirc_x a^i x - \alpha_i$ alle funzioni $\bigcirc_x b^j x - \beta_j$, trovandoci così in un caso in cui le ipotesi del teorema 13.9 (o del lemma 13.8) sono banalmente soddisfatte.

Bibliografia

- M. Aigner:** Diskrete Mathematik. Vieweg 2004.
- W. Alt:** Nichtlineare Optimierung. Vieweg 2002.
- I. Baroncelli:** Ottimizzazione genetica. Tesi, Ferrara 1991.
- K. Bennett/O. Mangasarian:** Robust linear programming discrimination of two linearly inseparable sets.
Opt. Meth. Software 1 (1992), 23-34.
- D. Bertsekas:** Nonlinear programming. Athena Scientific 1999.
- S. Boyd/L. Vandenberghe:** Convex optimization. Cambridge UP 2008.
- R. Cottle/E. Johnson/R. Wets:** George B. Dantzig (1914-2005).
Notices AMS March 2007, 344-362.
- G. Dantzig:** Linear programming. Operations Res. 50/1 (2002), 42-47.
- J. Eaton/D. Bateman/S. Hauberg:** GNU Octave. Network Theory 2008.
- S. Gass:** The life and times of the father of linear programming.
OR/MS Today August 2005, 10p.
- C. Geiger/C. Kanzow:** Theorie und Numerik restringierter Optimierungsaufgaben. Springer 2002.
- J. Gross/J. Yellen:** Graph theory and its applications.
Chapman & Hall 2006.
- F. Jarre/J. Stoer:** Optimierung. Springer 2004.
- D. Jungnickel:** Optimierungsmethoden. Springer 1999.
- O. Mangasarian/W. Street/W. Wolberg:** Breast cancer diagnosis and prognosis via linear programming. Op. Res. 43/4 (1995), 570-578.
- J. Matoušek/B. Gärtner:** Understanding and using linear programming.
Springer 2007, 220p. Eur 41.
- J. Nocedal/S. Wright:** Numerical optimization. Springer 2006.