

Marco Casu

🌀 Ottimizzazione 🌀



SAPIENZA
UNIVERSITÀ DI ROMA

Facoltà di Ingegneria dell'Informazione, Informatica e Statistica
Dipartimento di Informatica

Questo documento è distribuito sotto la licenza [GNU](#), è un resoconto degli appunti (eventualmente integrati con libri di testo) tratti dalle lezioni del corso di Ottimizzazione per la laurea triennale in Informatica. Se dovessi notare errori, ti prego di segnalarmeli.



INDICE

1	Flussi nei Grafi	3
1.1	Definizione e Grafo Residuo	3
1.2	Tagli $s - t$	5
1.3	Percorso Minimo nell'Aumento del Flusso	7
1.4	Cammini Edge-Disjoint in un Grafo	9
2	Programmazione Lineare	10
2.1	Insiemi Convessi	10
2.2	Applicazioni della Programmazione Lineare	13
2.3	Il Metodo del Simplexso	15
2.3.1	Soluzioni Ammissibili Basiche	16
2.3.2	Ricerca sul Poliedro	20
2.3.3	La Procedura di Risoluzione	25
2.4	Il Tableau del Simplexso	28
2.4.1	Trovare una Base Ammissibile	30
2.5	Soluzioni Degenere	32
2.6	Problemi del Metodo del Simplexso	33
2.6.1	Regola del Pivot	33
2.6.2	Cicli e Variabili Fickle	34
3	Dualità	37
3.1	Dualità Forte e Debole	37
3.2	Lemma di Farkas	40
3.2.1	interpretazione Geometrica del Lemma di Farkas	42
3.3	Il Teorema degli Slack Complementari	45
3.3.1	Alcune Applicazioni	47
4	Metodi Polinomiali per la Risoluzione di un LP	48
4.1	Metodo degli Ellissoidi	49
4.1.1	Tecnicalità Implementative	52
4.2	Accenno al Metodo dei Punti Interni	53
5	Programmazione Intera	54
5.1	Inviluppo Convesso dei Punti Interi	54
5.1.1	Il Teorema di Gomory	57
5.2	Il Matching Perfetto di Peso Minimo	58
5.2.1	PM per i Grafi Bipartiti	58
5.2.2	PM per i Grafi non Bipartiti	59

CAPITOLO

1

FLUSSI NEI GRAFI

1.1 Definizione e Grafo Residuo

Definizione 1 Una **network** o **rete** $G = (V, E, c, s, t)$ è un particolare grafo diretto, in cui V ed E sono i vertici e gli archi, tali per cui è soddisfatta la condizione

$$\forall (u, v) \in E(G), \quad \exists (v, u) \in E(G)$$

$c : E(G) \rightarrow \mathbb{R}^+$ è una funzione detta **capacità**, s e t sono due particolari vertici in $V(G)$ denominati **source** e **sink**.

Definizione 2 Data una network $G = (V, E, c, s, t)$, un **flusso** per G è una funzione $f : E(G) \rightarrow \mathbb{R}$ tale per cui valgono le seguenti

1. **skew-simmetria**: $f(u, v) = -f(v, u), \quad \forall (u, v) \in E(G)$
2. **capacità rispettata**: $f(u, v) \leq c(u, v), \quad \forall (u, v) \in E(G)$
3. **conservatività del flusso**: $\sum_{(u, v) \in E(G)} f(u, v) = 0, \quad \forall v \in V(G) \setminus \{s, t\}$

Denominiamo flusso uscente dal vertice v la somma del flusso (positivo) valutato su tutti gli archi che hanno v come primo membro (che collegano v ad un'altro vertice). Analogamente (ma in maniera opposta) si definisce il flusso entrante. Dato un flusso f per una network G si definisce il **valore del flusso** la somma del flusso uscente da s

$$\text{val}(f) = \sum_{(s, u) \in E(G)} f(s, u)$$

La terza proprietà, di conservazione del flusso, asserisce che il flusso uscente da un nodo deve essere identico al flusso entrante, sia x un vertice fissato in $V(G)$

$$\sum_{\substack{(u, x) \in E(G) \\ f(u, x) > 0}} f(u, x) = - \left(\sum_{\substack{(x, u) \in E(G) \\ f(x, u) < 0}} f(x, u) \right)$$

Definizione 3 Sia $G = (V, E, c, s, t)$ una network e f un flusso per G , il **grafo residuo** è il grafo diretto G' definito come segue

- $\forall v \in V(G), v \in V(G')$
- $(u, v) \in E(G) \wedge f(u, v) < c(u, v) \implies (u, v) \in E(G')$

Inoltre è definita una funzione $r : E(G') \rightarrow \mathbb{R}^+$ detta **capacità residua** definita come segue

$$r(u, v) = c(u, v) - f(u, v)$$

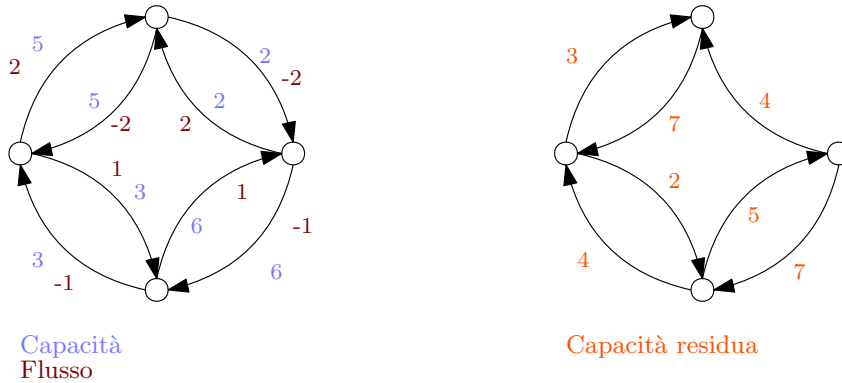


Figura 1.1: Capacità residua del flusso (evidenziato in rosso)

Si assuma che esista un cammino P in G' da s a t , si consideri il residuo minimo valutato sugli archi contenuti nel cammino

$$\alpha = \min_{(u,v) \in E(P)} r(u, v)$$

Si definisce una funzione $f' : E(G) \rightarrow \mathbb{R}$ come segue

$$f'(u, v) = \begin{cases} f(u, v) + \alpha & \text{se } (u, v) \in E(P) \\ f(u, v) - \alpha & \text{se } (v, u) \in E(P) \\ f(u, v) & \text{altrimenti} \end{cases}$$

Proposizione 1 f' è un flusso per G .

Dimostrazione : Sia (u, v) un arco in G , se $(u, v) \notin E(P)$, allora $f'(u, v) = f(u, v)$ e conseguentemente $f'(v, u) = f(v, u)$, quindi la proprietà di skew simmetria è preservata. Differentemente, se $(u, v) \in E(P)$ si avrebbe che $f'(u, v) = f(u, v) + \alpha$ e $f'(v, u) = f(v, u) - \alpha = -f(u, v) - \alpha = -(f(u, v) + \alpha)$, quindi il nuovo flusso rispetta la proprietà di skew-simmetria.

Per ogni arco $(u, v) \in E(P)$ si ha che $f'(u, v) = f(u, v) + \alpha$, α è (per definizione) minore o uguale a $r(u, v)$ quindi

$$f'(u, v) \leq f(u, v) + r(u, v)$$

Ma essendo che $f(u, v) + r(u, v) = c(u, v)$, f' rispetta la capacità.

Se $x \notin V(P)$ si avrebbe che $f'(x, u) = f(x, u)$ per ogni u adiacente ad x , allora

$$\sum_{(x,u) \in E(G)} f(x, u) = 0$$

Assumendo che $x \in V(P)$, vi è un arco uscente da x il cui flusso è aumentato di α , vi è quindi (per definizione di f') un'arco entrante in x il cui flusso è diminuito di α , quindi è ancora vero che

$$\sum_{\substack{(u,x) \in E(G) \\ f'(u,x) > 0}} f'(u, x) = - \left(\sum_{\substack{(x,u) \in E(G) \\ f'(x,u) < 0}} f'(x, u) \right)$$

la proprietà di conservazione del flusso è rispettata. ■

Il valore del nuovo flusso è uguale al valore del flusso di partenza aumentato di α

$$\text{val}(f') = \text{val}(f) + \alpha$$

Algorithm 1 Ford–Fulkerson**Require:** network $G = (V, E, c, s, t)$ si definisce un flusso f tale che $f(u, v) = 0, \forall (u, v) \in E(G)$ si definisce il grafo residuo G' dato il flusso f **while** Esiste un cammino P in G' da s a t **do** si definisce la funzione delle capacità residue $r : E(G') \rightarrow \mathbb{R}$ $\alpha = \min_{(u,v) \in E(P)} r(u, v)$ Si definisce un flusso $f' = f$ **for** $(u, v) \in E(P)$ **do** $f'(u, v) = f(u, v) + \alpha$ $f'(v, u) = f(v, u) - \alpha$ **end for****end while**

Dato che un singolo arco (s, u) per qualche u è necessariamente presente nel cammino P da s a t , ed il valore di f' su (s, u) è stato aumentato di α . La proposizione 1 delinea una procedura per la ricerca di un flusso ottimale (di valore massimo) per una network. Alla fine dell'esecuzione, il flusso f' sarà ottimale per la network data.

Osservazione 1 Se le capacità della network sono numeri interi, l'algoritmo termina. Se invece le capacità sono numeri reali, l'algoritmo potrebbe non terminare.

1.2 Tagli $s - t$

Data una network $G = (V, E, c, s, t)$, ed un flusso f per G , si consideri un'insieme $\mathcal{U} \subset V(G)$ tale che

- $s \in \mathcal{U}$
- $t \notin \mathcal{U}$

Tale insieme è detto **insieme di taglio**, si consideri ora il flusso uscente dai vertici presenti in \mathcal{U}

$$\sum_{\substack{(u,x) \in E(G) \\ \text{t.c. } u \in \mathcal{U}}} f(u, x)$$

Per la proprietà di conservazione del flusso si ha che il flusso uscente da ogni vertice diverso da s è nullo, ed il flusso uscente dal vertice s è il valore del flusso.

$$\sum_{\substack{(u,x) \in E(G) \\ \text{t.c. } u \in \mathcal{U}}} f(u, x) = \sum_{(s,x) \in E(G)} f(s, x) = \text{val}(f)$$

La sommatoria a sinistra può essere riscritta come la somma del flusso uscente dai vertici in \mathcal{U} verso i vertici in \mathcal{U} , e del flusso uscente dai vertici in \mathcal{U} verso i vertici che non sono contenuti in \mathcal{U}

$$\sum_{\substack{(u,x) \in E(G) \\ \text{t.c. } u \in \mathcal{U}}} f(u, x) = \sum_{\substack{(u,x) \in E(G) \\ \text{t.c. } u, x \in \mathcal{U}}} f(u, x) + \sum_{\substack{(u,x) \in E(G) \\ \text{t.c. } u \in \mathcal{U} \\ x \notin \mathcal{U}}} f(u, x)$$

Per la proprietà di skew-simmetria il flusso uscente dai vertici in \mathcal{U} verso i vertici in \mathcal{U} è nullo

$$\sum_{\substack{(u,x) \in E(G) \\ \text{t.c. } u \in \mathcal{U}}} f(u, x) = \sum_{\substack{(u,x) \in E(G) \\ \text{t.c. } u \in \mathcal{U} \\ x \notin \mathcal{U}}} f(u, x)$$

Conclusione : Il valore di f è uguale alla somma dei flussi uscenti dai vertici in \mathcal{U} verso i vertici non contenuti in \mathcal{U} . Questa proprietà è invariante rispetto la scelta di \mathcal{U} , a patto che rispetti le proprietà inizialmente elencate (deve contenere s ma non t).

Definizione 4 si definisce **capacità di taglio** la somma delle capacità degli archi che collegano i vertici in \mathcal{U} ai vertici in $V(G) \setminus \mathcal{U}$

$$c_t = \sum_{\substack{(u,x) \in E(G), \\ u \in \mathcal{U}, \\ x \notin \mathcal{U}}} c(u,x)$$



Osservazione 2 il valore massimale del flusso è limitato dalla capacità di taglio

$$\text{val}(f) \leq c_t$$

Proposizione 2 Data una network G , se esiste un flusso f^* ed un'insieme di taglio \mathcal{U} tali che

$$\text{val}(f) = \sum_{\substack{(u,x) \in E(G), \\ u \in \mathcal{U}, \\ x \notin \mathcal{U}}} c(u,x)$$

ossia, il valore del flusso è identico alla capacità di taglio, allora f^* è un flusso ottimale.

L'algoritmo di Ford-Fulkerson restituisce un flusso ottimale f^* , da questo è possibile individuare l'insieme di taglio \mathcal{U} associato, in particolare, se G^* è il grafo residuo della network rispetto il flusso dato in output f^* , allora l'insieme di taglio sarà composto da tutti i nodi raggiungibili da s in G^* , chiaramente, fra questi non vi sarà t , data la definizione dell'algoritmo, che termina proprio quando non vi è un cammino da s a t .

Si consideri adesso una network G , di cui f^* è il flusso ottimale trovato tramite l'algoritmo 1. Sia \mathcal{U} l'insieme di taglio dato dai nodi raggiungibili da s nel grafo residuo G^* .

Osservazione 3 Per ogni arco $(x,y) \in E(G)$ con $x \in \mathcal{U}$ e $y \notin \mathcal{U}$, si avrà che

$$f^*(x,y) = c(x,y)$$

Il valore del flusso è uguale alla somma delle capacità degli archi che collegano i vertici in \mathcal{U} a quelli fuori da \mathcal{U}

$$\sum_{\substack{(x,y) \in E(G) \\ x \in \mathcal{U} \\ y \notin \mathcal{U}}} f^*(x,y) = \sum_{\substack{(x,y) \in E(G) \\ x \in \mathcal{U} \\ y \notin \mathcal{U}}} c(x,y) = \text{val}(f^*)$$

La proposizione 2 non implica che non ci possa essere una network il cui flusso ottimale a valore strettamente minore della capacità di taglio di uno specifico insieme \mathcal{U} , si consideri l'immagine in figura 1.2, in cui è applicata la notazione sugli archi *capacità/flusso*, la capacità di taglio è data dalla somma delle capacità sugli archi evidenziati, ed è uguale a 4, nonostante questo, il flusso ottimale per la network in questione ha valore 1.

Nonostante ciò, esiste sempre un insieme \mathcal{U} contenente s e non t la cui capacità di taglio è uguale al valore del flusso ottimale per la network data, tale insieme può essere trovato adoperando l'algoritmo di Ford-Fulkerson nella procedura precedentemente elencata.

Osservazione 4 L'algoritmo di Ford-Fulkerson, termina sempre se le capacità della network sono numeri in \mathbb{Q} .

Dimostrazione : Se le capacità c_i sono numeri razionali allora esiste un numero naturale $N \in \mathbb{N}$ tale che ogni capacità è della forma $c_i = \frac{a_i}{N}$, ad ogni iterazione dell'algoritmo il valore del flusso aumenta di almeno $\frac{1}{N}$, quindi in un numero finito di passi raggiungerà il valore ottimale. ■

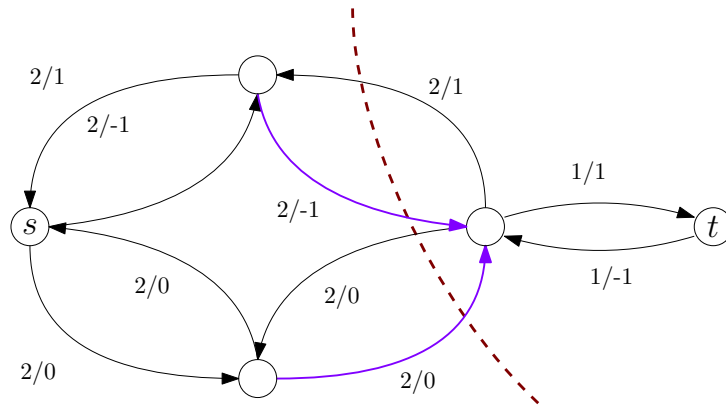


Figura 1.2: network con taglio sui vertici

1.3 Percorso Minimo nell'Aumento del Flusso

Durante la computazione dell'algoritmo di Ford-Fulkerson, viene scelto un qualsiasi percorso che connetta s a t nel grafo residuo, tale scelta comporta un aumento del valore del flusso, ma una scelta differente di percorso potrebbe far sì che l'aumento in quella iterazione sia maggiore, e che il numero finale di iterazioni per trovare il flusso ottimale sia minore. Il seguente esempio mostra l'inefficienza dell'algoritmo 1, si consideri la network in figura 1.3 (alcuni archi sono stati omessi). Il flusso massimale ha valore $2M$,

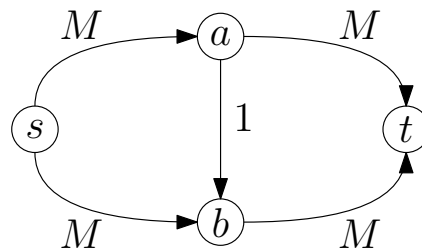


Figura 1.3: Sugli archi sono indicate le capacità

nonostante ciò, se ad ogni iterazione dell'algoritmo venisse selezionato il percorso $s \rightarrow a \rightarrow b \rightarrow t$, allora l'aumento del valore sarebbe uguale ad uno, e sarebbero necessarie $2M$ iterazioni, diversamente, la scelta del percorso $s \rightarrow a \rightarrow t$ implicherebbe già solo alla prima iterazione un'aumento pari ad M .

La complessità computazionale in questo caso dipende linearmente da M , tale valore è però codificato in binario (occupando $\log M$ spazio), quindi l'algoritmo di Ford-Fulkerson è esponenziale nelle dimensioni dell'input. È possibile considerare una rivisitazione dell'algoritmo 1, in cui ad ogni iterazione viene selezionato il percorso più breve (minor numero di archi) da s a t nel grafo residuo. Tale algoritmo rivisitato è noto con il nome di **Edmonds-Karp**.

Algorithm 2 Edmonds-Karp

Require: network $G = (V, E, c, s, t)$

si definisce un flusso f tale che $f(u, v) = 0, \forall (u, v) \in E(G)$

si definisce il grafo residuo G' dato il flusso f

while Esiste un cammino P in G' da s a t **do**

P = cammino più breve da s a t in G'

 si definisce la funzione delle capacità residue $r : E(G') \rightarrow \mathbb{R}$

$\alpha = \min_{(u,v) \in E(P)} r(u, v)$

 Si definisce un flusso $f' = f$

for $(u, v) \in E(P)$ **do**

$f'(u, v) = f(u, v) + \alpha$

$f'(v, u) = f(v, u) - \alpha$

end for

end while



Osservazione 5 Se G è un grafo diretto e P è il percorso più breve fra due vertici x ed y , allora $\forall z \in V(P)$, si ha che il sotto cammino $x \rightarrow z$ in P è anch'esso un percorso più breve.

Proposizione 3 Sia $G = (V, E, c, s, t)$ una network. Sia G_i il grafo residuo all' i -esima iterazione dell'algoritmo 2, e $G_{i'}$ il grafo residuo all' i' -esima iterazione, con $i' > i$, allora

$$\text{dist}_{G_i}(s, u) \leq \text{dist}_{G_{i'}}(s, u) \quad (1.1)$$

La distanza dal vertice source s rispetto ogni altro vertice aumenta in maniera monotona ad ogni passo dell'algoritmo.

Dimostrazione : Supponiamo che esiste un nodo $v \in G$ tale che $\text{dist}_{G_i}(s, v) > \text{dist}_{G_{i'}}(s, v)$, si assume inoltre che la distanza $\text{dist}_{G_{i'}}(s, v)$ sia la più piccola possibile (v è il nodo più vicino ad s in $G_{i'}$). Sia w il penultimo vertice del cammino $P' = u_1, u_2 \dots u_k$ in $G_{i'}$, con $u_1 = s$ e $u_k = v$. Ne segue che



$$\text{dist}_{G_i}(s, v) > \text{dist}_{G_{i'}}(s, v) = \text{dist}_{G_{i'}}(s, w) + 1 \geq \text{dist}_{G_i}(s, w) + 1 \quad (1.2)$$

Nota : nella dimostrazione si sta assumendo che la proposizione non sia valida per il nodo v , ma che sia valida per il nodo w , da qui è verificata la disuguaglianza a destra nell'equazione 1.2.

Ciò implica che l'arco (w, v) è presente in $G_{i'}$ ma non in G_i , se così non fosse sarebbe vero che $\text{dist}_{G_{i'}}(s, w) \geq \text{dist}_{G_i}(s, w) + 1$, e quindi $w = u_i$ e $v = u_{i+1}$ per qualche i , ma questa è una contraddizione dato che v segue w nel cammino P' , quindi l'asserto è verificato. ■

Teorema 1 Nell'algoritmo di Edmonds-Karp, applicato su una network G , il numero totale di incrementi del valore del flusso è al più $n \cdot m$, con $n = |V(G)|$ e $m = |E(G)|$. Tale affermazione è valida anche se le capacità sugli archi sono numeri reali.

Dimostrazione : Sia G_i il grafo residuo nell' i -esima iterazione dell'algoritmo 2, analogamente, sia f_i il flusso valutato anch'esso durante l' i -esima iterazione. Chiaramente $G_0 = G$ e $f_0(e) = 0$, $\forall e$.

Definizione 5 Durante l'esecuzione dell'algoritmo 2, un'arco (u, v) è detto **critico** in i se

- $(u, v) \in G_i$
- $(u, v) \notin G_{i+1}$

Se P_i è il percorso minimo da s a t considerato nell' i -esima iterazione, e (u, v) è critico in i , per definizione dell'algoritmo si ha che $(u, v) \in E(P_i)$.

Lemma 1 Sia (u, v) un'arco di una network G , durante l'esecuzione dell'algoritmo 2, l'arco (u, v) può essere considerato critico al più $\frac{n}{2}$ volte.

Dimostrazione Lemma : Siano

$$\pi(1) < \pi(2) < \dots < \pi(L)$$

gli indici delle iterazioni in cui (u, v) è critico, con $L \leq \frac{n}{2}$, chiaramente

$$(u, v) \in E(P_{\pi(i)})$$

per qualche $1 \leq i \leq L$. Chiaramente

$$\text{dist}_{G_{\pi(i)}}(s, v) = \text{dist}_{G_{\pi(i)}}(s, u) + 1$$

Se (u, v) è critico in $\pi(i)$ ed in $\pi(i+1)$, allora deve esistere un iterazione i' compresa fra queste

$$\pi(i) < i' < \pi(i+1)$$

In cui l'arco (u, v) è stato re-inserito nel grafo residuo, quindi il flusso su (u, v) in tale iterazione è diminuito, necessariamente (per skew-simmetria) il flusso su (v, u) è aumentato, quindi quest'ultimo arco si trovava sul percorso da s a t nell'iterazione i' .

$$(v, u) \in E(P_{i'})$$

Date le precedenti osservazioni, si deducono le seguenti disuguaglianze

$$\text{dist}_{G_{i'}}(s, u) = \text{dist}_{G_{i'}}(s, v) + 1 \quad (1.3)$$

$$\text{dist}_{G_{i'}}(s, v) + 1 \geq \text{dist}_{G_{\pi(i)}}(s, v) + 1 \quad (1.4)$$

$$\text{dist}_{G_{\pi(i)}}(s, v) + 1 = \text{dist}_{G_{\pi(i)}}(s, u) + 2 \quad (1.5)$$

$$\Downarrow \quad (1.6)$$

$$\text{dist}_{G_{\pi(i+1)}}(s, u) \geq \text{dist}_{G_{\pi(i)}}(s, u) + 2 \quad (1.7)$$

Essendo che la distanza fra due vertici è limitata da $n = |V(G)|$, si ha che

$$\text{dist}_{G_{\pi(i)}}(s, u) \leq n - 1$$

- la distanza fra s ed u è al più $n - 1$
- la distanza fra s ed u aumenta almeno di due in due iterazioni differenti in cui (u, v) è critico

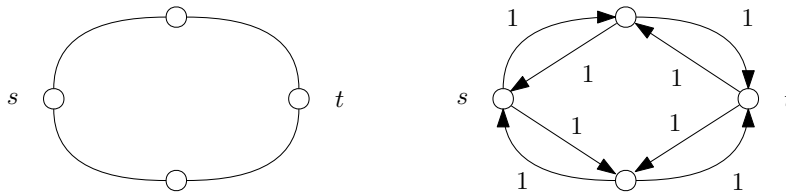
La conclusione è che non possono esistere più di $\frac{n}{2}$ indici $\pi(i)$ in cui (u, v) è critico. \square

La dimostrazione del teorema 1 segue in maniera naturale, ad ogni iterazione un'arco è critico, essendo che ci sono m archi ed ognuno può essere critico al più $\frac{n}{2}$ volte, il numero totale di aumenti del flusso è al più $\frac{1}{2}nm$. \blacksquare

1.4 Cammini Edge-Disjoint in un Grafo

In questa sezione verrà esposta un'applicazione dell'algoritmo di ricerca del flusso massimo. Sia G un grafo non diretto, si definisce la *network associata* a G , il grafo diretto \vec{G} tale che

- $u \in V(G) \implies u \in V(\vec{G})$
- $(u, v) \in E(G) \implies \begin{cases} (u, v) \in E(\vec{G}) \\ (v, u) \in E(\vec{G}) \end{cases}$
- $\forall (u, v) \in \vec{G} \quad , c(u, v) = 1$



Si può anche risalire in maniera naturale ad un grafo non diretto associato ad una network.

Definizione 6 Dato un grafo non diretto G e due nodi s, t , un'insieme di cammini da s a t è **edge-disjoint** se non condividono alcun arco.

Verrà mostrato come, dato un grafo G , il numero massimo di cammini edge-disjoint è uguale al valore del flusso ottimale nella network associata.

Definizione 7 Dato un flusso f per una network \vec{G} , si definisce **supporto del flusso** l'insieme

$$W = \{(u, v) \in E(\vec{G}) \text{ t.c. } f(u, v) \neq 0\}$$

Proposizione 4 Se nel supporto W di un flusso f per una network \vec{G} gli archi compongono k cammini diretti da s a t , allora ci sono k cammini edge-disjoint da s a t nel grafo non diretto associato.

Dimostrazione : **TODO : continuare**

PROGRAMMAZIONE LINEARE

2.1 Insiemi Convessi

La programmazione lineare consiste nella ricerca di un vettore (ingresso di una funzione lineare) in cui tale funzione assume il valore massimo, all'interno di un dominio definito da un'insieme di vincoli lineari. La funzione da massimizzare è detta **funzione obiettivo**, un'esempio di programma lineare può essere il seguente

$$x_1 + x_2$$

soggetto ai vincoli

$$\begin{aligned} x_1 &\geq 0 \\ x_2 &\geq 0 \\ x_2 - x_1 &\leq 1 \\ x_1 + 6x_2 &\leq 15 \\ 4x_1 - x_2 &\leq 10 \end{aligned}$$

Un punto è **ammissibile** se soddisfa tutti i vincoli lineari. L'insieme di punti ammissibili si può rappresentare su un piano in questo caso, essendo un sotto-insieme di \mathbb{R}^2 . La funzione obiettivo essendo lineare



Figura 2.1: Insieme dei punti ammissibili

si può rappresentare come prodotto scalare fra due vettori \mathbf{c} e \mathbf{x}

$$\mathbf{c}^T \mathbf{x} = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 + x_2$$

Può risultare utile rappresentare sul piano anche il vettore \mathbf{c} e la retta equivalente al sottospazio

$$\text{span}(\mathbf{c}) = \{\alpha\mathbf{c} \mid \alpha \in \mathbb{R}\}$$

Si consideri una retta y' perpendicolare alla retta definita da $\text{span}(\mathbf{c})$, i punti di y' che intersecano



l'insieme delle soluzioni ammissibili condividono la stessa immagine se valutati sulla funzione obiettivo. Un'interpretazione geometrica del problema può essere la seguente

Massimizzare la funzione obiettivo equivale a trovare il massimo β tale che l'iperpiano definito da $\mathbf{c}^T \mathbf{x} = \beta$ perpendicolare alla retta $\alpha\mathbf{c}$ interseca l'insieme dei punti ammissibili.

Si definisce **soluzione ottimale** ogni soluzione ammissibile che massimizza la funzione obiettivo. Il

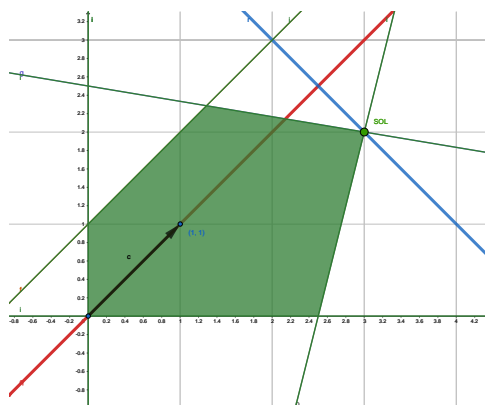


Figura 2.2: Il punto $(3, 2)$ è una soluzione ottimale per il programma lineare

numero di soluzioni di un programma lineare può variare fra i seguenti casi

1. Vi è un'unica soluzione ottimale.
2. Vi sono infinite soluzioni ottimali.
3. Non ci sono soluzioni ammissibili, nessun punto soddisfa tutti i vincoli lineari.
4. Non ci sono soluzioni ottimali perché il problema non è limitato, ciò avviene se l'insieme dei punti ammissibili non è chiuso.

Generalmente la funzione obiettivo $f : \mathbb{R}^n \rightarrow \mathbb{R}$ è del tipo

$$c_1x_1 + c_2x_2 + \cdots + c_nx_n$$

e gli m vincoli sono della forma

$$\begin{aligned} a_{11}x_1 + a_{21}x_2 + \cdots + a_{n1}x_n &\leq b_1 \\ a_{12}x_1 + a_{22}x_2 + \cdots + a_{n2}x_n &\leq b_2 \\ &\vdots \\ a_{1m}x_1 + a_{2m}x_2 + \cdots + a_{nm}x_n &\leq b_m \end{aligned}$$

Osservazione 6 Massimizzare $\mathbf{c}^T \mathbf{x}$ equivale a minimizzare $-\mathbf{c}^T \mathbf{x}$.

Si può quindi assumere che un generico problema di programmazione lineare riguardi la massimizzazione di una funzione del tipo $\mathbf{c}^T \mathbf{x}$ per un fissato $\mathbf{c} \in \mathbb{R}^n$. Inoltre, ogni vincolo del tipo

$$a_{1i}x_1 + a_{2i}x_2 + \cdots + a_{ni}x_n \leq b_i \quad (2.1)$$

è soddisfatto dagli stessi punti che soddisfano

$$-a_{1i}x_1 - a_{2i}x_2 - \cdots - a_{ni}x_n \geq b_i$$

Quindi si può assumere che ogni vincolo sia scritto nella forma 2.1. Inoltre ogni vincolo di uguaglianza equivale a due disuguaglianze. Date le precedenti osservazioni, si può definire un'insieme di m vincoli in maniera compatta tramite una matrice $m \times n$ ed un vettore $\mathbf{b} \in \mathbb{R}^m$.

$$A\mathbf{x} \leq \mathbf{b}$$

Inoltre ogni vincolo di positività del tipo $x_i \geq 0$ equivale al vincolo $-x_i \leq 0$. Se in un programma lineare una variabile x_i può assumere qualsiasi valore in \mathbb{R} , si può sostituire con la differenza di due nuove variabili

$$x_i = z_i - z'_i$$

ed imporre i vincoli

$$z_i, z'_i \geq 0$$

in tal modo è possibile, per ogni variabile del programma lineare, imporre la positività. Ciò è utile per definire una *forma standard* per un programma lineare.

Definizione 8 Un programma lineare in **forma standard** è un problema di ottimizzazione del tipo

$$\begin{aligned} \max \quad & \mathbf{c}^T \mathbf{x} \\ & A\mathbf{x} \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{aligned}$$

Dove

$$\begin{aligned} \mathbf{x}, \mathbf{c} &\in \mathbb{R}^n \\ A &\in \text{Mat}(m \times n) \\ \mathbf{b} &\in \mathbb{R}^m \end{aligned}$$

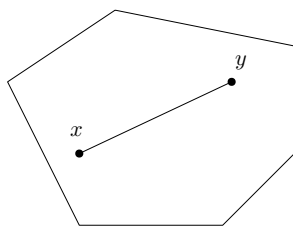
L'esistenza di soluzioni ammissibili dipende esclusivamente dalla matrice A . È impossibile che un programma lineare abbia un numero finito di soluzioni diverso da 0 e 1. In seguito verrà dimostrato che se un programma lineare ha due punti ammissibili che sono soluzione, allora ha infinite soluzioni.

Definizione 9 Dati due punti $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, si definisce **segmento di linea** fra i punti l'insieme

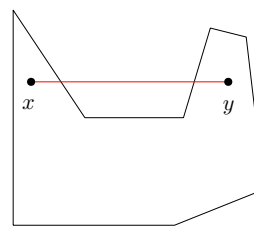
$$\{\alpha \mathbf{x} + (1 - \alpha) \mathbf{y} \text{ t.c. } \alpha \in [0, 1]\}$$

Definizione 10 Un sotto-insieme $X \subseteq \mathbb{R}^n$ è **convesso** se la seguente è verificata

$$\forall \mathbf{x}, \mathbf{y} \in X, \quad \{\alpha \mathbf{x} + (1 - \alpha) \mathbf{y} \text{ t.c. } \alpha \in [0, 1]\} \subseteq X$$



sotto-insieme di \mathbb{R}^2 convesso



sotto-insieme di \mathbb{R}^2 non convesso

Proposizione 5 L'insieme dei punti ammissibili di un programma lineare è convesso.



Dimostrazione : Siano $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ due punti ammissibili, sia $\alpha \in [0, 1]$ fissato, si considera il punto (appartenente al segmento)

$$\alpha \mathbf{x} + (1 - \alpha) \mathbf{y}$$

si ha che

$$A(\alpha \mathbf{x} + (1 - \alpha) \mathbf{y}) = \quad (2.2)$$

$$\alpha A\mathbf{x} + (1 - \alpha) A\mathbf{y} = \quad (2.3)$$

$$\alpha \mathbf{b} + (1 - \alpha) \mathbf{b} = \mathbf{b} \quad (2.4)$$

quindi Il punto soddisfa gli m vincoli

$$A(\alpha \mathbf{x} + (1 - \alpha) \mathbf{y}) \leq \mathbf{b}$$

Inoltre essendo che

$$\alpha \geq 0, \mathbf{x} \geq \mathbf{0}, \mathbf{y} \geq \mathbf{0}$$

si ha che

$$\alpha \mathbf{x} + (1 - \alpha) \mathbf{y} \geq \mathbf{0}$$

Quindi $\alpha \mathbf{x} + (1 - \alpha) \mathbf{y}$ soddisfa tutti i vincoli del programma lineare, ed appartiene quindi ai punti ammissibili. ■

Proposizione 6 *Se per un programma lineare esistono due soluzioni ottimali distinte $\mathbf{x}^*, \mathbf{y}^* \in \mathbb{R}^n$, allora tale programma lineare ha infinite soluzioni ottimali.*

Dimostrazione : Sia \mathbf{z} un punto sul segmento delineato da $\mathbf{x}^*, \mathbf{y}^*$

$$\mathbf{z} = \alpha \mathbf{x}^* + (1 - \alpha) \mathbf{y}^* \quad \text{per qualche } \alpha \in [0, 1]$$

si ha che

$$\mathbf{c}^T \mathbf{z} = \quad (2.5)$$

$$\mathbf{c}^T (\alpha \mathbf{x}^* + (1 - \alpha) \mathbf{y}^*) = \quad (2.6)$$

$$\alpha \mathbf{c}^T \mathbf{x}^* + (1 - \alpha) \mathbf{c}^T \mathbf{y}^* \quad (2.7)$$

Essendo che \mathbf{x}^* e \mathbf{y}^* sono entrambe soluzioni ottimali, si ha che $\mathbf{c}^T \mathbf{x}^* = \mathbf{c}^T \mathbf{y}^*$

$$\alpha \mathbf{c}^T \mathbf{x}^* + (1 - \alpha) \mathbf{c}^T \mathbf{y}^* = \quad (2.8)$$

$$\alpha \mathbf{c}^T \mathbf{x}^* + (1 - \alpha) \mathbf{c}^T \mathbf{x}^* = \mathbf{c}^T \mathbf{x}^* \quad (2.9)$$

Quindi anche \mathbf{z} è soluzione, essendo che quest'ultimo è un generico punto sul segmento, tutti i punti del segmento sono soluzioni. ■

2.2 Applicazioni della Programmazione Lineare

Un classico esempio di applicazione riguarda la scelta di una dieta che soddisfi dei vincoli nutrizionali minimizzando il costo degli alimenti. Vi è un'insieme di n alimenti x_1, \dots, x_n ed ognuno di questi ha un costo c_i , si vuole minimizzare

$$\sum_{i=1}^n c_i x_i$$

tenendo conto dei vincoli nutrizionali del tipo

$$a_{1i} x_1 + a_{2i} x_2 + \dots + a_{ni} x_n \geq b_i$$

Un'altro esempio riguarda la ricerca di un flusso ottimale per una network $G = (V, E, c, s, t)$, per ogni arco (i, j) vi sarà una variabile x_{ij} che rappresenta il flusso su tale arco, si vuole massimizzare la somma del flusso uscente dal vertice source s

$$\sum_{(s,j) \in E(G)} x_{sj}$$

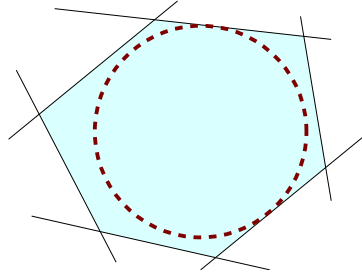


Figura 2.3: Cerchio inscritto nel poligono

I vincoli riguardano la skew simmetria, la conservazione del flusso e il vincolo delle capacità, ad esempio:

$$\begin{aligned} x_{ij} &\leq c_{ij} \\ x_{ij} &= -x_{ji} \end{aligned}$$

Un'esempio interessante di problema che si riconduce alla programmazione lineare riguarda la ricerca del cerchio di raggio massimo che può essere inscritto in un poligono. Si assume che i lati del poligono possono essere descritti da un'insieme di rette del tipo $y = mx + b$ con $m \neq 0$.

Si consideri una delle rette $y = mx + b$, si vuole trovare una retta perpendicolare a questa passante per un fissato punto (x_0, y_0) . Questa ha equazione

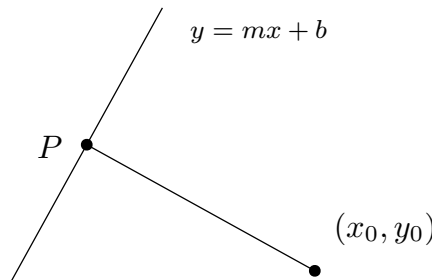
$$y = \frac{y_0 - x}{m} + y_0$$

Il punto di intersezione è $P = (x', y')$ dove

$$x' = \frac{x_0 + my_0}{m^2 + 1} - mb \quad (2.10)$$

$$y' = m \frac{x_0 + my_0 - mb}{m^2 + 1} + b \quad (2.11)$$

La distanza fra P e (x_0, y_0) , applicando la norma euclidea, risulta essere

Figura 2.4: retta perpendicolare passante per (x_0, y_0)

- $\frac{b + mx_0 - y_0}{\sqrt{1 + m^2}}$ se (x_0, y_0) si trova sotto la retta
- $\frac{-b - mx_0 + y_0}{\sqrt{1 + m^2}}$ se (x_0, y_0) si trova sopra la retta

Il programma lineare riguarda la ricerca di un cerchio che massimizzi il raggio r (distanza fra (x_0, y_0) e P) variando fra il possibile centro (x_0, y_0) all'interno del poligono, rispettando i vincoli del tipo

- $\frac{b_l + mx_0 - y_0}{\sqrt{1 + m^2}} \geq r, \quad \forall l \in 1, \dots, k$
- $\frac{-b_l - mx_0 + y_0}{\sqrt{1 + m^2}} \geq r \quad \forall l \in k + 1, \dots, n$

2.3 Il Metodo del Simplexso

Il metodo del simplexso è un'algoritmo per la risoluzione dei programmi lineari. Questi devono però assumere una forma differente da quella standard introdotta nella definizione 8. Un programma lineare è definito come segue

$$\begin{aligned} \max \quad & \mathbf{c}^T \mathbf{x} \\ \text{s.t.} \quad & A\mathbf{x} \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{aligned}$$

Per ogni generica disuguaglianza del tipo

$$a_{1i}x_1 + a_{2i}x_2 + \dots + a_{ni}x_n \leq b_i$$

Si considera una nuova variabile $y_i \geq 0$ detta **slack**, e la disuguaglianza diventa un'uguaglianza come segue

$$a_{1i}x_1 + a_{2i}x_2 + \dots + a_{ni}x_n + y_i = b_i$$

Vi sarà quindi un vettore $\mathbf{y} \in \mathbb{R}^m$ che viene introdotto nel problema, che assume la seguente forma

$$\begin{aligned} \max \quad & [\mathbf{c}^T \mid 0 \dots 0] \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \\ \text{s.t.} \quad & [A \mid \text{Id}_m] \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \mathbf{b} \\ & \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \geq \mathbf{0} \end{aligned}$$

Dove Id_m è la matrice identità $m \times m$. Un problema LP di questo tipo è detto in **forma di equazione**, ed il metodo del simplexso opera su un programma lineare di tale forma. In generale, si scrive

$$\begin{aligned} \max \quad & \mathbf{c}^T \mathbf{x} \\ \text{s.t.} \quad & A\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{aligned}$$

Assunzione : Il programma lineare considerato ha almeno una soluzione ammissibile. Ossia $\exists \bar{\mathbf{x}}$ tale che $A\bar{\mathbf{x}} = \mathbf{b}$ e $\bar{\mathbf{x}} \geq \mathbf{0}$. Data la matrice A , si considerino due righe di essa, ossia A_i e A_j , sia A' un'altra matrice identica a A , in cui la riga j -esima è sostituita con la riga $A_i + A_j$

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{i1} & \dots & a_{in} \\ \vdots & & \vdots \\ a_{j1} & \dots & a_{jn} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \quad A' = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{i1} & \dots & a_{in} \\ \vdots & & \vdots \\ a_{j1} + a_{i1} & \dots & a_{jn} + a_{in} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \quad (2.12)$$

Se \mathbf{x} soddisfa $A\mathbf{x} = \mathbf{b}$ allora soddisfa anche $A'\mathbf{x} = \mathbf{b}'$ dove \mathbf{b}' è identico a \mathbf{b} ma il j -esimo elemento è sommato all' i -esimo

$$\mathbf{b}' = [b_1 \dots b_i \dots b_i + b_j \dots b_n]^T$$

Le operazioni sulle righe di A non cambiano l'insieme delle soluzioni, si assume che le righe siano *linearmente indipendenti*, che il rango di A sia m , e che lo span delle righe di A è un sottospazio di \mathbb{R}^n .

Proposizione 7 *Il rango delle colonne di A è m , la dimensione del sottospazio di \mathbb{R}^m dato dallo span delle colonne è m .*

Dimostrazione : **TODO**

Il problema descritto è

$$A\mathbf{x} = \mathbf{b} \quad (2.13)$$

$$\mathbf{x} \geq \mathbf{0} \quad (2.14)$$

$$\text{esiste una soluzione} \quad (2.15)$$

$$\text{rango righe di } A = \text{rango colonne di } A = m \quad (2.16)$$

Considerando $\ker(A) = \{\mathbf{x} \text{ t.c. } A\mathbf{x} = \mathbf{0}\}$ si ha un sottospazio di dimensione $n - m$ di \mathbb{R}^n , lo spazio delle soluzioni di $A\mathbf{x} = \mathbf{b}$ è un **sottospazio affine** e si ottiene sommando \mathbf{b} a tutti i vettori di $\ker(A)$.

2.3.1 Soluzioni Ammissibili Basiche

Esempio : Si consideri il triangolo delineato dai punti $(0, 0, 3)$, $(0, 3, 0)$, $(3, 0, 0)$, tutti i punti di tale triangolo sono i vettori \mathbf{x} che soddisfano $A\mathbf{x} = \mathbf{b}$. Si ha $A = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$ e $b = 3$

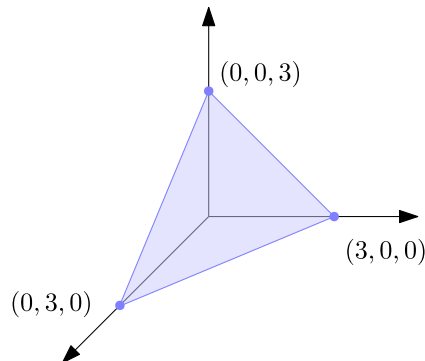


Figura 2.5: triangolo in \mathbb{R}^3

$$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 3 \implies x + y + z = 3$$

I 3 vertici del triangolo rappresentano delle soluzioni "speciali", ottenute ponendo 2 delle 3 variabili uguali a zero e risolvendo per le altre, tali soluzioni sono dette *basiche*.

In generale, $A \in \text{Mat}(m \times n)$ e $\mathbf{b} \in \mathbb{R}^m$, si vogliono porre un numero massimale di variabili pari a zero e risolvere per le altre, si pongono esattamente $n - m$ variabili nulle.

Sia $\mathcal{B} \subseteq \{1, \dots, n\}$ l'insieme degli indici i per cui la variabile x_i non è posta a zero, e sia $A_{\mathcal{B}}$ la sotto-matrice di A composta dalle colonne il cui indice è relativo ad una variabile diversa da zero.

$$\begin{bmatrix} a_{1i} & \dots & a_{mi} \end{bmatrix}^T \text{ è una colonna di } A_{\mathcal{B}} \iff i \in \mathcal{B}$$

Sia $\mathbf{b}_{\mathcal{B}}$ definito in maniera intuitiva, come il vettore contenente la componente j -esima di \mathbf{b} se $j \in \mathcal{B}$. Si considera il problema derivato

$$A_{\mathcal{B}}\mathbf{x}_{\mathcal{B}} = \mathbf{b}_{\mathcal{B}} \quad A_{\mathcal{B}} \in \text{Mat}(m \times m) \quad (2.17)$$

Se $A_{\mathcal{B}}$ non è **singolare** (il determinante è diverso da zero) allora esiste una soluzione $\bar{\mathbf{x}}$ per il problema derivato, se $\bar{\mathbf{x}} \geq 0$, allora è anche una soluzione per il problema originale, ponendo a zero tutte le componenti della soluzione i cui indici non sono in \mathcal{B} .

- se $\bar{\mathbf{x}} = [x_{k1} \ x_{k2} \ \dots \ x_{km}]^T$ è soluzione di $A_{\mathcal{B}}\mathbf{x}_{\mathcal{B}} = \mathbf{b}_{\mathcal{B}}$
- si ha $\{k1, k2, \dots, km\} = \mathcal{B}$
- allora il vettore $\bar{\mathbf{x}}'$ con
 - $x_i = x_{kj}$ se $i = kj$ per qualche $j \leq m$, ossia $i \in \mathcal{B}$
 - $x_i = 0$ se $i \notin \mathcal{B}$

è soluzione di $A\mathbf{x} = \mathbf{b}$ se $\bar{\mathbf{x}}' \geq 0$.

Definizione 11 Dato un programma lineare

$$\begin{aligned} \max \quad & \mathbf{c}^T \mathbf{x} \\ \text{s.t.} \quad & A\mathbf{x} \leq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{aligned}$$

Una soluzione ammissibile $\bar{\mathbf{x}}$ è detta **basica** (BFS) se esiste un sotto-insieme di indici $\mathcal{B} \subseteq \{1, 2, \dots, n\}$ tale per cui

- $A_{\mathcal{B}}$ non è singolare
- $\bar{x}_j = 0 \ \forall j \notin \mathcal{B}$

Esempio : Data la matrice

$$A = \begin{bmatrix} 1 & 6 & 3 & -3 & 4 \\ 0 & 1 & 3 & 2 & 4 \end{bmatrix} \quad (2.18)$$

Ed il sistema di equazioni

$$A\mathbf{x} = \begin{bmatrix} 6 \\ 6 \end{bmatrix}$$

Per la base $\mathcal{B} = \{2, 4\}$ si ha

$$A_{\mathcal{B}} = \begin{bmatrix} 6 & -3 \\ 1 & 2 \end{bmatrix}$$

La matrice non è singolare, per il sistema

$$\begin{bmatrix} 6 & -3 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_2 \\ x_4 \end{bmatrix} = \begin{bmatrix} 6 \\ 6 \end{bmatrix}$$

una soluzione è

$$\begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

Quindi una soluzione ammissibile basica per il problema originale è

$$\bar{\mathbf{x}} = [0 \quad 2 \quad 0 \quad 2 \quad 0]^T$$

Se considerassimo la base $\mathcal{B}' = \{3, 5\}$, si avrebbe la sotto-matrice

$$A_{\mathcal{B}'} = \begin{bmatrix} 3 & 4 \\ 3 & 4 \end{bmatrix}$$

Essendo singolare, per tale base non esiste alcuna soluzione basica.

Proposizione 8 Sia $A\mathbf{x} = \mathbf{b}$, $\mathbf{x} \geq \mathbf{0}$ un programma lineare, e sia $\bar{\mathbf{x}}$ una BFS (soluzione ammissibile basica), si consideri l'insieme di indici

$$K = \{i \text{ t.c. } \bar{x}_i > 0\}$$

Le colonne della matrice A_K sono linearmente indipendenti.

Dimostrazione : La dimostrazione è banale, esiste una base \mathcal{B} a cui $\bar{\mathbf{x}}$ fa riferimento per cui $K \subseteq \mathcal{B}$, essendo le colonne di $A_{\mathcal{B}}$ linearmente indipendenti, anche un loro sotto-insieme, ossia quelle di A_K lo sono. ■

Proposizione 9 Si consideri un programma lineare, sia $\bar{\mathbf{x}}$ una soluzione ammissibile e

$$K = \{i \text{ t.c. } \bar{x}_i > 0\}$$

$\bar{\mathbf{x}}$ è una BFS se e solo se le colonne di A_K sono linearmente indipendenti.

Dimostrazione : Verrà dimostrata solo una delle due implicazioni, si assume che $\bar{\mathbf{x}}$ è una soluzione ammissibile e che le colonne di A_K siano linearmente indipendenti. Per dimostrare che $\bar{\mathbf{x}}$ sia una BFS è necessario trovare una base che contenga K .

Sia \mathcal{B}^* un'insieme definito come segue

1. $K \subseteq \mathcal{B}^*$
2. Le colonne di \mathcal{B}^* sono linearmente indipendenti
3. \mathcal{B}^* è l'insieme massimale che rispetti le due condizioni precedenti

Bisogna mostrare che \mathcal{B}^* ha m elementi in modo tale che $A_{\mathcal{B}^*}$ sia una matrice quadrata $m \times m$.

Si assuma che $|\mathcal{B}^*| < m$, ciò significherebbe che le restanti colonne di A sono contenute nello span delle colonne di $A_{\mathcal{B}^*}$, quindi A ha meno di m colonne linearmente indipendenti, ma è noto che il rango di A sia m , quindi l'assunzione $|\mathcal{B}^*| < m$ porta ad una contraddizione $\implies |\mathcal{B}^*| = m$. ■

Osservazione 7 Ogni base ha almeno una BFS associata.

Conclusione : Esiste un numero finito di BFS. Ogni base è un sotto-insieme di m elementi preso da $\{1, 2, \dots, n\}$, quindi esistono al più $\binom{n}{m}$ basi distinte, quindi esistono al più $\binom{n}{m}$ BFS.

Il seguente risultato è fondamentale in quanto stabilisce che un programma lineare (che riguarda la ricerca di una soluzione fra un'insieme non numerabile) si può ridurre ad una ricerca fra un'insieme finito di possibili soluzioni ottimali.

Teorema 2 Si consideri un programma lineare in forma di equazione

$$\begin{aligned} \max \quad & \mathbf{c}^T \mathbf{x} \\ \text{s.t.} \quad & A\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{aligned}$$

1. Se esiste almeno una soluzione ammissibile ed il problema è limitato (l'insieme delle soluzioni ammissibili è compatto), allora esiste una soluzione ottimale.
2. Se esiste una soluzione ottimale, allora esiste anche una BFS che è a sua volta ottimale.

La dimostrazione del teorema 2 richiede alcuni passi preliminari.

Statement \star : Se la funzione obiettivo di un LP in forma di equazione è limitata superiormente, allora per ogni soluzione ammissibile \mathbf{y} esiste una BFS \mathbf{z} tale che $\mathbf{c}^T \mathbf{z} \geq \mathbf{c}^T \mathbf{y}$.

Claim 1 : Se \star è vera, allora il teorema è dimostrato.

Dimostrazione Claim 1 : Si consideri la BFS \mathbf{y} che massimizza la funzione obiettivo rispetto tutte le altre BFS. Assumiamo che \mathbf{y} non sia ottimale, allora esiste \mathbf{y}^* (che non è una BFS) tale che

$$\mathbf{c}^T \mathbf{y}^* \geq \mathbf{c}^T \mathbf{y}$$

Ma per il claim \star esiste una BFS \mathbf{z} tale per cui

$$\mathbf{c}^T \mathbf{z} \geq \mathbf{c}^T \mathbf{y}^* \geq \mathbf{c}^T \mathbf{y}$$

Ciò va in contraddizione con l'assunzione che \mathbf{y} sia massimale fra le BFS, quindi non può esistere tale $\mathbf{y}^* \implies \mathbf{y}$ è una BFS ottimale \implies il claim è dimostrato : $\star \implies$ teorema 2. ■

Per dimostrare il teorema è quindi sufficiente dimostrare lo Statement \star . Sia \mathbf{y} un'arbitraria soluzione ammissibile, sia \mathbf{z} un'altra soluzione ammissibile tale che

1. $\mathbf{c}^T \mathbf{z} \geq \mathbf{c}^T \mathbf{y}$
2. \mathbf{z} ha un numero massimale di componenti uguali a zero

Tale \mathbf{z} esiste dato che nel caso peggiore si può considerare $\mathbf{z} = \mathbf{y}$. Si consideri l'insieme K degli indici delle componenti positive di \mathbf{z}

$$K = \{j \text{ t.c. } z_j > 0\}$$

La proposizione 8 afferma che se le colonne di A_K sono linearmente indipendenti \mathbf{z} è una BFS, ciò dimostrerebbe lo Statement \star .

Si assuma che le colonne di A_K siano linearmente dipendenti. Sia $k = |K|$. Esistono dei coefficienti $\alpha_1, \dots, \alpha_k$ tali per cui

$$\alpha_1 \begin{bmatrix} A_K^1 \end{bmatrix} + \dots + \alpha_k \begin{bmatrix} A_K^k \end{bmatrix} = \mathbf{0} \implies \quad (2.19)$$

$$A_K \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{bmatrix} = \mathbf{0} \quad (2.20)$$

Si consideri un particolare vettore \mathbf{w} definito come segue



- la componente j -esima di \mathbf{w} contiene α_j se $j \in K$, α_j è il coefficiente moltiplicato alla colonna j -esima di A_K nella combinazione lineare dell'equazione 2.19.
- Se $j \notin K$, allora la componente j -esima di \mathbf{w} è 0.

Data l'equazione 2.20, risulta chiaro che il prodotto della matrice originale A e \mathbf{w} sia uguale al prodotto fra A_K ed il vettore dei coefficienti $\alpha_1 \dots, \alpha_k$.

$$A\mathbf{w} = A_K \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{bmatrix} = \mathbf{0} \quad (2.21)$$

Riconsiderando la soluzione \mathbf{z} , si può sommare a \mathbf{w} e moltiplicare per la matrice A

$$A(\mathbf{z} + \mathbf{w}) = A\mathbf{z} + A\mathbf{w} = A\mathbf{z} + \mathbf{0} = \mathbf{b} + \mathbf{0} = \mathbf{b}$$

Essendo che \mathbf{z} è soluzione il prodotto fra quest'ultimo ed A è uguale al vettore \mathbf{b} relativo ai vincoli del programma lineare. Nonostante $\mathbf{z} + \mathbf{w}$ soddisfi i vincoli, non è certo che sia soluzione, in quanto non è certo se abbia o no componenti negative (si ricordi che una soluzione deve essere maggiore o uguale al vettore $\mathbf{0}$).

Si noti come per definizione di \mathbf{w} , questo ha una componente uguale a zero in una data posizione, se in quella data posizione la soluzione \mathbf{z} ha la componente uguale a zero, quindi \mathbf{w} ha *almeno tanti* zeri quanti quelli di \mathbf{z} .

Claim 2 : Il vettore \mathbf{w} soddisfa le seguenti

1. $\mathbf{c}^T \mathbf{w} \geq 0$
2. $\exists j \in K$ t.c. $w_j < 0$

Dimostrazione : Se per \mathbf{w} il punto (2) non dovesse essere soddisfatto, si potrebbe considerare $-\mathbf{w}$, questo soddisfa tutte le condizioni per cui è stato definito \mathbf{w} : $A(-\mathbf{w}) = A\mathbf{w} = \mathbf{0}$. Tale sostituzione si può considerare anche per soddisfare il punto (1), nel caso in cui si dovesse verificare che $\mathbf{c}^T \mathbf{w} < 0$.

Se la condizione (2) non dovesse essere soddisfatta, allora si consideri il vettore (in funzione di $t \in \mathbb{R}^+$)

$$\mathbf{z}(t) = \mathbf{z} + t\mathbf{w}$$

Tale vettore soddisfa i vincoli del programma lineare

$$A\mathbf{z}(t) = A(\mathbf{z} + t\mathbf{w}) = A\mathbf{z} + tA\mathbf{w} = \mathbf{b} + t \cdot \mathbf{0} = \mathbf{b}$$

Inoltre essendo la (2) non soddisfatta, si ha che, per ogni possibile $t \in \mathbb{R}^+$

$$z(t)_j = z_j + tw_j \geq 0$$

questo perché

1. $z_j \geq 0$ essendo \mathbf{z} una soluzione
2. t è non negativo perché varia in \mathbb{R}^+
3. w_j non è negativo dato che la (2) non è soddisfatta

Essendo $\mathbf{z}(t)$ sempre maggiore o uguale a zero si ha:

$$\mathbf{c}^T \cdot \mathbf{z}(t) = \mathbf{c}^T (\mathbf{z} + t\mathbf{w}) = \mathbf{c}^T \mathbf{z} + t\mathbf{c}^T \mathbf{w} \quad (2.22)$$

Attenzione, essendo il punto (1) soddisfatto, $\mathbf{c}^T \mathbf{w}$ è positivo, quindi dato il termine $t\mathbf{c}^T \mathbf{w}$, la funzione $\mathbf{c}^T \cdot \mathbf{z}(t)$ non è limitata, ciò è una contraddizione, quindi le due assunzioni del claim sono vere. ■

Dato ciò, esiste un indice $j \in K$ per cui $w_j < 0$, si consideri nuovamente il vettore

$$\mathbf{z}(t) = \mathbf{z} + t\mathbf{w}$$

Avendo verificato che questo soddisfa i vincoli $\forall t$, ci si pone la domanda, che valore assume nella funzione obiettivo?

$$\mathbf{c}^T \mathbf{z}(t) = \mathbf{c}^T \mathbf{z} + t\mathbf{c}^T \mathbf{w}$$

Sappiamo che

- $\mathbf{c}^T \mathbf{z} \geq \mathbf{c}^T \mathbf{y}$ per la scelta di \mathbf{z}
- $t\mathbf{c}^T \mathbf{w} \geq 0$ per il claim

Ne consegue che

$$\mathbf{c}^T \mathbf{z}(t) \geq \mathbf{c}^T \mathbf{y}$$

Per la chiarezza della dimostrazione, si ricordi che \mathbf{z} è stato scelto per soddisfare le seguenti

1. $\mathbf{c}^T \mathbf{z} \geq \mathbf{c}^T \mathbf{y}$
2. \mathbf{z} ha un numero massimale di componenti uguali a zero

$$\mathbf{z}(t) = \begin{bmatrix} 0 \\ \vdots \\ z_i \\ \vdots \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ \vdots \\ w_i \\ \vdots \\ 0 \end{bmatrix}$$

sia w_i una componente negativa di \mathbf{w} , se $t = -\frac{z_i}{w_i}$ allora

$$\mathbf{z}\left(-\frac{z_i}{w_i}\right) = \begin{bmatrix} 0 \\ \vdots \\ z_i \\ \vdots \\ 0 \end{bmatrix} - \frac{z_i}{w_i} \begin{bmatrix} 0 \\ \vdots \\ w_i \\ \vdots \\ 0 \end{bmatrix} \implies z\left(-\frac{z_i}{w_i}\right)_i = z_i - z_i = 0$$

Il vettore $\mathbf{z}\left(-\frac{z_i}{w_i}\right)$ ha una componente nulla in più rispetto a \mathbf{z} , se si considera

$$t^* = \min_{i \text{ t.c. } w_i < 0} \left(-\frac{z_i}{w_i}\right)$$

il vettore $\mathbf{z}(t^*)$ risulterà non negativo, quindi:

$$\begin{aligned} A\mathbf{z}(t^*) &= \mathbf{b} \\ \mathbf{z}(t^*) &\geq 0 \end{aligned}$$

Il vettore $\mathbf{z}(t^*)$ è una soluzione ammissibile ed ha uno 0 in più rispetto a \mathbf{z} , ma questa è una contraddizione (data la scelta di \mathbf{z}), quindi è impossibile che le colonne di A_K siano linearmente dipendenti $\implies \mathbf{z}$ è una BFS, lo Statement \star è dimostrato. ■

Corollario : Se un programma lineare ha una funzione obiettivo limitata, allora esiste un'algoritmo che trova una soluzione ottimale in un numero finito di passi.

2.3.2 Ricerca sul Poliedro

Si consideri il seguente programma lineare

$$\begin{aligned} \max \quad & x_1 + x_2 \\ -x_1 + x_2 &\leq 1 \\ x_1 &\leq 3 \\ x_2 &\leq 2 \\ x_1, x_2 &\geq 0 \end{aligned}$$

L'insieme delle soluzioni ammissibili è mostrato in figura 2.3.2. Per trasformare il problema in forma di equazione occorre aggiungere 3 variabili slack x_3, x_4, x_5 .

$$\begin{aligned} \max \quad & \mathbf{c}^T \mathbf{x} \\ -x_1 + x_2 + x_3 &= 1 \\ x_1 + x_4 &= 3 \\ x_2 + x_5 &= 2 \\ x_1, x_2, x_3, x_4, x_5 &\geq 0 \end{aligned}$$



Figura 2.6: Insieme definito dai vincoli

in forma matriciale:

$$A = \begin{bmatrix} -1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix} \quad \mathbf{c} = [1 \quad 1 \quad 0 \quad 0 \quad 0]^T$$

Si consideri la base $\mathcal{B} = \{1, 2, 4\}$:

$$A_{\mathcal{B}} = \begin{bmatrix} -1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

è non singolare, si può risolvere il sistema di equazioni

$$\begin{bmatrix} -1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \implies \begin{cases} x_1 = 1 \\ x_2 = 2 \\ x_4 = 2 \end{cases} \quad (2.23)$$

La BFS è $\mathbf{x} = [1 \quad 2 \quad 0 \quad 2 \quad 0]^T$. Le variabili slack x_3, x_5 della soluzione \mathbf{x} sono poste a zero, ciò significa che tale punto soddisfa i vincoli del problema originale (in forma di uguaglianza), in particolare, soddisfa le uguaglianze del primo e del terzo vincolo

$$\begin{aligned} -x_1 + x_2 &= 1 \\ x_2 &= 2 \end{aligned}$$

Tale punto può essere geometricamente individuato nell'insieme delle soluzioni ammissibili del problema originale, ed equivale al punto su cui si intersecano le due rette che definiscono il primo ed il terzo vincolo, come mostrato in figura 2.3.2.



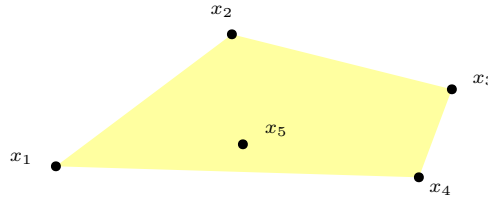
Figura 2.7: punto di intersezione

Definizione 12 Un'insieme $\mathcal{X} \subset \mathbb{R}^n$ è **aperto** se $\forall \mathbf{x} \in \mathcal{X}, \exists \epsilon > 0$ tale che $\{\mathbf{y} \in \mathbb{R}^n \mid \|\mathbf{y} - \mathbf{x}\| \leq \epsilon\} \subset \mathcal{X}$.

Definizione 13 Un'insieme $\mathcal{X} \subset \mathbb{R}^n$ è **chiuso** se il suo complemento $\mathbb{R}^n \setminus \mathcal{X}$ è un'insieme aperto.

Osservazione 8 L'unione di due insiemi aperti è un'insieme aperto. L'intersezione di due insiemi chiusi è un'insieme chiuso.

Definizione 14 Dato un sotto-insieme di punti $I \subset \mathbb{R}^n$, si definisce il suo **inviluppo convesso** l'intersezione di tutti i sotto-insiemi convessi di \mathbb{R}^n contenenti I . Alternativamente, si può dire che l'inviluppo convesso di I è il più piccolo insieme convesso contenente I .

Figura 2.8: inviluppo convesso dei punti x_1, x_2, \dots, x_5

Definizione 15 Dati n punti $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in \mathbb{R}^m$, una loro **combinazione convessa** è un punto \mathbf{z} definito come segue

- $\mathbf{z} = \sum_{i=1}^n \alpha_i \mathbf{x}_i$
- $\alpha_i \geq 0 \quad \forall i$
- $\sum_{i=1}^n \alpha_i = 1$

Osservazione 9 Una combinazione convessa fra due punti è un segmento di linea.

Proposizione 10 Siano $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ dei punti in \mathbb{R}^m , sia C l'inviluppo convesso di tali punti, e sia \tilde{C} l'insieme di tutte le combinazioni convesse

$$\tilde{C} = \left\{ \sum_{i=1}^n \alpha_i \mathbf{x}_i \text{ t.c. } \alpha_i \geq 0 \quad \forall i, \quad \sum_{i=1}^n \alpha_i = 1 \right\}$$

si ha che $C = \tilde{C}$

Dimostrazione : La dimostrazione procederà classicamente con una doppia inclusione. Si vuole mostrare come prima cosa che $C \subseteq \tilde{C}$, essendo C l'intersezione di tutti gli insiemi convessi contenenti i punti, ed essendo che \tilde{C} contiene ogni punto \mathbf{x}_i , è sufficiente mostrare che \tilde{C} sia convesso.

Siano $\mathbf{z}_1, \mathbf{z}_2 \in \tilde{C}$, ossia della forma

$$\begin{aligned} \mathbf{z}_1 &= \sum_i \alpha_i \mathbf{x}_i, & \alpha_i &\geq 0, \quad \sum_i \alpha_i = 1 \\ \mathbf{z}_2 &= \sum_i \beta_i \mathbf{x}_i, & \beta_i &\geq 0, \quad \sum_i \beta_i = 1 \end{aligned}$$

Un generico punto sul segmento di $\mathbf{z}_1, \mathbf{z}_2$ è

$$t\mathbf{z}_1 + (1-t)\mathbf{z}_2, \quad t \in [0, 1]$$

esplicitando

$$t \sum_i \alpha_i \mathbf{x}_i + (1-t) \sum_i \beta_i \mathbf{x}_i = \tag{2.24}$$

$$\sum_i t\alpha_i \mathbf{x}_i + \sum_i (1-t)\beta_i \mathbf{x}_i = \tag{2.25}$$

$$\sum_i (t\alpha_i + (1-t)\beta_i) \mathbf{x}_i \tag{2.26}$$

Bisogna mostrare che $\sum_i (t\alpha_i + (1-t)\beta_i) \mathbf{x}_i$ è una combinazione convessa, essendo che

$$t \geq 0, \quad \alpha_i \geq 0, \quad \beta_i \geq 0$$

è immediato che, per ogni i si ha che $t\alpha_i + (1-t)\beta_i \geq 0$, inoltre

$$\sum_i t\alpha_i + (1-t)\beta_i = \tag{2.27}$$

$$\sum_i t\alpha_i + \sum_i (1-t)\beta_i = \tag{2.28}$$

$$t \sum_i \alpha_i + (1-t) \sum_i \beta_i \tag{2.29}$$

Essendo che per ipotesi $\sum_i \alpha_i = \sum_i \beta_i = 1$, si ha che

$$t \sum_i \alpha_i + (1-t) \sum_i \beta_i = t + (1-t) = 1 \quad (2.30)$$

Quindi ogni punto sul segmento $t\mathbf{z}_1 + (1-t)\mathbf{z}_2$, $t \in [0, 1]$ è a sua volta una combinazione convessa di $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \implies \tilde{C}$ è convesso $\implies C \subseteq \tilde{C}$.

Si vuole mostrare ora che $\tilde{C} \subseteq C$, sia \mathbf{z} una combinazione convessa, $\mathbf{z} = \sum_i \alpha_i \mathbf{x}_i$, si procede per induzione sul numero di coefficienti α_i il cui valore è diverso da zero.

- *Caso base* : Solamente un coefficiente α_i è diverso da zero, sia questo α_j (j fissato), allora

$$\sum_i \alpha_i \mathbf{x}_i = 0 \cdot \mathbf{x}_0 + \dots 1 \cdot \mathbf{x}_j + \dots 0 \cdot \mathbf{x}_n = \mathbf{x}_j$$

Essendo che i punti sono contenuti nell'involuppo convesso, si ha che $\mathbf{z} \in C$

- *Secondo caso base* : Se il numero di coefficienti di \mathbf{z} diversi da zero fosse 2, allora la combinazione convessa sarebbe del tipo

$$\mathbf{z} = t\mathbf{x}_i + (1-t)\mathbf{x}_j, \quad \text{per qualche } i, j$$

anche in questo caso \mathbf{z} si troverebbe in C , dato che C è convesso e $\mathbf{x}_i, \mathbf{x}_j \in C$, come mostrato nel caso base.

- *Ipotesi induttiva* : Le combinazioni convesse con $k-1$ coefficienti diversi da zero sono in C .
- *Passo induttivo* : Sia \mathbf{z} una combinazione convessa con k coefficienti diversi da zero. È sufficiente mostrare che \mathbf{z} si trova sul segmento di linea di due punti contenuti in C . Sia j un fissato indice tale per cui il coefficiente α_j di \mathbf{z} è diverso da zero, per definizione si ha che

$$\sum_{i=1, i \neq j}^n \alpha_i = 1 - \alpha_j$$

Questo comporta che

$$\frac{1}{1 - \alpha_j} \sum_{i=1, i \neq j}^n \alpha_i = 1$$

Si considerino n coefficienti $\alpha'_1, \dots, \alpha'_n$ definiti come segue

$$\alpha'_i = \begin{cases} 0 & \text{se } i = j \\ \frac{\alpha_i}{1 - \alpha_j} & \text{altrimenti} \end{cases}$$

Si consideri la combinazione convessa \mathbf{z}' data da tali coefficienti

$$\mathbf{z}' = \sum_{i=1}^n \alpha'_i \mathbf{x}_i$$

Chiaramente, \mathbf{z}' ha $k-1$ coefficienti diversi da zero, quindi per ipotesi induttiva $\mathbf{z}' \in C$. Si consideri il seguente punto sul segmento fra \mathbf{z}' e \mathbf{x}_j :

$$(1 - \alpha_j)\mathbf{z}' + \alpha_j \mathbf{x}_j$$

Essendo sul segmento, tale punto è contenuto in C , esplicitando:

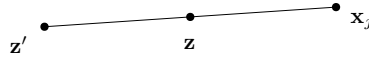
$$(1 - \alpha_j)\mathbf{z}' + \alpha_j \mathbf{x}_j = \quad (2.31)$$

$$(1 - \alpha_j) \sum_{i=1}^n \alpha'_i \mathbf{x}_i + \alpha_j \mathbf{x}_j = \quad (2.32)$$

$$\sum_{i=1}^n (1 - \alpha_j) \alpha'_i \mathbf{x}_i + \alpha_j \mathbf{x}_j = \quad (2.33)$$

$$\sum_{i=1, i \neq j}^n \alpha_i \mathbf{x}_i + \alpha_j \mathbf{x}_j = \mathbf{z} \quad (2.34)$$

Ma allora \mathbf{z} si trova sul segmento di linea fra \mathbf{z}' e \mathbf{x}_j , quindi $\mathbf{z} \in C \implies \tilde{C} \subseteq C$, questo completa la dimostrazione, $C = \tilde{C}$. ■

Figura 2.9: locazione geometrica di \mathbf{z}

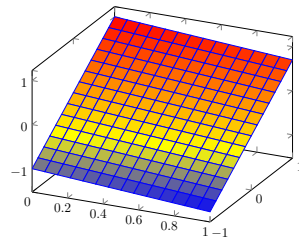
Definizione 16 Un **iperpiano** in \mathbb{R}^n è un sottospazio affine di dimensione $n - 1$ definito dall'insieme dei punti che soddisfano un'equazione del tipo

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

Ogni iperpiano definisce due *mezzi spazi*, ossia due insiemi convessi e chiusi, la cui unione comprende tutto \mathbb{R}^n , definiti dai punti che soddisfano le equazioni

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n \leq b$$

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n \geq b$$

Figura 2.10: iperpiano in \mathbb{R}^3

Definizione 17 Un **poliedro** è l'intersezione di un numero finito di mezzi-spazi definiti da iperpiani. La dimensione del poliedro P è uguale alla dimensione del più piccolo sotto-spazio affine di \mathbb{R}^n contenente P .

Definizione 18 Un **politopo** P è un poliedro limitato, ossia, $\exists c \in \mathbb{R}$ tale che $\forall \mathbf{x} \in P, \|\mathbf{x}\| \leq c$.

Definizione 19 Dato un poliedro P , un punto \mathbf{v} è un **vertice** di P se esiste un iperpiano

$$\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n \text{ t.c. } a_1x_1 + a_2x_2 + \cdots + a_nx_n = b\}$$

tale per cui

- $\mathbf{v} \in \mathcal{X}$
- $\forall \bar{\mathbf{x}} \in P, \bar{\mathbf{x}} = [\bar{x}_1, \dots, \bar{x}_n]^T \neq \mathbf{v}$ si ha

$$a_1\bar{x}_1 + a_2\bar{x}_2 + \cdots + a_n\bar{x}_n < b$$

Definizione 20 Dato un poliedro P , una **faccia** è un'insieme $\mathcal{U} \subseteq P$ tale per cui esiste un iperpiano \mathcal{X} tale che $\mathcal{X} \cap P = \mathcal{U}$, e per cui, ogni punto di P si trova in uno dei due mezzi spazi definiti da \mathcal{X}

La dimensione di una faccia è uguale alla dimensione del più piccolo sottospazio affine che la contiene.

Definizione 21 Un **angolo** è una faccia di dimensione 1.

Teorema 3 Se P è il poliedro rappresentante l'insieme delle soluzioni ammissibili di un programma lineare in forma di equazione, allora \mathbf{v} è un vertice di P se e solo se \mathbf{v} è una soluzione ammissibile basica per il programma lineare.



Dimostrazione : Verranno dimostrati entrambe le implicazione del *se e solo se*.

\Rightarrow : Si vuole dimostrare che il vertice di un poliedro è soluzione di un programma lineare. Sia \mathbf{v} un vertice di P , per definizione, esiste un iperpiano $a_1x_1 + \dots + a_nx_n = b$ tale per cui P è contenuto in una delle due metà definite da esso, sia questa $a_1x_1 + \dots + a_nx_n \leq b$, inoltre

$$a_1v_1 + \dots + a_nv_n = b$$

e

$$a_1x_1 + \dots + a_nx_n < b, \quad \forall \mathbf{x} \in P, \quad \mathbf{x} \neq \mathbf{v}$$

Quindi, considerando il programma lineare

$$\max [a_1 \ a_2 \ \dots \ a_n] \mathbf{x} \quad (2.35)$$

$$[a_1 \ a_2 \ \dots \ a_n] \mathbf{x} \leq b \quad (2.36)$$

$$\mathbf{x} \in P \quad (2.37)$$

\mathbf{v} è l'unica soluzione ottima, per il teorema 2, allora \mathbf{v} è una BFS.

\Leftarrow : Sia \mathbf{v} una BFS per un dato programma lineare, il cui insieme delle soluzioni ammissibili è il poliedro P , a \mathbf{v} è associata una base $\mathcal{B} \subset \{1, 2, \dots, n\}$, si consideri vettore $\tilde{\mathbf{c}}$ definito come segue

$$\tilde{\mathbf{c}} = \begin{bmatrix} \tilde{c}_1 \\ \vdots \\ \tilde{c}_n \end{bmatrix} \quad \tilde{c}_i = \begin{cases} 0 & \text{se } i \in \mathcal{B} \\ -1 & \text{se } i \notin \mathcal{B} \end{cases} \quad (2.38)$$

Essendo che ogni componente i -esima di \mathbf{v} è nulla se $i \notin \mathcal{B}$, è immediato che

$$\tilde{\mathbf{c}}^T \mathbf{v} = 0$$

Inoltre, preso un qualsiasi altro punto $\mathbf{x} \in P$, se $\exists j$ tale che $x_j > 0$, allora $\tilde{\mathbf{c}}^T \mathbf{x} < 0$.

Osservazione 10 $\forall \mathbf{x} \in P, \quad \tilde{\mathbf{c}}^T \mathbf{x} \leq 0$

Quindi, considerando l'iperpiano

$$\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n \text{ t.c. } \tilde{\mathbf{c}}^T \mathbf{x} = 0\}$$

si noti come

- Per l'osservazione 10 tutti i punti del poliedro P sono contenuti in una delle metà definite dall'iperpiano \mathcal{X}
- $\tilde{\mathbf{c}}^T \mathbf{v} = 0$

Se \mathbf{v} fosse l'unico punto per cui $\tilde{\mathbf{c}}^T \mathbf{v} = 0$ allora sarebbe per definizione un vertice di P . Si assuma che esiste un $\mathbf{y} \in P$ tale per cui $\tilde{\mathbf{c}}^T \mathbf{y} = 0$, ciò significa che $\forall j \notin \mathcal{B}, y_j = 0$, questo significa che, considerando la matrice $A_{\mathcal{B}}$, ed il sistema di equazioni

$$A_{\mathcal{B}} \mathbf{x}_{\mathcal{B}} = \mathbf{b}_{\mathcal{B}}$$

il vettore $\mathbf{y}_{\mathcal{B}}$ (sotto-vettore di \mathbf{y} le cui componenti sono quelle di indice contenuto in \mathcal{B}) è soluzione, ma anche $\mathbf{v}_{\mathcal{B}}$ è soluzione, essendo $A_{\mathcal{B}}$ una matrice quadrata non singolare, il sistema ammette un'unica soluzione, quindi $\mathbf{v}_{\mathcal{B}} = \mathbf{y}_{\mathcal{B}} \implies \mathbf{v} = \mathbf{y}$, ma allora \mathbf{v} è l'unica soluzione per cui $\tilde{\mathbf{c}}^T \mathbf{v} = 0$, ne consegue che è un vertice. ■

2.3.3 La Procedura di Risoluzione

Le proposizioni ed i teoremi presentati nei paragrafi precedenti dovrebbero aver fornito un'idea di come un programma lineare si riduce alla ricerca delle soluzioni ottimali fra i vertici del poliedro definito dall'insieme delle soluzioni ammissibili.

Si consideri il programma lineare definito all'inizio della sezione 2.3.2

$$\begin{aligned} \max \quad & x_1 + x_2 \\ \text{s.t.} \quad & -x_1 + x_2 \leq 1 \\ & x_1 \leq 3 \\ & x_2 \leq 2 \\ & x_1, x_2 \geq 0 \end{aligned}$$



Il poliedro in questione è mostrato in figura 2.3.2. La soluzione ottimale si trova sul vertice in alto a destra, ossia $\mathbf{x} = [3, 2]^T$, presenteremo la procedura del metodo del simplesso su tale programma lineare.

Prima di procedere è necessario aggiungere delle variabili slack e trasformare il problema in forma di equazione:

$$\begin{aligned} \max \mathbf{c}^T \mathbf{x} \\ -x_1 + x_2 + x_3 &= 1 \\ x_1 + x_4 &= 3 \\ x_2 + x_5 &= 2 \\ x_1, x_2, x_3, x_4, x_5 &\geq 0 \end{aligned}$$

in forma matriciale:

$$A = \begin{bmatrix} -1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix} \quad \mathbf{c} = [1 \quad 1 \quad 0 \quad 0 \quad 0]^T$$

Passo 1 : Si parte sempre da una possibile base, sia questa $\mathcal{B}_1 = \{3, 4, 5\}$, si riscrive il problema risolvendo il sistema di equazioni definito da A per le variabili di base

$$\begin{aligned} x_3 &= 1 + x_1 - x_2 \\ x_4 &= 3 - x_1 \\ x_5 &= 2 - x_2 \end{aligned}$$

Il valore della funzione obiettivo è $\mathbf{c}^T \mathbf{x} = z = x_1 + x_2$, si scrive tale equazione insieme alle equazioni del sistema, costruendo una tabella (la cui definizione formale verrà fornita in seguito):

$x_3 = 1 + x_1 - x_2$
$x_4 = 3 - x_1$
$x_5 = 2 - x_2$
$z = x_1 + x_2$

Essendo x_1, x_2 due variabili non di base, queste sono poste uguali a zero, il valore della funzione obiettivo è quindi 0, e la BFS associata a tale base si ottiene risolvendo il sistema di equazioni

$$\begin{cases} x_3 = 1 + 0 - 0 \\ x_4 = 3 - 0 \\ x_5 = 2 - 0 \end{cases} \implies \begin{cases} x_3 = 1 \\ x_4 = 3 \\ x_5 = 2 \end{cases}$$

$$\mathcal{B}_1 = \{3, 4, 5\} \implies \text{BFS} = [0 \quad 0 \quad 1 \quad 3 \quad 2]^T \implies z = 0$$

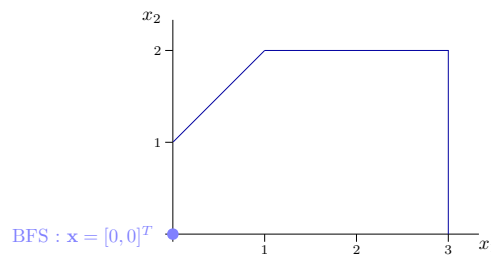


Figura 2.11: BFS con \mathcal{B}_1

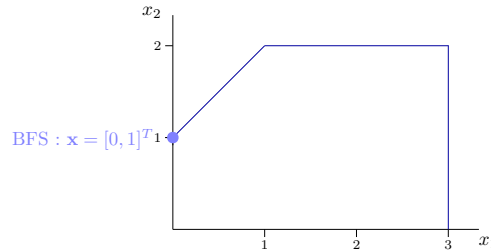
Passo 2 : Si considera adesso una nuova base, partendo da quella iniziale, si tira fuori una variabile per inserirne un'altra, in questo caso lo scambio avviene fra x_2 e x_3 , considerando $\mathcal{B}_2 = \{2, 4, 5\}$, la tabella diviene

$x_2 = 1 + x_1 - x_3$
$x_4 = 3 - x_1$
$x_5 = 2 - x_2$
$z = x_1 + 1 + x_1 - x_3$

Essendo $x_1 = x_3 = 0$, la funzione obiettivo assume valore $z = 1$, la BFS associata si ottiene risolvendo il sistema di equazioni

$$\begin{cases} x_2 = 1 + 0 - 0 \\ x_4 = 3 - 0 \\ x_5 = 2 - x_2 \end{cases} \implies \begin{cases} x_2 = 1 \\ x_4 = 3 \\ x_5 = 2 - 1 \end{cases} \implies \begin{cases} x_2 = 1 \\ x_4 = 3 \\ x_5 = 1 \end{cases}$$

$$\mathcal{B}_2 = \{2, 4, 5\} \implies \text{BFS} = [0 \ 1 \ 0 \ 3 \ 1]^T \implies z = 1$$

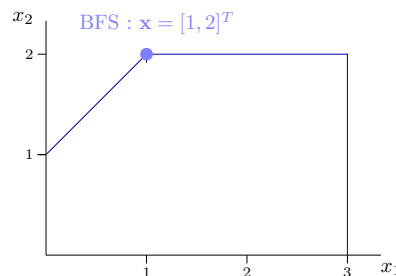
Figura 2.12: BFS con \mathcal{B}_2

Passo 3 : Si vuole incrementare il valore della funzione obiettivo, si scambia la variabile x_1 con la variabile x_5 , ottenendo la base $\mathcal{B}_3 = \{1, 2, 4\}$, la tabella considerata è

$x_1 = 1 + x_3 - x_5$
$x_2 = 2 - x_5$
$x_4 = 2 - x_3 + x_5$
$z = 3 + x_3 - 2x_5$

Risolvendo il sistema di equazioni e sostituendo x_3, x_5 con 0, si ha

$$\mathcal{B}_3 = \{1, 2, 4\} \implies \text{BFS} = [1 \ 2 \ 0 \ 2 \ 0]^T \implies z = 3$$

Figura 2.13: BFS con \mathcal{B}_3

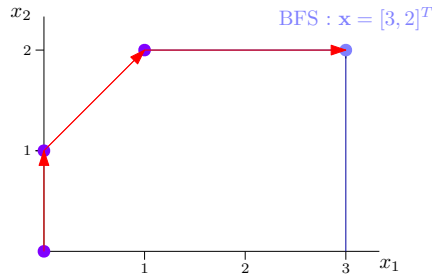
Passo 4 : Il seguente passo è l'ultimo, si sostituisce x_4 con x_3 , considerando la base $\mathcal{B}_4 = \{1, 2, 3\}$, la tabella risulta essere

$x_1 = 3 - x_4$
$x_2 = 2 - x_5$
$x_3 = x_5 - x_4 + 2$
$z = 5 - x_4 - x_5$

Questa risulta essere la soluzione ottimale, z non può essere incrementato in nessun modo dato che è uguale a $5 - x_4 - x_5$, e x_4, x_5 variano in \mathbb{R}^+ , quindi $z \leq 5$, la funzione obiettivo è massimizzata e la soluzione ottimale si ottiene risolvendo il sistema di equazioni

$$\mathcal{B}_4 = \{1, 2, 3\} \implies \text{BFS} = [3 \ 2 \ 2 \ 0 \ 0]^T \implies z = 5$$

Si osservi come geometricamente, il "cammino" fra le diverse BFS date dalle basi considerate equivale ad un cammino sui vertici del poliedro. Durante l'esecuzione dell'algoritmo, ad ogni "passo", si deve rimuovere una variabile dalla base ed inserirne un'altra, si definisce *regola del pivot* la regola con la quale si selezionano le variabili da scambiare (verranno formalizzate in seguito).

Figura 2.14: BFS con \mathcal{B}_4 (soluzione ottimale)

2.4 Il Tableau del Simplexso

Definizione 22 Dato un LP in forma d'equazione

$$\begin{aligned} \max \quad & \mathbf{c}^T \mathbf{x} \\ \text{s.t.} \quad & A\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{aligned}$$

Dove $A \in \text{Mat}(m \times n)$, ed una base ammissibile \mathcal{B} , si definisce **tableau del simplexso** un sistema di $m+1$ equazioni nelle variabili \mathbf{x} che ha lo stesso insieme di soluzioni di

$$\begin{aligned} A\mathbf{x} &= \mathbf{b} \\ z &= \mathbf{c}^T \mathbf{x} \end{aligned}$$

In forma matriciale si rappresenta come segue

$\mathbf{x}_{\mathcal{B}} = \mathbf{p} + Q\mathbf{x}_N$
$z = z_0 + \mathbf{r}^T \mathbf{x}_N$

In questo contesto $N = \{1, 2, \dots, n\} \setminus \mathcal{B}$, ed \mathbf{x}_N è il vettore \mathbf{x} comprendente solamente le componenti di indice $i \in N$. \mathbf{p} è un vettore in \mathbb{R}^m , \mathbf{r} è un vettore in \mathbb{R}^{n-m} , z_0 è un numero reale e Q è una matrice $m \times (n-m)$. Data la base \mathcal{B} , si denota $\mathcal{T}(\mathcal{B})$ il tableau ad essa relativo. Si può trovare la BFS associata a \mathcal{B} ponendo

$$\begin{aligned} x_i &= 0 \quad \forall i \in N \\ x_i &= p_i \quad \forall i \in \mathcal{B} \end{aligned}$$

Ed il valore della funzione obiettivo con tale BFS è z_0 .

Osservazione 11 Se $\mathbf{r} \leq \mathbf{0}$ la soluzione è ottimale ed il valore massimizzato della funzione obiettivo è z_0 .

Lemma 2 Per ogni base ammissibile \mathcal{B} , le componenti del tableau sono univocamente definite ed assumono i seguenti valori

$$Q = -A_{\mathcal{B}}^{-1} \cdot A_N \tag{2.39}$$

$$\mathbf{p} = A_{\mathcal{B}}^{-1} \mathbf{b} \tag{2.40}$$

$$z_0 = \mathbf{c}_{\mathcal{B}}^T A_{\mathcal{B}}^{-1} \mathbf{b} \tag{2.41}$$

$$\mathbf{r} = \mathbf{c}_N - (\mathbf{c}_{\mathcal{B}}^T A_{\mathcal{B}}^{-1} A_N)^T \tag{2.42}$$

Dimostrazione : Essendo che

$$\mathbf{x}_{\mathcal{B}} = [x_i : i \in \mathcal{B}]^T$$

$$\mathbf{x}_N = [x_i : i \notin \mathcal{B}]^T$$

La seguente identità è verificata

$$A\mathbf{x} = A_{\mathcal{B}}\mathbf{x}_{\mathcal{B}} + A_N\mathbf{x}_N = \mathbf{b}$$

quindi

$$A_{\mathcal{B}}\mathbf{x}_{\mathcal{B}} = \mathbf{b} - A_N\mathbf{x}_N$$



Essendo che A_B è per definizione non singolare (essendo B una base ammissibile), esiste l'inverso

$$\mathbf{x}_B = A_B^{-1} \mathbf{b} - A_B^{-1} \cdot A_N \mathbf{x}_N \quad (2.43)$$

Da qui

- $A_B^{-1} \mathbf{b} = \mathbf{p}$
- $-A_B^{-1} \cdot A_N = Q$

inoltre

$$z = \mathbf{c}^T \mathbf{x} = \mathbf{c}_B^T \mathbf{x}_B + \mathbf{c}_N^T \mathbf{x}_N$$

sostituendo \mathbf{x}_B come nell'equazione 2.43 si ha

$$z = \mathbf{c}_B^T (A_B^{-1} \mathbf{b} - A_B^{-1} \cdot A_N \mathbf{x}_N) + \mathbf{c}_N^T \mathbf{x}_N$$

con alcune semplici manipolazioni algebriche:

$$z = \mathbf{c}_B^T A_B^{-1} \mathbf{b} + (\mathbf{c}_N - \mathbf{c}_B^T A_B^{-1} A_N) \mathbf{x}_N$$

Da qui

- $z_0 = \mathbf{c}_B^T A_B^{-1} \mathbf{b}$
- $\mathbf{r}^T = \mathbf{c}_N - \mathbf{c}_B^T A_B^{-1} A_N$

■

Si osservi come il metodo del simplesso richiede una base ammissibile dalla quale partire, dato un LP in forma standard, si può trasformare in forma d'equazione aggiungendo le variabili slack, la matrice originale verrà affiancata da una matrice identità di m colonne (m numero originale di vincoli), queste sono banalmente linearmente indipendenti, quindi quella composta dalle variabili slack è sempre una valida base dalla quale partire.

Come procedere nel caso generale di un problema che si trova già in forma di equazione?

$$\begin{aligned} A\mathbf{x} &= \mathbf{b} \\ \mathbf{x} &\geq \mathbf{0} \end{aligned}$$

Nota : Si può assumere che \mathbf{b} abbia ogni componente positiva, se così non dovesse essere, il vincolo

$$a_{1,j}x_1 + \dots + a_{n,j}x_n = b_j$$

si potrebbe trasformare in

$$-a_{1,j}x_1 - \dots - a_{n,j}x_n = -b_j$$

Al problema originale si possono aggiungere delle variabili slack per ogni vincolo

$$A' \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = [A | \text{Id}_m] \begin{bmatrix} x_1 \\ \vdots \\ x_n \\ y_1 \\ \vdots \\ y_m \end{bmatrix} = \mathbf{b} \quad \begin{aligned} \mathbf{x} &\geq 0 \\ \mathbf{y} &\geq 0 \end{aligned}$$

In tal modo si ha una BFS semplice associata ad una base contenente le variabili slack.

Osservazione 12 *Se esiste una soluzione per $A' \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \mathbf{b}$ con le variabili slack uguali a zero, allora esiste una soluzione per il problema originale.*

Si può fare risolvendo il seguente LP con una nuova funzione obiettivo

$$\begin{aligned} \max & -y_1 - y_2 \cdots - y_m \\ A' \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} &= \mathbf{b} \\ \mathbf{x}, \mathbf{y} &\geq 0 \end{aligned} \quad (2.44)$$

chiaramente dato $\mathbf{y} \geq 0$ il valore minimo che può assumere \mathbf{y} è zero. Quindi se il nuovo problema 2.44 ha soluzione ottimale diversa da zero, allora non esiste una soluzione per il problema originale.

Il ragionamento appena espresso porta al seguente risultato.

Proposizione 11 *Il sistema di equazioni*

$$\begin{aligned} A\mathbf{x} &= \mathbf{b} \\ \mathbf{x} &\geq 0 \end{aligned}$$

ha soluzione se e solo se il seguente programma lineare

$$\begin{aligned} \max & -y_1 - y_2 \cdots - y_m \\ A' \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} &= \mathbf{b} \\ \mathbf{x}, \mathbf{y} &\geq 0 \end{aligned}$$

ha come valore ottimale 0.

Nelle sezioni precedenti abbiamo visto come la ricerca di soluzioni ammissibili riguarda i punti contenuti in un poliedro (definito dai vincoli lineari), un problema correlato ma sostanzialmente diverso, riguarda lo stabilire se un dato poliedro sia vuoto o no. Si potrebbe ipotizzare, che tale problema sia simile alla risoluzione di un programma lineare, ma più semplice, in quanto un LP riguarda sia la ricerca di una soluzione ammissibile (punti nel poliedro), sia la ricerca del punto specifico che massimizza la funzione obiettivo. La seguente proposizione enuncia che tale ipotesi è falsa.

Proposizione 12 *Sia ϕ una funzione oracolo, che dato un poliedro definito da un sistema di disequazioni lineari, assume valore 1 se e solo se il poliedro non è vuoto (il sistema ha soluzione), altrimenti, assume valore 0. Se ϕ esiste, è possibile risolvere un qualsiasi programma lineare tramite un numero finito di valutazioni di ϕ .*

Piuttosto che una dimostrazione rigorosa sarà data una prova informale, sia

$$\begin{aligned} \max & \mathbf{c}^T \mathbf{x} \\ A\mathbf{x} &\leq \mathbf{b} \\ \mathbf{x} &\geq 0 \end{aligned}$$

il programma lineare in questione, si assume che gli elementi di A e di \mathbf{c} siano numeri razionali e non reali. Si costruisce, dato un arbitrario $\alpha \in \mathbb{Q}$ il seguente poliedro

$$P_\alpha = \begin{cases} \mathbf{c}^T \mathbf{x} \leq \alpha \\ A\mathbf{x} \leq \mathbf{b} \\ \mathbf{x} \geq 0 \end{cases}$$

Si utilizza l'oracolo ϕ , se P_α è non vuoto, ossia se $\phi(P_\alpha) = 1$, allora il valore massimizzato del programma lineare è maggiore o uguale ad α , altrimenti, se $\phi(P_\alpha) = 0$, il valore è minore. Si può fare ricerca binaria sui valori di α , data l'assunzione iniziale sul fatto che i numeri coinvolti siano razionali, la ricerca termina in un numero finito di passi.

2.4.1 Trovare una Base Ammissibile

Si consideri il seguente programma lineare

$$\begin{aligned} \max & -12x_1 - 3x_2 - 4x_3 \\ -4x_1 - 2x_2 - 3x_3 &\leq -2 \\ -8x_1 - x_2 - 2x_3 &\leq -3 \\ x_1, x_2, x_3 &\geq 0 \end{aligned}$$

Si vuole risolvere con il metodo del simplesso, si trasforma quindi in forma di equazione:

$$\begin{aligned}
 \max \quad & -12x_1 - 3x_2 - 4x_3 \\
 -4x_1 - 2x_2 - 3x_3 + x_4 = & -2 \\
 -8x_1 - x_2 - 2x_3 + x_5 = & -3 \\
 x_1, x_2, x_3, x_4, x_5 \geq & 0
 \end{aligned} \tag{2.45}$$

Si procede tentando con la base iniziale (banale) $\mathcal{B} = \{4, 5\}$, si noti però che tale base è inammissibile

$$\begin{aligned}
 x_4 &= -2 \\
 x_5 &= -3 \\
 x_4, x_5 &\geq 0
 \end{aligned}$$

È necessario trovare una base ammissibile, questo è un compito computazionalmente complesso tanto quanto risolvere un programma lineare, de facto, nei casi in cui

- La base banale iniziale di un LP trasformato con le slack variable non è ammissibile
- Il problema da risolvere è dato direttamente in forma di equazione

è necessario trovare una base ammissibile dalla quale iniziare il metodo del simplesso. Il seguente teorema fornisce un metodo a tale scopo.

Teorema 4 *Dato un LP in forma di equazione*

$$\begin{aligned}
 \max \quad & c_1x_1 + \cdots + c_nx_n \\
 a_{1,1}x_1 + \cdots + a_{1,n}x_n &= b_1 \\
 & \vdots \\
 a_{m,1}x_1 + \cdots + a_{m,n}x_n &= b_m \\
 \mathbf{x} &\geq \mathbf{0}
 \end{aligned}$$

Si può considerare il seguente **problema ausiliario**

$$\begin{aligned}
 \max \quad & -x_{n+1} - \cdots - x_{n+m} \\
 a_{1,1}x_1 + \cdots + a_{1,n}x_n \bullet x_{n+1} &= b_1 \\
 & \vdots \\
 a_{m,1}x_1 + \cdots + a_{m,n}x_n \bullet x_{n+m} &= b_m \\
 \mathbf{x} &\geq \mathbf{0}
 \end{aligned}$$

Dove il simbolo \bullet nell' i -esimo vincolo è uguale a $+$ se $b_i \geq 0$, ed è uguale a $-$ se $b_i < 0$. Si ha la base $\mathcal{B} = \{n+1, n+2, \dots, n+n\}$ è ammissibile per il problema ausiliario, e qualsiasi BFS ottimale per tale problema, è associata ad una base la cui restrizione sulle variabili $1 \dots, n$ è ammissibile per il problema originale.

Si consideri nuovamente il problema 2.45, applicando il teorema 4 si ottiene il programma ausiliario

$$\begin{aligned}
 \max \quad & -x_6 - x_7 \\
 -4x_1 - 2x_2 - 3x_3 + x_4 - x_6 &= -2 \\
 -8x_1 - x_2 - 2x_3 + x_5 - x_7 &= -3 \\
 x_1, x_2, x_3, x_4, x_5, x_6, x_7 &\geq 0
 \end{aligned} \tag{2.46}$$

la base $\mathcal{B} = \{6, 7\}$ è ammissibile, e la soluzione ottimale è la BFS associata alla base $\mathcal{B} = \{1, 2\}$, quest'ultima è ammissibile per il problema originale, e si può cominciare il metodo del simplesso partendo da questa.

2.5 Soluzioni Degenere

Alcune BFS sono problematiche quando valutate durante il metodo del simplesso e possono portare ad una situazione di stallo in cui si considerano in maniera ciclica le stesse basi a ripetizione.

Definizione 23 Dato un LP, sia $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ una BFS, se l'insieme

$$\{x_i \text{ t.c. } x_i > 0\}$$

ha cardinalità minore di $n - m$ (alcune variabili di base sono uguali a zero), allora \mathbf{x} è una soluzione **degenere**.

Si consideri il seguente esempio

$$\begin{aligned} \max x_2 \\ x_1 &\leq 1 \\ -x_1 + x_2 &\leq 0 \\ x_1, x_2 &\geq 0 \end{aligned} \quad (2.47)$$

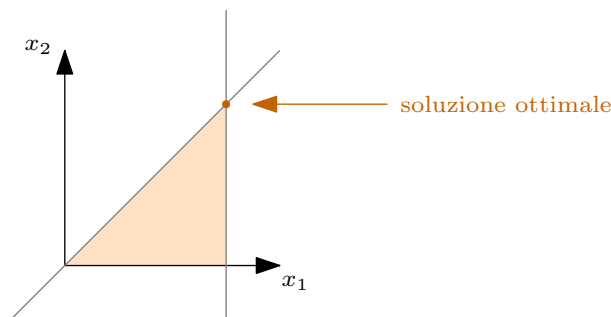


Figura 2.15: Poliedro delle soluzioni ammissibili

La soluzione ottimale è (banalmente) $\bar{\mathbf{x}} = [1, 1]^T$, il problema in forma di equazione assume la seguente forma

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ -1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (2.48)$$

La base di partenza è quella composta dalle variabili slack, ossia $\mathcal{B} = \{3, 4\}$, ponendo $x_1, x_2 = 0$ si risolve per x_3, x_4 trovando la BFS

$$[0 \ 0 \ 1 \ 0]^T$$

Questa è per definizione degenere, il tableau per \mathcal{B} è il seguente

$x_3 = 1 - x_1$
$x_4 = 0 + x_1 - x_2$
$z = 0 + x_2$

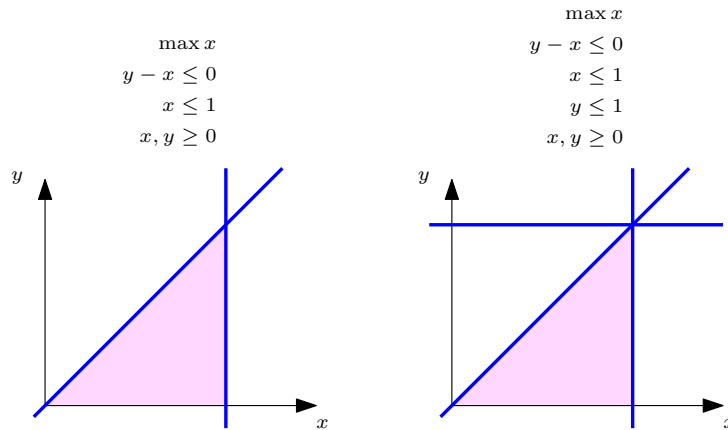
si osservi come c'è un'unica possibilità di sostituzione per il pivot, si sostituisce x_2 con x_4 , la nuova base è $\mathcal{B} = \{2, 3\}$

$x_2 = 0 + x_1 - x_4$
$x_3 = 1 - x_1$
$z = 0 + x_1 - x_4$

La BFS in questo caso è identica, si noti come lo scambio delle variabili non ha fatto incrementare il valore di z , questa è la conseguenza di una soluzione degenere. In questo caso, scambiando x_1 con x_3 si ottiene la soluzione ottimale. Si noti che una BFS degenere può essere associata a più di una base (ciò non si verifica in ogni caso).

Osservazione 13 Una soluzione è degenere se non è univocamente definita, ossia, esistono diversi insiemi di vincoli che identificano la stessa soluzione.

Le variabili degenere sono causa di una descrizione *ridondante* di un poliedro, che coinvolge iperpiani non necessari (se non considerati, il sistema di equazioni descrive lo stesso poliedro).

Figura 2.16: Il vincolo $y \leq 1$ è ridondante per il poliedro

2.6 Problemi del Metodo del Simplexso

2.6.1 Regola del Pivot

Nella forma generale un tableau del simplexso è definito come segue

$$\begin{array}{|l} \mathbf{x}_B = \mathbf{p} + Q\mathbf{x}_N \\ z = z_0 + \mathbf{r}^T \mathbf{x}_N \end{array}$$

Si è visto come se $\mathbf{r} \leq \mathbf{0}$ allora z_0 è la soluzione ottimale. Quando si deve scegliere quale variabile far entrare nella base durante un passo di computazione nel metodo del simplexso, è possibile scegliere una qualsiasi variabile x_i tale per cui $r_i > 0$. Per comodità, si assumano le seguenti notazioni per gli indici della base e del complemento della base:

$$\begin{aligned} \mathcal{B} &= \{k_1, k_2, \dots, k_m\} \\ \mathcal{N} &= \{l_1, l_2, \dots, l_{n-m}\} \end{aligned}$$

Essendo $\mathbf{x}_B = \mathbf{p} + Q\mathbf{x}_N$, si ha che

$$x_{k_i} = p_i + \sum_{j=1}^{n-m} Q_{i,j} x_{l_j}$$

Sia l_β l'indice che entra nella base in uno step del simplexso

$$l_\beta \in \{l_1, l_2, \dots, l_{n-m}\} \text{ tale che } r_\beta > 0$$

Bisogna selezionare secondo un dato criterio, la variabile nella base che deve essere rimossa per fare posto a x_{l_β} . Il criterio è il seguente, la variabile che esce è x_{k_α} , α è scelto in modo da minimizzare

$$-\frac{p_\alpha}{Q_{\alpha,\beta}}$$

Ricapitolando, lo scambio

$$\begin{array}{ll} x_{l_\beta} & \text{entrante} \\ x_{k_\alpha} & \text{uscente} \end{array}$$

è valido se

- $r_\beta > 0$
- α minimizza $-\frac{p_\alpha}{Q_{\alpha,\beta}}$ fra tutti gli i tali che $Q_{i,\beta} < 0$

Una **regola del pivot** descrive il criterio di scelta delle variabili da scambiare fra tutti i possibili scambi validi, alcune regole note sono le seguenti

- **Coefficiente più grande** : Si sceglie x_{l_β} entrante che massimizza r_β , la scelta di x_{k_α} uscente è arbitraria fra quelle valide.
- **Incremento maggiore** : Per tutti i possibili l_j per cui $r_j > 0$ si calcola

$$\min_{i \text{ t.c. } Q_{i,j} < 0} -\frac{p_i}{Q_{i,j}}$$

e si sceglie β che massimizza

$$-\frac{p_i}{Q_{i,j}} r_j$$

Entrambe queste regole sono computazionalmente valide, seppure esistono dei casi limite in cui il simplesso va in loop a causa di variabili degeneri. La seguente regola è computazionalmente inefficiente, ma garantisce la terminazione del metodo del simplesso.

- **Regola di Bland** : Si sceglie x_{l_β} tale che $r_\beta > 0$ e l_β è l'indice minimale.

2.6.2 Cicli e Variabili Fickle

Si consideri un LP in forma di equazione, con tableau

$\mathbf{x}_B = \mathbf{p} + Q\mathbf{x}_N$
$z = z_0 + \mathbf{r}^T \mathbf{x}_N$

e $\mathbf{p}, \mathbf{r}, Q$ univocamente definiti da una base ammissibile \mathcal{B} , denotando gli indici

$$\begin{aligned}\mathcal{B} &= \{k_1, k_2, \dots, k_m\} \\ N &= \{l_1, l_2, \dots, l_{n-m}\}\end{aligned}$$

In un passo di computazione, secondo la regola di Bland, se x_{l_β} è la variabile che entra nella base, la nuova base sarà

$$(\mathcal{B} \setminus \{k_j\}) \cup \{l_\beta\}$$

per cui j è minimale. Un **ciclo** nel simplesso si manifesta sotto-forma di una sequenza di basi $\mathcal{B}_1, \dots, \mathcal{B}_t$ che si ripetono nei passi di computazione, il tableau della base \mathcal{B}_i porta alla base \mathcal{B}_{i+1} , fino alla base \mathcal{B}_t il cui tableau porta alla base \mathcal{B}_1 .

$$\mathcal{T}(\mathcal{B}_1) \longrightarrow \mathcal{T}(\mathcal{B}_2) \longrightarrow \mathcal{T}(\mathcal{B}_3) \longrightarrow \dots \longrightarrow \mathcal{T}(\mathcal{B}_t) \longrightarrow \mathcal{T}(\mathcal{B}_1)$$

Definizione 24 Nel contesto dell'esecuzione del metodo del simplesso, una variabile x_u è detta **fickle** se, per qualche i, k, j con $j > k > i$, si ha che

$$\begin{aligned}u &\in \mathcal{B}_i \\ u &\notin \mathcal{B}_k \\ u &\in \mathcal{B}_j\end{aligned}$$

Quindi viene prima rimossa dalla base, per poi venire re-inserita in un passo di computazione successivo.

Il seguente lemma è fondamentale per la dimostrazione (cruciale) della correttezza del metodo del simplesso quando la regola del pivot applicata è la regola di Bland.

Lemma 3 Siano $\mathcal{T}(\mathcal{B}_1) \rightarrow \mathcal{T}(\mathcal{B}_2) \rightarrow \dots \mathcal{T}(\mathcal{B}_t) \rightarrow \mathcal{T}(\mathcal{B}_1)$ le basi che susseguendosi formano un ciclo durante l'esecuzione del metodo del simplesso con una determinata regola del pivot. Sia F l'insieme degli indici delle variabili fickle per tale esecuzione. Esiste un'unica soluzione ammissibile $\bar{\mathbf{x}}$ comune per tutte le basi $\mathcal{B}_1, \dots, \mathcal{B}_t$, per cui

$$x_i = 0, \quad \forall i \in F$$

Dimostrazione : Il valore della funzione obiettivo durante il simplesso è crescente, essendo che ci si trova in un ciclo, questo rimane identico, denotiamo il tableau associato alla base \mathcal{B}_j come segue

$\mathbf{x}_{\mathcal{B}_j} = \mathbf{p}^j + Q^j \mathbf{x}_{N_j}$
$z = z_0 + \mathbf{r}^j{}^T \mathbf{x}_{N_j}$

e denotiamo \mathbf{x}^j la BFS associata a \mathcal{B}_j . z_0 è invariante per ogni j . Siano β e α gli indici delle variabili entrante ed uscente

$$\mathcal{B}_{j+1} = (\mathcal{B}_j \setminus \{\alpha\}) \cup \{\beta\}$$

Il lemma afferma che la BFS è la stessa per tutte le basi e che le variabili fickle sono uguali a zero. Si vuole dimostrare la prima affermazione del lemma, ossia che

$$\mathbf{x}^j = \mathbf{x}^{j+1}$$

per ogni j . Sappiamo che le variabili non di base in due passi di computazione differenti, ossia in N_j e N_{j+1} , sono uguali a zero (esclusa β che viene inserita in N_{j+1})

$$x_i^j = x_i^{j+1} = 0, \quad i \in N_j \setminus \{\beta\}$$

Inoltre, sappiamo che il valore della funzione obiettivo non aumenta, essendo

$$x_\alpha^j = p_\alpha^j + \begin{bmatrix} Q_{\alpha,1}^j & Q_{\alpha,2}^j & \cdots & Q_{\alpha,n-m}^j \end{bmatrix}^T \cdot \mathbf{x}_{N_j}$$

necessariamente si ha che $p_\alpha^j = 0$, quindi

$$x_\alpha^j = \begin{bmatrix} Q_{\alpha,1}^j & Q_{\alpha,2}^j & \cdots & Q_{\alpha,n-m}^j \end{bmatrix}^T \cdot \mathbf{x}_{N_j}$$

TODO : Finire la dimostrazione

Teorema 5 *Il metodo del simplesso con la regola di Bland termina sempre, e non entra mai in un ciclo.*

Dimostrazione : Si procede per assurdo, si assuma che durante l'esecuzione del metodo del simplesso per un generico LP con regola di Bland ci sia un ciclo:

$$\mathcal{T}(B_1) \rightarrow \mathcal{T}(B_2) \rightarrow \dots \mathcal{T}(B_t) \rightarrow \mathcal{T}(B_1)$$

Sia F l'insieme degli indici delle variabili *fickle*, sia $v = \max(F)$. Si considerino due particolari basi :

- Denotiamo \mathcal{B} una base \mathcal{B}_i per cui $v \notin \mathcal{B}_i$ e $v \in \mathcal{B}_{i+1}$
- Denotiamo \mathcal{B}' una base \mathcal{B}_j per cui $v \in \mathcal{B}_j$ e $v \notin \mathcal{B}_{j+1}$

\mathcal{B} è la base precedente a quella in cui entra v , e \mathcal{B}' è l'ultima base in cui si trova v prima di essere rimosso. Denotiamo i due relativi tableau

$\mathbf{x}_{\mathcal{B}} = \mathbf{p} + Q\mathbf{x}_N$	$\mathbf{x}_{\mathcal{B}'} = \mathbf{p}' + Q'\mathbf{x}_{N'}$
$z = z_0 + \mathbf{r}^T \mathbf{x}_N$	$z = z_0 + \mathbf{r}'^T \mathbf{x}_{N'}$

Esplicitando gli elementi delle basi si ha

$$\mathcal{B} = \{k_1, \dots, k_m\} \quad N = \{l_1, \dots, l_{n-m}\}$$

$$\mathcal{B}' = \{k'_1, \dots, k'_m\} \quad N' = \{l'_1, \dots, l'_{n-m}\}$$

$v \in N \implies v = l_\beta$ per qualche β . Sia k_α l'indice che viene rimosso da \mathcal{B} per far entrare v , similmente, $v = k'_{\alpha'}$ per qualche α' che esce da \mathcal{B}' per far entrare $l'_{\beta'}$ per qualche β' . Dato che si sta usando la regola di Bland, l_β è il più piccolo indice (per una valida scelta) di pivot che viene rimosso da N .

$$r_\beta > 0, \quad r_i \leq 0 \quad \forall i \text{ t.c. } l_i \in F \cap N$$

$\implies v$ lascia $\mathcal{B}' \implies v = k'_{\alpha'}$, ed è l'unico indice di $F \cap \mathcal{B}'$ che soddisfa le condizioni per essere rimosso da \mathcal{B}' .

$$Q'[i, \beta'] < 0 \quad \text{e} \quad -\frac{p'_i}{Q'[i, \beta']} \text{ è minimale}$$

ma p'_i è zero, p'_i è il valore di \bar{x}_v per la BFS \bar{x} associata alla base \mathcal{B}' , x_v è una variabile fickle, è quindi uguale a zero per il lemma 3. Le altre variabili fickle in \mathcal{B}' corrispondono a valori di \mathbf{p}' uguali a zero,

quindi v è l'unica variabile fickle in \mathcal{B}' tale per cui $Q[\alpha', \beta'] < 0$. Si costruisce un'ulteriore programma lineare basato su quello originale, come segue:

$$\begin{aligned}
 & \max \mathbf{c}^T \mathbf{x} \text{ stessa funzione obiettivo} \\
 & A\mathbf{x} = \mathbf{b} \\
 & \mathbf{x}_{F \setminus \{v\}} \geq 0 \\
 & x_v \leq 0 \\
 & \mathbf{x}_{N \setminus F} \leq 0 \\
 & \mathbf{x}_{\mathcal{B} \setminus F} \text{ non è vincolato}
 \end{aligned} \tag{2.49}$$

Se i due seguenti claim sono veri, le assunzioni della dimostrazione portano ad una contraddizione.

Claim 1 : $\bar{\mathbf{x}}$ è una soluzione ottimale per il problema 2.49.

Dimostrazione Claim 1 : Sappiamo che $\bar{\mathbf{x}}$ è ammissibile per il problema originale, quindi $A\bar{\mathbf{x}} = \mathbf{b}$ è soddisfatta, inoltre $\bar{x}_v \leq 0$ dato che tutte le variabili fickle sono uguali a zero. Chiaramente $\bar{\mathbf{x}}_N$ e $\bar{\mathbf{x}}_F$ sono nulli, quindi $\bar{\mathbf{x}}$ è ammissibile per il problema 2.49. Sia \mathbf{y} una soluzione ammissibile per 2.49, il valore della funzione obiettivo con \mathbf{y} è

$$\mathbf{c}^T \mathbf{y} = z_0 + \mathbf{r}^T \mathbf{y}_N$$

essendo \mathbf{y} ammissibile si ha $\mathbf{y}_{N \setminus F} = \mathbf{0}$.

- $x_{l_j} \geq 0$ se $l_j \in F \setminus \{v\}$
- $r_i \leq 0$, $\forall i$ t.c. $l_i \in F \setminus \{v\}$
- $r_\beta > 0$

Quindi ogni termine di $\mathbf{r}^T \mathbf{y}_N$ ha ogni componente minore o uguale di zero, quindi $\mathbf{c}^T \mathbf{y} \leq z_0 \implies \mathbf{c}^T \bar{\mathbf{x}} \geq \mathbf{c}^T \mathbf{y} \implies \bar{\mathbf{x}}$ è ottimale per il problema 2.49. \square

Claim 2 : Il programma lineare 2.49 non è limitato.

Dimostrazione Claim 2 : Sia $\bar{x}(t)$ definita come segue: tutte le sue componenti sono nulle eccetto la variabile u -esima, dove u è l'indice entrante in \mathcal{B}' **TODO : finire dimostrazione**

CAPITOLO

3

DUALITÀ

3.1 Dualità Forte e Debole

Definizione 25 Sia

$$\begin{aligned} \max \mathbf{c}^T \mathbf{x} \\ A\mathbf{x} \leq \mathbf{b} \\ \mathbf{x} \geq 0 \end{aligned}$$

Un programma lineare che in tal contesto è detto problema **primario**. Il suo problema **duale** è il seguente programma lineare

$$\begin{aligned} \min \mathbf{b}^T \mathbf{y} \\ A^T \mathbf{y} \geq \mathbf{c} \\ \mathbf{y} \geq 0 \end{aligned}$$

Teorema 6 (dualità debole) L'ottimo del problema duale è un limite superiore per la funzione obiettivo del problema primario.

$$\text{ottimo primario} \leq \text{ottimo duale}$$

Se esiste una soluzione per il duale, il primario è limitato, se il problema duale è illimitato, si avrebbe

$$\text{ottimo primario} \leq -\infty$$

ciò è impossibile, il problema primario in tal caso non può essere ammissibile e limitato.

Teorema 7 (dualità forte) data una coppia di programmi lineari primario-duale, esattamente una delle seguenti condizioni si verifica

- entrambi i problemi sono inammissibili
- il problema primario è ammissibile ma illimitato, il problema duale è inammissibile
- il problema duale è ammissibile ma illimitato, il problema primario è inammissibile
- entrambi i problemi sono ammissibili e limitati, e condividono lo stesso ottimo

La dimostrazione verrà riportata in seguito. La seguente tabella riassume le possibilità dei casi riguardanti le soluzioni fra problema primario e duale.

		primario		
		ammissibile e limitato	ammissibile e illimitato	non ammissibile
duale	ammissibile e limitato	○	×	×
	ammissibile e illimitato	×	×	○
	non ammissibile	×	○	○

Figura 3.1: In verde sono riportati i possibili casi

In generale, l' i -esimo vincolo del programma primario non deve essere necessariamente nella forma " \leq ", in base a questo, l' i -esima variabile del programma duale deve essere confinata in un certo intervallo della retta come segue

$$\begin{cases} A_i^T \mathbf{x} \leq b_i \implies y_i \geq 0 \\ A_i^T \mathbf{x} \geq b_i \implies y_i \leq 0 \\ A_i^T \mathbf{x} = b_i \implies y_i \in \mathbb{R} \end{cases}$$

equivalentemente, la forma dell' i -esimo vincolo del programma duale dipende dalla regione alla quale appartiene l' i -esima variabile del programma primario

$$\begin{cases} x_i \geq 0 \implies A_i \mathbf{y} \geq c_i \\ x_i \leq 0 \implies A_i \mathbf{y} \leq c_i \\ x_i \in \mathbb{R} \implies A_i \mathbf{y} = c_i \end{cases}$$

Proposizione 13 Se P è un programma lineare e D è il suo duale, allora il duale di D è P .

Claim: il teorema 8 è sufficiente a dimostrare il teorema 7.

Teorema 8 Sia P un programma lineare e D il suo duale, se P è ammissibile e limitato, anche D lo è, e hanno lo stesso ottimo.

Dimostrazione: Sia

$$P = \begin{cases} \max \mathbf{c}^T \mathbf{x} \\ A\mathbf{x} \leq \mathbf{b} \\ \mathbf{x} \geq 0 \end{cases}$$

il problema primario, con $\mathbf{x} \in \mathbb{R}^n$, si assuma che sia ammissibile e limitato, quindi che esiste un ottimo per P ed il metodo del simplesso eseguito su P fornirà una BFS che massimizza la funzione obiettivo. Si consideri il seguente problema P^* , equivalente a P ma in forma di equazione

$$P^* = \begin{cases} \max \bar{\mathbf{c}}^T \bar{\mathbf{x}} \\ \bar{A}\bar{\mathbf{x}} \leq \mathbf{b} \\ \bar{\mathbf{x}} \geq 0 \end{cases}$$

con

$$\begin{aligned} \bar{\mathbf{c}} &= [\mathbf{c} \mid 0 \dots 0]^T \in \mathbb{R}^{n+m} \\ \bar{A} &= [A \mid Id_m] \\ \bar{\mathbf{x}} &\in \mathbb{R}^{n+m} \end{aligned}$$

Sia \mathcal{B} la base associata alla soluzione ottimale per P^* , il tableau associato è il seguente

$$\begin{aligned} \bar{\mathbf{x}}_{\mathcal{B}} &= \bar{\mathbf{p}} + \bar{Q}\bar{\mathbf{x}}_N \\ z &= z_0 + \bar{\mathbf{r}}^T \bar{\mathbf{x}}_N \end{aligned}$$

la soluzione ottimale è $\bar{\mathbf{x}}^*$ tale che

$$\begin{aligned} \bar{\mathbf{x}}_N^* &= \mathbf{0} \\ \bar{\mathbf{x}}_{\mathcal{B}}^* &= \bar{\mathbf{p}} \end{aligned}$$

Dato il lemma 2:

$$\bar{\mathbf{p}} = \bar{A}_B^{-1} \mathbf{b}$$

La soluzione ottimale per il problema originale P è uguale alle prime n componenti di $\bar{\mathbf{x}}^*$. Si definisce un termine \mathbf{y}^* come segue

$$\mathbf{y}^* = (\bar{\mathbf{c}}_B^T \bar{A}_B^{-1})^T \quad (3.1)$$

Lo scopo sta nel dimostrare che \mathbf{y}^* è ammissibile per il duale di P , e che $\mathbf{b}^T \mathbf{y}^* = \mathbf{c}^T \mathbf{x}^*$.

Si ha che $\mathbf{c}^T \mathbf{x}^* = \bar{\mathbf{c}}^T \bar{\mathbf{x}}^*$ perché gli ultimi m elementi di \mathbf{c} sono nulli. Inoltre $\bar{\mathbf{c}}^T \bar{\mathbf{x}}^* = \bar{\mathbf{c}}_B^T \bar{\mathbf{p}}$ dato che $\bar{\mathbf{p}}$ equivale alle componenti non nulle di $\bar{\mathbf{x}}^*$. Da qui si ha

$$\bar{\mathbf{c}}_B^T \bar{\mathbf{p}} = \bar{\mathbf{c}}_B^T \bar{A}_B^{-1} \mathbf{b} \quad (3.2)$$

per definizione $\mathbf{y}^* = (\bar{\mathbf{c}}_B^T \bar{A}_B^{-1})^T$

$$\bar{\mathbf{c}}_B^T \bar{A}_B^{-1} \mathbf{b} = \mathbf{y}^{*T} \mathbf{b} = \mathbf{b}^T \mathbf{y}^* \implies \quad (3.3)$$

$$\mathbf{c}^T \mathbf{x}^* = \mathbf{b}^T \mathbf{y}^* \quad (3.4)$$

Bisogna mostrare che \mathbf{y}^* è ammissibile per il duale di P , ossia che

$$\begin{aligned} A^T \mathbf{y}^* &\geq \mathbf{c} \\ \mathbf{y}^* &\geq \mathbf{0} \end{aligned}$$

Si può considerare la versione con le variabili slack e dimostrare direttamente che

$$\begin{bmatrix} A^T \\ - \\ Id_m \end{bmatrix} \mathbf{y}^* \geq \begin{bmatrix} \mathbf{c} \\ - \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Si può riscrivere

$$\bar{A}^T \mathbf{y}^* \geq \bar{\mathbf{c}}$$

Si ha che

$$\bar{A}^T \mathbf{y}^* = \bar{A}^T (\bar{\mathbf{c}}_B^T \bar{A}_B^{-1})^T = (\bar{\mathbf{c}}_B^T \bar{A}_B^{-1} \bar{A})^T$$

si pone per definizione

$$\mathbf{w} = (\bar{\mathbf{c}}_B^T \bar{A}_B^{-1} \bar{A})^T$$

è necessario mostrare che $\mathbf{w} \geq \bar{\mathbf{c}}$. Verrà dimostrato separando l'affermazione in due affermazioni, ossia

- $\mathbf{w}_N \geq \bar{\mathbf{c}}_N$
- $\mathbf{w}_B \geq \bar{\mathbf{c}}_B$

Si ha che

$$\mathbf{w}_N = (\bar{\mathbf{c}}_B^T \bar{A}_B^{-1} \bar{A}_N)^T$$

si osservi la definizione di \mathbf{r} nel lemma 2, ne consegue

$$\mathbf{r} = \bar{\mathbf{c}}_N - (\bar{\mathbf{c}}_B^T \bar{A}_B^{-1} \bar{A}_N)^T \implies$$

quindi

$$\mathbf{x}_N = \bar{\mathbf{c}}_N - \mathbf{r} \geq \bar{\mathbf{c}}_N$$

perché \mathbf{r} è un vettore di componenti minori di zero. quindi è dimostrato che $\mathbf{w}_N \geq \bar{\mathbf{c}}_N$.

Si consideri ora \mathbf{w}_B , questo è uguale a

$$\mathbf{w}_B = (\bar{\mathbf{c}}_B^T \bar{A}_B^{-1} \bar{A}_B)^T$$

ma $\bar{A}_B^{-1} \bar{A}_B^T$ è uguale alla matrice identità, quindi

$$\mathbf{w}_B = \bar{\mathbf{c}}_B^T$$

Entrambi i punti sono dimostrati, quindi \mathbf{y}^* è ammissibile per il problema duale, e tale soluzione trova un'ottimo identico all'ottimo del problema primario P , quindi per il teorema della dualità debole, \mathbf{y}^* è una soluzione ottimale per il duale. ■



3.2 Lemma di Farkas

Abbiamo visto come il teorema della dualità forte stabilisce importanti risultati riguardo le soluzioni di una coppia di problemi duale-primario. La seguente sezione riguarda un'interpretazione geometrica del teorema 8.

Lemma 4 (Farkas) Sia A una matrice $m \times n$ e $\mathbf{b} \in \mathbb{R}^m$, esattamente una fra le seguenti è vera

1. $\exists \mathbf{x} \in \mathbb{R}^n$ tale che $A\mathbf{x} = \mathbf{b}$, $\mathbf{x} \geq 0$
2. $\exists \mathbf{y} \in \mathbb{R}^m$ tale che $A\mathbf{y}^T \geq \mathbf{0}$, $\mathbf{y}^T \mathbf{b} < 0$

Dimostrazione: Si consideri il seguente programma lineare

$$P = \begin{cases} \max \mathbf{0} \\ A\mathbf{x} = \mathbf{b} \\ \mathbf{x} \geq 0 \end{cases}$$

Il problema duale è

$$D = \begin{cases} \min \mathbf{b}^T \mathbf{y} \\ A^T \mathbf{y} \geq \mathbf{0} \\ \mathbf{y} \in \mathbb{R}^m \end{cases}$$

D è ammissibile perché $\mathbf{y} = \mathbf{0}$ è una soluzione, per il teorema della dualità forte, una delle seguenti è verificata

- P è inammissibile se D è illimitato
- P è ammissibile e ha lo stesso ottimo di D

Se P fosse inammissibile, ossia il punto (1) del lemma non fosse verificato, allora $\mathbf{b}^T \mathbf{y}$ tenderebbe $-\infty$, decrescendo indefinitamente, quindi esisterebbe un \mathbf{y} che rende negativa la funzione obiettivo, ma allora il punto (2) del lemma di Farkas è verificato.

Se P fosse ammissibile, ossia il punto (1) del lemma fosse verificato, si avrebbe come ottimo 0, quindi non esisterebbe un \mathbf{y} che rende negativa $\mathbf{b}^T \mathbf{y}$, quindi il punto (2) del lemma di Farkas non sarebbe verificato. ■

Tale lemma fornisce un certificato per verificare l'ammissibilità di un programma lineare, se il punto 2 del lemma di Farkas è verificato per un programma lineare in forma di equazione, allora la regione ammissibile è vuota. Tale lemma può essere dimostrato sia utilizzando la dualità forte, sia non utilizzandola. Verrà data in seguito una dimostrazione che non si avvale del teorema 7.

La seguente proposizione è una variante del lemma di Farkas.

Proposizione 14 se $A \in \text{Mat}(m \times n)$ e $\mathbf{b} \in \mathbb{R}^m$, le seguenti sono equivalenti

- $A\mathbf{x} = \mathbf{b}$ ha una soluzione non negativa se e solo se $\forall \mathbf{y} \in \mathbb{R}^m$ tale che $A^T \mathbf{y} \geq \mathbf{0}$ si ha $\mathbf{b}^T \mathbf{y} \geq 0$
- $A\mathbf{x} \leq \mathbf{b}$ ha una soluzione non negativa se e solo se per ogni $\mathbf{y} \in \mathbb{R}_+^m$ tale che $A^T \mathbf{y} \geq \mathbf{0}$ si ha $\mathbf{b}^T \mathbf{y} \geq 0$
- $A\mathbf{x} \leq \mathbf{b}$ ha una soluzione se e solo se esiste una soluzione $\mathbf{y} \in \mathbb{R}_+^m$ tale che $A^T \mathbf{y} = \mathbf{0}$ e $\mathbf{b}^T \mathbf{y} \geq 0$

Si può utilizzare il secondo punto di tale proposizione per dimostrare il teorema 8.

Dimostrazione (Teorema 8): Siano P, D una coppia di problemi primario-duale

$$P = \begin{cases} \max \mathbf{c}^T \mathbf{x} \\ A\mathbf{x} \leq \mathbf{b} \\ \mathbf{x} \geq 0 \end{cases} \quad D = \begin{cases} \min \mathbf{b}^T \mathbf{y} \\ A^T \mathbf{y} \geq \mathbf{c} \\ \mathbf{y} \geq 0 \end{cases}$$

si vuole mostrare che se P ha un ottimo, allora D ha lo stesso ottimo.



Sia \mathbf{x}^* la soluzione ottimale di P , e sia $\gamma = \mathbf{c}^T \mathbf{x}^*$ il valore massimizzato, il sistema di disequazioni

$$\begin{aligned} A\mathbf{x} &\leq \mathbf{b} \\ \mathbf{c}^T \mathbf{x} &\geq \gamma \end{aligned}$$

ha come unica soluzione \mathbf{x}^* , invece il sistema

$$\begin{aligned} A\mathbf{x} &\leq \mathbf{b} \\ \mathbf{c}^T \mathbf{x} &\geq \gamma + \varepsilon \end{aligned}$$

non ha soluzione per ogni $\varepsilon > 0$. Si definiscono

$$\hat{A} = \begin{bmatrix} A \\ -\mathbf{c}^T \end{bmatrix} \quad \hat{\mathbf{b}}_\varepsilon = \begin{bmatrix} \mathbf{b} \\ -\gamma - \varepsilon \end{bmatrix} \quad (3.5)$$

- $\hat{A}\mathbf{x} \leq \hat{\mathbf{b}}_0$ ha una soluzione, ed è \mathbf{x}^*
- $\hat{A}\mathbf{x} \leq \hat{\mathbf{b}}_\varepsilon$ con $\varepsilon > 0$ non ha soluzione

Si consideri il secondo punto della proposizione 14, ossia

$A\mathbf{x} \leq \mathbf{b}$ ha una soluzione non negativa se e solo se per ogni $\mathbf{y} \in \mathbb{R}_+^m$ tale che $A^T \mathbf{y} \geq \mathbf{0}$ si ha $\mathbf{b}^T \mathbf{y} \geq 0$

Dato che non esiste una soluzione per il sistema $\hat{A}\mathbf{x} \leq \hat{\mathbf{b}}_\varepsilon$, allora esiste un $\hat{\mathbf{y}} \in \mathbb{R}^m$ tale che $\hat{\mathbf{y}} \geq \mathbf{0}$ e

$$\hat{A}^T \hat{\mathbf{y}} \geq \mathbf{0} \quad (3.6)$$

$$\hat{\mathbf{y}}^T \hat{\mathbf{b}}_\varepsilon < 0 \quad (3.7)$$

Sia

$$\hat{\mathbf{y}} = \begin{bmatrix} \hat{\mathbf{u}} \\ z \end{bmatrix}$$

con $\hat{\mathbf{u}}$ un vettore e z uno scalare. allora

$$\hat{A}^T \hat{\mathbf{y}} \geq \mathbf{0} \implies \quad (3.8)$$

$$A^T \hat{\mathbf{u}} - \mathbf{c}^T z \geq \mathbf{0} \implies \quad (3.9)$$

$$A^T \hat{\mathbf{u}} \geq \mathbf{c}^T z \quad (3.10)$$

inoltre

$$\hat{\mathbf{y}}^T \hat{\mathbf{b}}_\varepsilon < 0 \implies \quad (3.11)$$

$$\hat{\mathbf{u}}^T \mathbf{b} + z(-\gamma - \varepsilon) < 0 \implies \quad (3.12)$$

$$\hat{\mathbf{u}}^T \mathbf{b} < z(\gamma + \varepsilon) \quad (3.13)$$

Attenzione, si ricordi che $\hat{\mathbf{y}} \geq \mathbf{0}$, quindi $z \geq 0$. Si consideri adesso il sistema

$$\hat{A}\mathbf{x} \leq \hat{\mathbf{b}}_0$$

dato che ha una soluzione, il secondo punto della proposizione 14 implica che per ogni \mathbf{y} tale che $A^T \mathbf{y} \geq \mathbf{0}$ si ha $\mathbf{y}^T \mathbf{b}_0 \geq 0$, ciò è valido anche per $\hat{\mathbf{y}}$

$$\hat{A}^T \hat{\mathbf{y}} \geq \mathbf{0} \implies \hat{\mathbf{y}}^T \mathbf{b}_0 \geq 0$$

ma $\hat{\mathbf{y}} = [\hat{\mathbf{u}} \quad z]^T$, quindi

$$\hat{\mathbf{y}}^T \mathbf{b}_0 \geq 0 \implies \quad (3.14)$$

$$[\hat{\mathbf{u}} \quad z] \begin{bmatrix} \mathbf{b} \\ -\gamma \end{bmatrix} \geq 0 \implies \quad (3.15)$$

$$\hat{\mathbf{u}}^T \mathbf{b} \geq z\gamma \quad (3.16)$$

Dato tale risultato ed il risultato precedente, sappiamo che $z > 0$ dato che

$$z\gamma \leq \hat{\mathbf{u}}^T \mathbf{b} < z(\gamma + \varepsilon) \quad (3.17)$$

essendo che $z \neq 0$, è possibile considerare il vettore

$$\mathbf{w} = \frac{1}{z} \mathbf{u}$$

Si ha che \mathbf{w} è ammissibile per il problema duale D :

$$A^T \mathbf{w} = A^T \left(\frac{1}{z} \mathbf{u} \right) \geq \mathbf{c} \frac{1}{z} z = \mathbf{c}$$

inoltre

$$\gamma \leq \mathbf{w}^T \mathbf{b} < \gamma + \varepsilon \implies \mathbf{w}^T \mathbf{b} = \gamma$$

\mathbf{w} è quindi la soluzione ottimale del problema duale. ■

3.2.1 interpretazione Geometrica del Lemma di Farkas

Definizione 26 Dati n punti $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n \in \mathbb{R}^m$, si definisce **cono convesso** generato da tali punti il seguente insieme

$$\left\{ \sum_{i=1}^n t_i \mathbf{a}_i : \{t_1, \dots, t_n\} \in \mathbb{R}_+^n \right\}$$

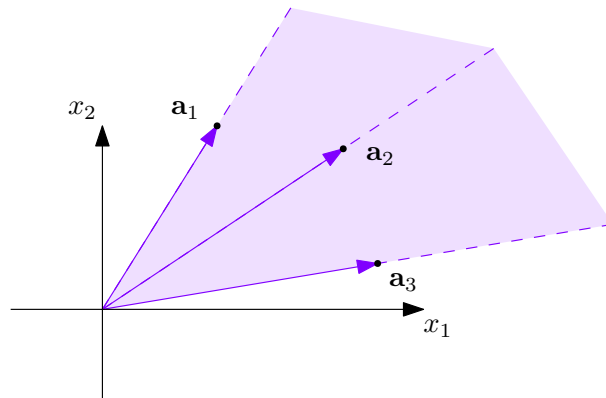


Figura 3.2: cono convesso in \mathbb{R}^2

Lemma 5 (Farkas Geometrico) Dati n punti $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n \in \mathbb{R}^m$, ed un punto $\mathbf{b} \in \mathbb{R}^m$, esattamente una fra le seguenti è vera

1. \mathbf{b} è nel cono convesso generato da $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$, ossia $\mathbf{b} \in \left\{ \sum_{i=1}^n t_i \mathbf{a}_i : \{t_1, \dots, t_n\} \in \mathbb{R}_+^n \right\}$
2. Esiste un iperpiano H passante per l'origine descritto da un vettore \mathbf{y}

$$H = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{y}^T \mathbf{x} = 0\}$$

tale per cui, per ogni i , $\mathbf{y}^T \mathbf{a}_i \geq 0$ e $\mathbf{y}^T \mathbf{b} < 0$, ossia, il cono convesso si trova in una delle due metà definite da H , e \mathbf{b} si trova nell'altra metà.

Presenteremo adesso alcuni richiami di analisi sugli insiemi reali necessari per la dimostrazione della versione geometrica del lemma di Farkas.

$Z \subseteq \mathbb{R}^n$ è un insieme *aperto* se per ogni $\mathbf{x} \in Z$, esiste una palla $B_\epsilon(\mathbf{x})$ centrata in \mathbf{x} di raggio ϵ (arbitrario) che è contenuta in Z . Si ricordi che

- l'unione di due insiemi aperti è un insieme aperto

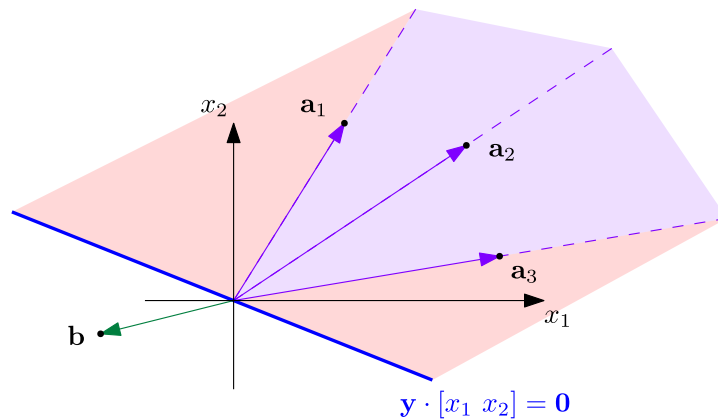


Figura 3.3: Secondo il lemma di Farkas, si può descrivere un iperpiano $\mathbf{y}^T \mathbf{x} = 0$ che divide il vettore \mathbf{b} dal cono convesso.

- l'intersezione di due insiemi aperti è un'insieme aperto

Lemma 6 se f è una funzione continua e biettiva su \mathbb{R}^n e $U \subseteq \mathbb{R}^n$ è aperto, l'insieme $f(U)$ è aperto. Se $Z \subset \mathbb{R}^n$ è un'insieme chiuso e limitato, allora $f(Z)$ è un'insieme chiuso.

Definizione 27 Dati n punti $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n \in \mathbb{R}^m$, il cono convesso generato da questi è **primitivo** se $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ sono linearmente indipendenti.

La dimostrazione del lemma 5 consiste nella dimostrazione di 5 proposizioni.

Proposizione 15 Se C è un cono convesso, allora C è l'unione di un'insieme finito di coni primitivi.

Dimostrazione: Siano $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n \in \mathbb{R}^m$ i punti che generano C , sia

$$\mathcal{I} = \{I \subseteq \{1, \dots, n\} \text{ t.c. } \{\mathbf{a}_i \mid i \in I\} \text{ sono linearmente indipendenti}\}$$

\mathcal{I} è un sotto-insieme dell'insieme potenza di $\{1, \dots, n\}$, ha quindi cardinalità $|\mathcal{I}| \leq 2^n$. Si denota C_I il cono primitivo generato dall'insieme $\{\mathbf{a}_i \mid i \in I\}$ con $I \in \mathcal{I}$. Si vuole mostrare che

$$\bigcup_{I \in \mathcal{I}} C_I = C$$

Sia \mathbf{x} un elemento fissato di C

$$\mathbf{x} = \sum \alpha_i \mathbf{a}_i, \quad \alpha_i \geq 0$$

Sia \mathbf{x} definito come segue: \mathbf{x} è l'elemento che massimizza il numero di coefficienti α_i uguali a zero, quindi ogni altro elemento di C è una combinazione lineare di \mathbf{a}_i che ha un numero di coefficienti nulli minore del numero di coefficienti nulli di \mathbf{x} .

Sia $I = \{i : \alpha_i > 0\}$, si pone per assurdo che i vettori \mathbf{a}_i con $i \in I$ non sono linearmente indipendenti, ossia

$$\exists \beta_i \in \mathbb{R} \quad \text{t.c.} \quad \sum_{i \in I} \beta_i \mathbf{a}_i = \mathbf{0} \quad (3.18)$$

Si consideri il particolare coefficiente

$$\alpha_i - t\beta_i$$

per qualche $t \in \mathbb{R}$, per definizione, $\alpha_i > 0$, scegliendo un valore di t molto piccolo, $\alpha_i - t\beta_i > 0$, se t cresce indefinitamente, ad un certo punto $\alpha_i - t\beta_i$ diverrà negativo. Si sceglie $t = \frac{\alpha_j}{\beta_j}$ con j l'indice che minimizza fra tutti il valore assoluto

$$\left| \frac{\alpha_j}{\beta_j} \right|$$

il vettore

$$\mathbf{x}^* = \sum_{i \in I} (\alpha_i - t\beta_i) \mathbf{a}_i = \sum_{i \in I} \left(\alpha_i - \frac{\alpha_j}{\beta_j} \beta_i \right) \mathbf{a}_i$$

è per definizione un elemento del cono convesso C dato che, per la specifica scelta di t , si ha che $(\alpha_i - t\beta_i) \geq 0$ per ogni $i \in I$, inoltre il vettore si può riscrivere

$$\begin{aligned}\mathbf{x}^* &= \sum_{i \in I} \alpha_i \mathbf{a}_i + t \sum_{i \in I} \beta_i \mathbf{a}_i = \\ &= \mathbf{x} + t \sum_{i \in I} \beta_i \mathbf{a}_i\end{aligned}$$

La scelta di t rende uno dei coefficienti $(\alpha_i - t\beta_i)$ uguali a zero, quindi \mathbf{x}^* ha un coefficiente nullo in più rispetto ad \mathbf{x} , ma questo per definizione massimizzava il numero di coefficienti nulli, è quindi impossibile che l'assunzione che i vettori \mathbf{a}_i con $i \in I$ non siano linearmente indipendenti sia vera, quindi \mathbf{a}_i con $i \in I$ sono linearmente indipendenti, allora \mathbf{x} è nel cono primitivo C_I , quindi $\bigcup_{I \in \mathcal{I}} C_I = C$. ■

Proposizione 16 *Ogni cono primitivo è un'insieme chiuso.*

Dimostrazione: Sia C_I un cono primitivo generato da \mathbf{a}_i (linearmente indipendenti) con $i \in I$. si consideri la base canonica di \mathbb{R}^n

$$\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

dove \mathbf{e}_i è il vettore composto da tutti zeri fatta eccezione per l' i -esima componente che è uguale ad 1. Sia \bar{C} il cono convesso generato da $\mathbf{e}_1, \dots, \mathbf{e}_n$, questo è chiuso perché è l'intersezione di un'insieme finito di mezzi spazi. Si consideri la funzione $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ definita come segue

$$f([x_i : i \in I]^T) = \sum_{i \in I} x_i \mathbf{a}_i$$

essendo che \mathbf{a}_i , $i \in I$ sono linearmente indipendenti, f è continua ed iniettiva, ma allora, essendo che $f(\bar{C}) = C_I$, per il lemma 6, C_I è chiuso.

Proposizione 17 *Ogni cono convesso è un'insieme chiuso. La dimostrazione è banale e segue dalle proposizioni 15 e 16.*

Proposizione 18 *Se $X \subset \mathbb{R}^n$ è un'insieme chiuso e $\mathbf{b} \notin X$, allora esiste un punto $\mathbf{z} \in X$ che minimizza la distanza euclidea con \mathbf{b} , ossia è il **più vicino** a \mathbf{b} .*

Dimostrazione: Sia \mathbf{x} un punto in X , si consideri la sfera $B_\varepsilon(\mathbf{b})$ centrata in \mathbf{b} di raggio

$$\varepsilon = \|\mathbf{b} - \mathbf{x}\|$$

$B_\varepsilon(\mathbf{b})$ è un'insieme chiuso, l'insieme $B_\varepsilon(\mathbf{b}) \cap X$ è chiuso e limitato, si consideri la funzione $f: \mathbb{R}^n \rightarrow \mathbb{R}$ definita come segue

$$f(\mathbf{y}) = \|\mathbf{b} - \mathbf{y}\|$$

f è continua in $B_\varepsilon(\mathbf{b}) \cap X$ ed ha un minimo, tale punto è proprio il punto più vicino a \mathbf{b} .

Proposizione 19 *Se C è un cono convesso di \mathbb{R}^n e $\mathbf{b} \notin C$, allora esiste $\mathbf{z} \in C$ tale che \mathbf{z} è l'elemento di C più vicino a \mathbf{b} . La dimostrazione segue dalle proposizioni 17 e 18.*

Dimostrazione (Lemma 5): Sia C il cono convesso, si assume che $\mathbf{b} \notin C$, per la proposizione esiste un punto nel cono che è il più vicino a \mathbf{b} .

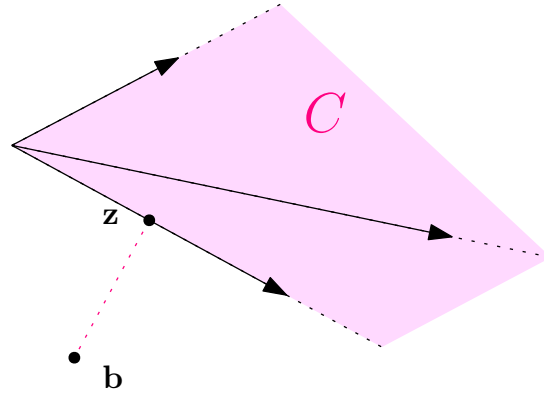
Si considera l'iperpiano

$$H = \{\mathbf{y} \in \mathbb{R}^m : (\mathbf{z} - \mathbf{b})^T \mathbf{y} = 0\}$$

Si noti che $(\mathbf{z} - \mathbf{b})$ è ortogonale a \mathbf{z} , se così non fosse, esisterebbe un punto in C più vicino a \mathbf{b} . Si vuole mostrare che $(\mathbf{z} - \mathbf{b})^T \mathbf{b} < 0$

$$\mathbf{z} - \mathbf{b} \neq 0 \implies (\mathbf{z} - \mathbf{b})^T (\mathbf{z} - \mathbf{b}) > 0 \implies \quad (3.19)$$

$$(\mathbf{z} - \mathbf{b})^T \mathbf{z} - (\mathbf{z} - \mathbf{b})^T \mathbf{b} > 0 \quad (3.20)$$

Figura 3.4: \mathbf{z} è il punto in C più vicino a \mathbf{b}

ma $(\mathbf{z} - \mathbf{b})^T \mathbf{z} = 0$ data l'ortogonalità, quindi

$$(\mathbf{z} - \mathbf{b})^T \mathbf{z} - (\mathbf{z} - \mathbf{b})^T \mathbf{b} > 0 \implies \quad (3.21)$$

$$-(\mathbf{z} - \mathbf{b})^T \mathbf{b} > 0 \implies (\mathbf{z} - \mathbf{b})^T \mathbf{b} < 0 \quad (3.22)$$

Questo dimostra che \mathbf{b} non si trova sull'iperpiano ma su uno dei due mezzi-spazi definiti da esso. Inoltre, per ogni \mathbf{a}_i si ha che

$$(\mathbf{z} - \mathbf{b})^T \mathbf{a}_i = \|\mathbf{z} - \mathbf{b}\| \cdot \|\mathbf{a}_i\| \cos \theta$$

ed ogni termine di tale prodotto è positivo, quindi $(\mathbf{z} - \mathbf{b})^T \mathbf{a}_i > 0$, allora i vettori \mathbf{a}_i , e conseguentemente tutto il cono convesso da loro generato, si trovano sull'altra metà di spazio definita da H . ■

3.3 Il Teorema degli Slack Complementari

Si consideri una coppia P, D di problemi primario-duale

$$P = \begin{cases} \max \mathbf{c}^T \mathbf{x} \\ A\mathbf{x} \leq \mathbf{b} \\ \mathbf{x} \geq 0 \end{cases} \quad D = \begin{cases} \min \mathbf{b}^T \mathbf{y} \\ A^T \mathbf{y} \geq \mathbf{c} \\ \mathbf{y} \geq 0 \end{cases} \quad (3.23)$$

I vincoli in P corrispondono a variabili in D , e vincoli in D corrispondono a variabili in P . Data una soluzione \mathbf{x}^* , un vincolo

$$x_1 a_{j1} + \dots x_n a_{jn} \leq b_j$$

si dice **slack** o **residuo** se

$$x_1^* a_{j1} + \dots x_n^* a_{jn} < b_j$$

Ossia il vincolo soddisfa la disuguaglianza in maniera *stretta*, ossia $A_j^T \mathbf{x}^*$ è strettamente minore di b_j . Se invece

$$x_1^* a_{j1} + \dots x_n^* a_{jn} = b_j$$

il vincolo per \mathbf{x}^* si dice **binding** o **saturo**. Tale denotazione si adotta anche ai vincoli del tipo $\mathbf{x} \geq \mathbf{0}$, infatti se

$$x_i^* > 0$$

la variabile x_i^* è slack, altrimenti se

$$x_i^* = 0$$

è binding. In tale contesto, il termine slack non è da confondere con le variabili slack che si aggiungono ad un programma lineare per farlo passare dalla forma canonica alla forma d'equazione. Si noti che la *slackness* di ogni variabile è riferita ad una specifica soluzione.

Il seguente teorema è fondamentale, descrive la relazione fra vincoli di un problema e variabili del suo duale nei termini della *slackness*.



Teorema 9 Siano P, D una coppia di problemi primario-duale definiti come in 3.23, entrambi ammissibili e limitati, sia \mathbf{x}^* la soluzione ottimale per P , e \mathbf{y}^* la soluzione ottimale per D , allora si verifica una delle seguenti

1. Se $x_i^* > 0$, ossia la variabile x_i è slack per \mathbf{x}^* , allora il vincolo i -esimo di D per \mathbf{y}^* è binding.
2. Se il j -esimo vincolo di P è slack per \mathbf{x}^* , allora $y_j^* = 0$, ossia la variabile y_j è binding per \mathbf{y}^* .
3. Se $y_j^* > 0$, ossia la variabile y_j è slack per \mathbf{y}^* , allora il vincolo j -esimo di P per \mathbf{x}^* è binding.
4. Se l' i -esimo vincolo di D è slack per \mathbf{y}^* , allora $x_i^* = 0$, ossia la variabile x_i è binding per \mathbf{x}^* .

Dimostrazione: per ogni soluzione ammissibile $\bar{\mathbf{x}}$ per P , e $\bar{\mathbf{y}}$ per D si ha

$$A\bar{\mathbf{x}} \leq \mathbf{b}$$

moltiplicando tutto per $\bar{\mathbf{y}}^T$

$$\bar{\mathbf{y}}^T A\bar{\mathbf{x}} \leq \bar{\mathbf{y}}^T \mathbf{b}$$

Analogamente, essendo

$$A^T \bar{\mathbf{y}} \geq \mathbf{c}$$

si ha

$$A^T \bar{\mathbf{y}} \bar{\mathbf{x}} \geq \mathbf{c}^T \bar{\mathbf{x}}$$

Quindi è valida la seguente

$$\mathbf{c}^T \bar{\mathbf{x}} \leq \bar{\mathbf{y}}^T A\bar{\mathbf{x}} \leq \bar{\mathbf{y}}^T \mathbf{b} \quad (3.24)$$

essendo che duale e primario hanno lo stesso ottimo

$$\mathbf{c}^T \mathbf{x}^* = \mathbf{b}^T \mathbf{y}^*$$

si ha che

$$\mathbf{c}^T \mathbf{x}^* = \mathbf{y}^{*T} A\mathbf{x}^* = \mathbf{b}^T \mathbf{y}^* \quad (3.25)$$

Considerando la prima disuguaglianza

$$\mathbf{c}^T \mathbf{x}^* = \mathbf{y}^{*T} A\mathbf{x}^* \implies \quad (3.26)$$

$$\mathbf{c}^T \mathbf{x}^* - \mathbf{y}^{*T} A\mathbf{x}^* = 0 \implies \quad (3.27)$$

$$(\mathbf{c}^T - \mathbf{y}^{*T} A)\mathbf{x}^* = 0 \implies \quad (3.28)$$

$$\sum_{i=1}^n (c_i - \mathbf{y}^{*T} A_i) x_i^* = 0 \quad (3.29)$$

ma $(c_i - \mathbf{y}^{*T} A_i)$ è minore o uguale di zero per ogni i , dato il vincolo $A^T \mathbf{y}^* \geq \mathbf{c}$, e x_i^* è maggiore o uguale a zero per ogni i , dato il vincolo $\mathbf{x} \geq 0$.

- ogni termine della sommatoria 3.29 è il prodotto fra un numero maggiore o uguale a zero per uno minore o uguale a zero.
- la somma totale è zero
- ogni prodotto $(c_i - \mathbf{y}^{*T} A_i) x_i^*$ della sommatoria è quindi uguale a zero

Quindi $x_i^* > 0$ (l' i -esima variabile di P è slack), necessariamente $c_i - \mathbf{y}^{*T} A_i = 0 \implies \mathbf{y}^{*T} A_i = c_i$ (l' i -esimo vincolo di D è binding).

Analogamente, se $c_i - \mathbf{y}^{*T} A_i < 0 \implies \mathbf{y}^{*T} A_i < c_i$ (l' i -esimo vincolo di D è slack), necessariamente $x_i^* = 0$ (l' i -esima variabile di P è binding). Quindi i punti (1) e (4) del teorema sono dimostrati, i restanti punti possono essere dimostrati in maniera analoga. ■

3.3.1 Alcune Applicazioni

Il teorema 9 può essere utilizzato per trovare la soluzione di un programma lineare, data una soluzione per il suo duale. Siano P e D una coppia di problemi primario-duale, sia \mathbf{x}^* la soluzione ottimale per P , se $x_i^* > 0$ allora l' i -esimo vincolo di D è binding, quindi in forma di equazione, inoltre, in base a quali vincoli di P sono binding, si possono porre le variabili della soluzione \mathbf{y}^* di D uguali a zero, costruendo un sistema di k equazioni in k variabili.

Esempio: Si consideri la seguente coppia di problemi primario duale:

$$P = \begin{cases} \max & 2x_1 + 4x_2 + 3x_3 + x_4 \\ & 3x_1 + x_2 + x_3 + 4x_4 \leq 12 \\ & x_1 - 3x_2 + 2x_3 + 3x_4 \leq 7 \\ & 2x_1 + x_2 + 3x_3 - x_4 \leq 10 \\ & x_1, x_2, x_3, x_4 \geq 0 \end{cases} \quad D = \begin{cases} \min & 12y_1 + 7y_2 + 10y_3 \\ & 3y_1 + y_2 + 2y_3 \geq 2 \\ & y_1 - 3y_2 + y_3 \geq 4 \\ & y_1 + 2y_2 + 3y_3 \geq 3 \\ & 4y_1 + 3y_2 - y_3 \geq 1 \\ & y_1, y_2, y_3 \geq 0 \end{cases} \quad (3.30)$$

Si consideri la soluzione ottimale $\mathbf{x}^* = [0 \ 10.4 \ 0 \ 0.4]^T$ per P , dato che $x_2^*, x_4^* > 0$, per il teorema 9 è vero che, se \mathbf{y}^* è la soluzione ottimale per D si ha

$$\begin{aligned} y_1 - 3y_2 + y_3 &= 4 \\ 4y_1 + 3y_2 - y_3 &= 1 \end{aligned}$$

Inoltre si noti come il secondo vincolo di P per \mathbf{x}^* è slack

$$\begin{aligned} x_1^* - 3x_2^* + 2x_3^* + 3x_4^* &= \\ -3 \cdot 10.4 + 3 \cdot 0.4 &= -28.6 < 7 \end{aligned}$$

Quindi $y_2^* = 0$, si ha il seguente sistema di equazioni

$$\begin{cases} y_1 - 3y_2 + y_3 = 4 \\ 4y_1 + 3y_2 - y_3 = 1 \\ y_2^* = 0 \end{cases} \implies \begin{cases} y_1 + y_3 = 4 \\ 4y_1 - y_3 = 1 \\ y_2^* = 0 \end{cases} \implies \mathbf{y}^* = [1 \ 0 \ 3]^T$$

Un'altra possibilità è quella di usare il teorema 9 per verificare che una data soluzione ammissibile sia ottimale. Proseguendo in tale maniera

1. Data una soluzione \mathbf{x}^* da verificare, si assume che sia ottimale
 2. Si calcola tramite il teorema 9 la soluzione \mathbf{y}^* ottimale per il suo duale
 3. Si verifica che le due soluzioni assumano lo stesso valore sulle funzioni obiettivo dei due problemi
- si consideri la seguente coppia di problemi primario duale

$$P = \begin{cases} \max & 2x_1 + 16x_2 + 12x_3 \\ & 2x_1 + x_2 - x_3 \leq 3 \\ & -3x_1 + 8x_2 + 2x_3 \leq 12 \\ & x_1, x_2, x_3 \geq 0 \end{cases} \quad D = \begin{cases} \min & 3y_1 + 12y_2 \\ & 2y_1 - 3y_2 \geq 2 \\ & y_1 + 8y_2 \geq 16 \\ & -y_1 + 2y_2 \geq 12 \\ & y_1, y_2 \geq 0 \end{cases} \quad (3.31)$$

Si vuole verificare che $\mathbf{x}^* = [6 \ 0 \ 12]^T$ sia ottimale, si assume che lo sia e si calcola la soluzione ottimale per D .

- Si noti come entrambi i vincoli di P per \mathbf{x}^* sono slack, quindi $y_1^*, y_2^* = 0$

la soluzione ottimale per D dovrebbe essere $\mathbf{y}^* = [0 \ 0]^T$, ma questa non soddisfa i vincoli di D , quindi è impossibile che \mathbf{x}^* sia ottimale.

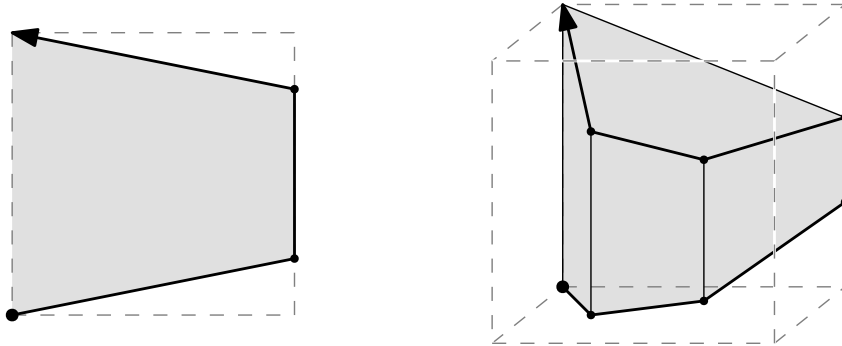
CAPITOLO

4

METODI POLINOMIALI PER LA RISOLUZIONE DI UN LP

Esiste uno specifico poliedro definito su \mathbb{R}^n che soddisfa le seguenti

- il poliedro è definito da $3n$ equazioni
- esiste un cammino sui vertici di lunghezza $2^n - 1$ che collega due punti



Tale risultato è stato dimostrato da *Klee* e *Minty*, l'implicazione è chiara, nel caso peggiore il metodo del sempliceo esegue un numero di passi esponenziali nella dimensione dell'input. È stato poi dimostrato che se venissero ordinate in maniera casuale le n variabili del problema, per poi applicare la regola di Bland, il valore atteso del numero di pivot controllati nell'esecuzione sarebbe $\exp(C\sqrt{n \ln(n)})$, dove C è una data costante.

Definizione 28 Il **diametro** di un poliedro è la massima distanza (in termini di cammino sui vertici) fra due vertici qualsiasi del poliedro.

Congettura di Hirsh ogni poliedro in \mathbb{R}^n definito da cn vincoli (c costante) ha un diametro minore o uguale a n^c .

Esistono degli algoritmi polinomiali per la risoluzione di un programma lineare, è necessario descrivere prima quale sia la dimensione dell'input di un LP. Un programma lineare è descritto da una matrice e due vettori

$$A \in \text{Mat}(m \times n), \mathbf{b} \in \mathbb{R}^m, \mathbf{c} \in \mathbb{R}^n$$

Quanti bit servono per descrivere tali elementi? Per descrivere un singolo numero intero α , sono necessari $\lceil \log_2(\alpha) + 1 \rceil$ bit, si denota tale grandezza $\langle \alpha \rangle$.

assunzione: Tutte le componenti di A , \mathbf{b} e \mathbf{c} sono numeri razionali

Tale assunzione è fondamentale per far sì che i numeri di A , \mathbf{b} e \mathbf{c} si possano rappresentare con un numero finito di bit. Un qualsiasi numero razionale $\frac{p}{q}$ si può rappresentare come frazione di due numeri interi, quindi

$$\langle \frac{p}{q} \rangle = \langle p \rangle + \langle q \rangle$$

le dimensioni di un vettore $\mathbf{x} \in \mathbb{R}^d$ si esprimono come somma delle dimensioni dei suoi componenti

$$\langle \mathbf{x} \rangle = \sum_{i=1}^d \langle x_i \rangle$$

analogamente, si descrivono le dimensioni della matrice A come segue

$$\langle A \rangle = \sum_{i=1}^n \sum_{j=1}^m \langle A_{ij} \rangle$$

dato un programma lineare P in forma canonica, la dimensione del problema è

$$\langle P \rangle = \langle A \rangle + \langle \mathbf{b} \rangle + \langle \mathbf{c} \rangle$$

L'algoritmo seguente farà un numero polinomiale in $\langle P \rangle$ di operazioni bit a bit. Non descriveremo un algoritmo che fa un numero di operazioni polinomiale in n, m , in quel caso, sarebbe **fortemente polinomiale**.

4.1 Metodo degli Ellissoidi

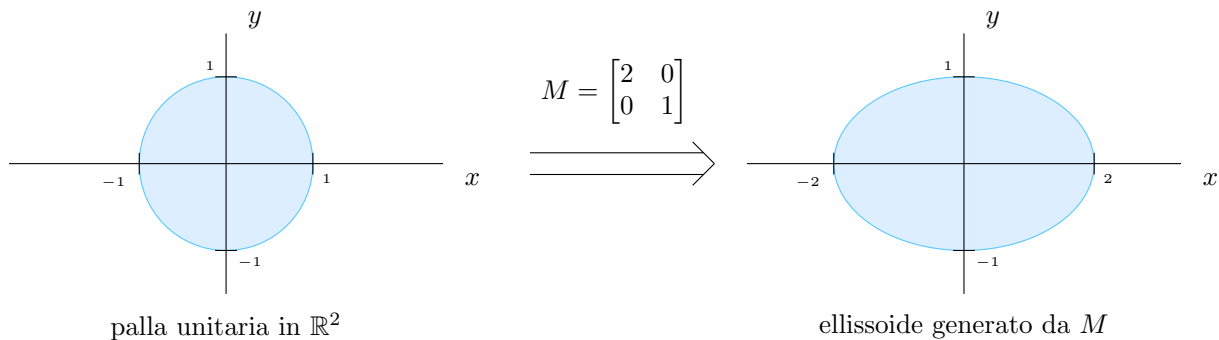
Sia B^n la palla unitaria di \mathbb{R}^n

$$B^n = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| \leq 1\}$$

Definizione 29 Data una matrice M non singolare ed un vettore \mathbf{s} (detto centro), un'**ellissoide** è il seguente insieme

$$E = \{M\mathbf{x} + \mathbf{s} : \mathbf{x} \in B^n\}$$

geometricamente, un'ellissoide è una sfera che è stata stirata dall'applicazione lineare M e centrata in \mathbf{s} .



Un ellissoide E si può definire anche in maniera differente

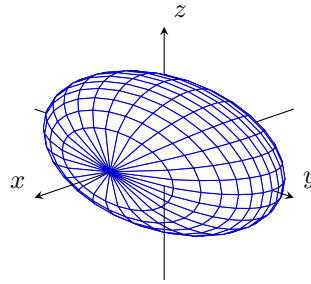
$$E = \{\mathbf{y} \in \mathbb{R}^n : M^{-1}(\mathbf{y} - \mathbf{s}) \in B^n\} =$$

$$E = \{\mathbf{y} \in \mathbb{R}^n : (\mathbf{y} - \mathbf{s})^T M^{-1T} M^{-1}(\mathbf{y} - \mathbf{s}) \leq 1\} =$$

$$E = \{\mathbf{y} \in \mathbb{R}^n : (\mathbf{y} - \mathbf{s})^T Q(\mathbf{y} - \mathbf{s}) \leq 1\}$$

dove $Q = M^{-1T} M^{-1}$ è una matrice quadrata **positiva semi-definita**, ossia

$$\mathbf{x}^T Q \mathbf{x} > 0, \forall \mathbf{x}$$


 Figura 4.1: ellissoide in \mathbb{R}^3

Teorema 10 Q è una matrice positiva semi-definita se e solo se esiste una matrice M non singolare tale che

$$Q = MM^T$$

Corollario: $E \subset \mathbb{R}^n$ è un ellissoide se e solo se esiste una matrice positiva semi-definita Q ed un vettore $\mathbf{s} \in \mathbb{R}^n$ tale che

$$E = \{\mathbf{y} \in \mathbb{R}^n : (\mathbf{y} - \mathbf{s})^T Q (\mathbf{y} - \mathbf{s}) \leq 1\}$$

Il metodo degli ellissoidi non risolverà direttamente un programma lineare, ma si occuperà di verificare se un dato poliedro è o non è vuoto, è stato già dimostrato in precedenza che i due problemi sono equivalenti. Una descrizione iniziale del funzionamento dell'algoritmo è la seguente

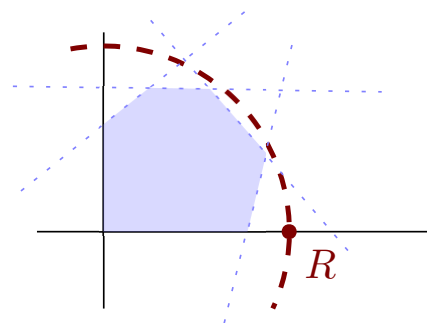
- L'algoritmo prende come input un poliedro P descritto da una matrice A ed un vettore \mathbf{b}

$$P = \{\mathbf{x} : A\mathbf{x} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$$

e due numeri razionali R, ε , l'assunzione è che P sia contenuto interamente nella palla $B^n(\mathbf{0}, R)$ di raggio R centrata nell'origine.

$$B^n(\mathbf{0}, R) = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| \leq R\}$$

- l'algoritmo genera una serie di t ellissoidi E_1, E_2, \dots, E_t tali che $P \subset E_i$ per ogni i
- l'algoritmo termina restituendo una delle seguenti risposte
 - è stato trovato un punto in P , ed è il centro di E_t
 - il volume di E_t è minore di ε


 Figura 4.2: l'assunzione che il poliedro sia contenuto nella palla di raggio R

In particolare, se l'algoritmo non trova un punto in P , restituisce un *certificato* che il poliedro non contiene una palla di raggio ε .

Nell'algoritmo, l'ellissoide iniziale E_0 è proprio $B^n(\mathbf{0}, R)$, denotiamo \mathbf{s}_i il centro dell' i -esimo ellissoide, in tal caso, $\mathbf{s}_0 = \mathbf{0}$. Si assuma di essere nel k -esimo passo, e si a disposizione l'ellissoide E_k centrato in \mathbf{s}_k . Si verifica che \mathbf{s}_k sia un'elemento di P

$$A\mathbf{s}_k \leq \mathbf{b}$$

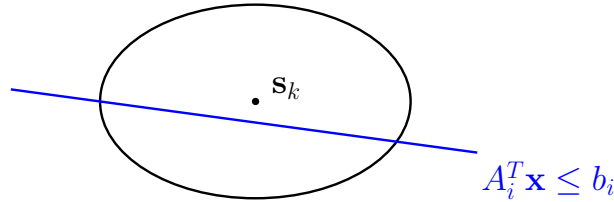
se ciò dovesse essere vero, si restituirebbe \mathbf{s}_k , altrimenti, si avrebbe un vincolo che tale punto non soddisfa

$$A_i^T \mathbf{s}_k > b_i \quad \text{per qualche } i$$

Il vincolo i -esimo definisce un iperpiano, e a sua volta un mezzo spazio

$$\{\mathbf{x} : A_i^T \mathbf{x} \leq b_i\}$$

\mathbf{s}_k violando il vincolo, non si trova nel mezzo-spazio appena definito.

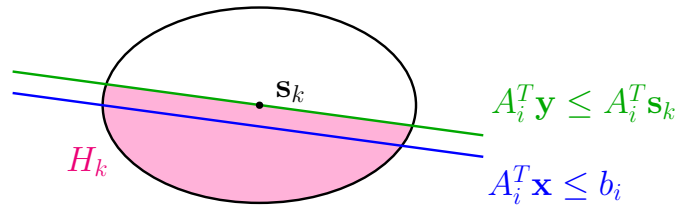


Si considera l'iperpiano parallelo a $\{\mathbf{x} : A_i^T \mathbf{x} = b_i\}$ ma passante per \mathbf{s}_k , ossia quello che genera il mezzo spazio

$$\{\mathbf{y} : A_i^T \mathbf{y} \leq A_i^T \mathbf{s}_k\}$$

e si considera l'insieme H_k che è intersezione di tale mezzo spazio e E_k

$$H_k = E_k \cap \{\mathbf{y} : A_i^T \mathbf{y} \leq A_i^T \mathbf{s}_k\}$$



Si sceglie il prossimo ellissoide E_{k+1} come il più piccolo ellissoide contenente H_k , e si avrà che $P \subset E_{k+1}$. La seguente proposizione non verrà dimostrata, ma è fondamentale per la descrizione dell'algoritmo.

Proposizione 20 Dato l'ellissoide $E_k \subset \mathbb{R}^n$ descritto da un centro \mathbf{s}_k che viola l' i -esimo vincolo di A

$$A_i^T \mathbf{s}_k > b_i$$

e Q_k matrice positiva semi-definita, Il più piccolo ellissoide contenete

$$H_k = E_k \cap \{\mathbf{y} : A_i^T \mathbf{y} \leq A_i^T \mathbf{s}_k\}$$

è univocamente definito dal centro \mathbf{s}_{k+1} e dalla matrice Q_{k+1} come segue

$$\mathbf{s}_{k+1} = \mathbf{s}_k - \frac{1}{n-1} \frac{1}{\sqrt{A_i Q_k A_i^T}} Q_k A_i^T \quad (4.1)$$

$$Q_{k+1} = \frac{n^2}{n^2+1} \left(Q_k - \frac{2}{n+1} \frac{Q_k A_i^T A_i Q_k}{A_i Q_k A_i^T} \right) \quad (4.2)$$

Il volume dell'ellissoide E_k è l'integrale della funzione identità su di esso

$$\text{vol}(E_k) = \int_{E_k} d\mathbf{x}$$

e passando da un'ellissoide al successivo nell'algoritmo tale volume decresce di un valore costante

$$\frac{\text{vol}(E_{k+1})}{\text{vol}(E_k)} \leq e^{-\frac{1}{2n+2}} < 1$$

Durante la successione di ellissoidi, se il volume dell'ellissoide corrente è minore del volume della palla di raggio ε dato in ingresso, l'algoritmo termina senza restituire una soluzione, garantendo però che P non è contenuto nella palla di raggio ε .

Proposizione 21 *L'algoritmo termina in un numero di passi polinomiale in funzione di $\langle A \rangle + \langle \mathbf{b} \rangle + \langle R \rangle + \langle \varepsilon \rangle$.*

Dimostrazione: Il volume di $E_0 = B^n(\mathbf{0}, R)$ è $c \cdot R^n$ per qualche costante c che dipende da n , inoltre per ogni k si ha

$$\text{vol}(E_k) \leq \text{vol}(E_{k+1})e^{-\frac{k}{2n+2}}$$

se $k \geq n(2n+2)\ln(\frac{R}{\varepsilon})$ si ha che

$$\text{vol}(E_k) \leq cR^n \cdot e^{\lceil n \cdot \ln(\frac{R}{\varepsilon}) \rceil} \quad (4.3)$$

$$cR^n \cdot e^{\lceil n \cdot \ln(\frac{R}{\varepsilon}) \rceil} = \frac{cR^n - \frac{R^n}{\varepsilon^n}}{e} \leq \varepsilon \quad (4.4)$$

Quindi $\lceil n(2n+2)\ln(\frac{R}{\varepsilon}) \rceil$ è il massimo numero di passi del metodo degli ellissoidi, ed è polinomiale in $\langle A \rangle + \langle \mathbf{b} \rangle + \langle R \rangle + \langle \varepsilon \rangle$. ■

4.1.1 Tecnicità Implementative

Ci sono alcune criticità nel metodo degli ellissoidi, il calcolo del centro dell'ellissoide successivo avviene considerando il termine

$$\mathbf{s}_{k+1} = \mathbf{s}_k - \frac{1}{n-1} \frac{1}{\sqrt{A_i Q_k A_i^T}} Q_k A_i^T$$

questo presenta una radice quadrata quindi viola l'assunzione che tutti i numeri rappresentati siano razionali, è necessario includere un'ulteriore argomento in input, ε_2 necessario per arrotondare \mathbf{s}_{k+1} in modo da far sì che ogni sua componente sia un numero razionale. Vi è un'altro problema, come viene scelto il valore di R iniziale per far sì che la palla $B^n(\mathbf{0}, R)$ contenga completamente il poliedro P ? Siano

φ e l due valori definiti come segue

$$\varphi = \langle A \rangle + \langle \mathbf{b} \rangle \quad l = 2^\varphi \quad (4.5)$$

Proposizione 22 *Assumendo che ogni numero considerato sia razionale, il problema $A\mathbf{x} \leq \mathbf{b}$, $\mathbf{x} \geq \mathbf{0}$ ha una soluzione se e solo se il problema*

$$\begin{cases} A\mathbf{x} \leq \mathbf{b} \\ \mathbf{x} \geq \mathbf{0} \\ x_i \leq l, \quad \forall i \end{cases}$$

ha soluzione.

Dimostrazione: sia \mathbf{x}^* una soluzione per il problema $A\mathbf{x} \leq \mathbf{b}$, $\mathbf{x} \geq \mathbf{0}$, essendo \mathbf{x}^* una soluzione ad un sistema di equazioni nei termini di A e \mathbf{b} , il numero di bit per rappresentare ogni x_i^* è al più $\langle A \rangle + \langle \mathbf{b} \rangle = \varphi$, quindi $x_i^* \leq 2^\varphi = l$. ■

Quindi, scegliendo $R = l$, si ha la certezza che ogni possibile punto del poliedro sia contenuto nella palla $B^n(\mathbf{0}, l)$.

Vi è un'ultima tecnicità da considerare, l'algoritmo potrebbe terminare senza restituire un punto in P anche se quest'ultimo è non vuoto, dato che termina appena l'ellissoide k -esimo considerato ha un volume minore del volume della palla di raggio ε , e l'unica certezza che si ha da tale risultato è che il poliedro P non contiene una palla di raggio ε .

Si definiscono

$$\eta = 2^{-5\varphi}$$

$$\mathbf{z} = [\eta \quad \eta \quad \dots \quad \eta]^T \in \mathbb{R}^n$$

e si sceglie $\varepsilon = 2^{-6\varphi}$. Ciò è funzionale dato che il sistema

$$A\mathbf{x} \leq \mathbf{b}$$

ha soluzione se e solo se il sistema

$$A\mathbf{x} \leq \mathbf{b} + \mathbf{z}$$

ha soluzione, in tal caso l'insieme delle soluzioni ammissibili contiene necessariamente una palla di raggio ε .

In conclusione, per verificare che un poliedro

$$P = \{\mathbf{x} : A\mathbf{x} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$$

sia o non sia vuoto, è sufficiente eseguire il metodo degli ellissoidi sul poliedro derivato

$$P' = \{\mathbf{x} : A\mathbf{x} \leq \mathbf{b} + \mathbf{z}, -l \leq \mathbf{x} \leq l\}$$

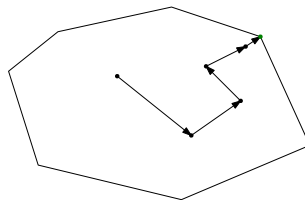
Se l'algoritmo restituisce un punto, allora P non è vuoto, altrimenti si ha la certezza che P sia vuoto. Tale metodo è polinomiale nelle dimensioni di $\langle A \rangle + \langle \mathbf{b} \rangle$.

4.2 Accenno al Metodo dei Punti Interni

Il metodo degli ellissoidi non si presta ad essere efficiente se implementato su un calcolatore, nell'effettivo vengono utilizzate delle rivisitazioni del simpleso oppure il metodo dei punti interni, quest'ultimo consiste nello svilupparsi di un cammino di punti contenuti nel poliedro P , muovendosi verso un punto ottimale per la funzione obiettivo. Il problema dato in input è quindi un programma lineare del tipo

$$P = \{\mathbf{x} : A\mathbf{x} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$$

$$\max \mathbf{c}^T \mathbf{x} \text{ su } P$$



Si considera una funzione $f(\eta, \mathbf{x})$ con $\mathbf{x} \in \mathbb{R}^n$ e $\eta \in [0, \infty)$ definita come segue

$$f(\eta, \mathbf{x}) = \mathbf{c}^T \mathbf{x} + \eta \sum_{i=1}^n \ln(b_i - A_i \mathbf{x}) \quad (4.6)$$

se $\eta = 0$ allora $f(\eta, \mathbf{x}) = f(0, \mathbf{x}) = \mathbf{c}^T \mathbf{x}$ è la funzione obiettivo, dal momento che il logaritmo non è definito in 0, $f(\eta, \mathbf{x})$ non è definita sulla frontiera di P , ossia nei punti che annullano l'argomento del logaritmo

$$b_i = A_i \mathbf{x} \implies b_i - A_i \mathbf{x} = 0$$

per ogni η fissato, $f(\eta, \mathbf{x})$ tende a $-\infty$ se \mathbf{x} si avvicina alla frontiera di P , se η tende a zero, $f(\eta, \mathbf{x})$ tende a $\mathbf{c}^T \mathbf{x}$.

Proposizione 23 per ogni η fissato, $f(\eta, \mathbf{x})$ ha un'unico massimo in P .

Senza entrare nel dettaglio, esistono dei metodi di ottimizzazione non lineare basati sui *moltiplicatori di Lagrange* per la ricerca di un massimo di $f(\eta, \mathbf{x})$, che porterà poi alla ricerca di un massimo per $\mathbf{c}^T \mathbf{x}$ su P .

CAPITOLO

5

PROGRAMMAZIONE INTERA

La programmazione intera riguarda la soluzione di un programma lineare, con il vincolo ulteriore che la soluzione deve essere un vettore di numeri interi, il generico programma intero è il seguente

$$\begin{aligned} \max \quad & \mathbf{c}^T \mathbf{x} \\ \text{A} \mathbf{x} \leq & \mathbf{b} \\ \mathbf{x} \geq & \mathbf{0} \\ \mathbf{x} \in & \mathbb{Z}^n \end{aligned} \tag{5.1}$$

Il noto **Knapsack Problem** è un classico esempio di problema che si può risolvere tramite la programmazione intera, vi è un'insieme di n oggetti, x_1, x_2, \dots, x_n , ad ogni oggetto x_i è associato un valore v_i ed un peso w_i , e si vuole trovare il sotto-insieme di oggetti I che massimizza il valore complessivo

$$\sum_{i \in I} v_i$$

con il vincolo che il peso complessivo sia minore o uguale ad una certa costante C .

$$\sum_{i \in I} w_i \leq C$$

Il problema si modella considerando un vettore \mathbf{x} che rappresenta gli oggetti, se $x_i = 1$ allora l' i -esimo oggetto è considerato, altrimenti no

$$x_i = \begin{cases} 1 & \text{se } i \in I \\ 0 & \text{altrimenti} \end{cases}$$

sia \mathbf{v} il vettore la cui i -esima componente v_i è il valore dell' i -esimo oggetto, e \mathbf{w} il vettore la cui i -esima componente w_i è il peso dell' i -esimo oggetto. Il problema si modella come segue

$$\begin{aligned} \max \quad & \mathbf{v}^T \mathbf{x} \\ \mathbf{w}^T \mathbf{x} \leq & C \\ \mathbf{x} \in & \{0, 1\}^n \end{aligned}$$

5.1 Inviluppo Convesso dei Punti Interi

In questa sezione verrà presentato un metodo per la risoluzione di un programma intero. La programmazione intera è tipicamente NP-completa, non esiste quindi un algoritmo efficiente per risolvere un programma intero.

L'insieme dei vincoli di un programma intero definisce un poliedro proprio come quello definito da un programma lineare

$$P = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$$

Si consideri A come l'insieme dei punti interi contenuti in P

$$A = \{\mathbf{x} \in \mathbb{Z}^n : \mathbf{x} \in P\} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k\}$$

è di *fondamentale importanza* l'involuppo convesso di A , che in tale contesto verrà indicato P_I , ossia l'insieme di tutte le combinazioni convesse di A

$$P_I = \left\{ \sum_{i=1}^k \alpha_i \mathbf{a}_i : \alpha_i \geq 0 \forall i, \sum_{i=1}^k \alpha_i = 1 \right\}$$

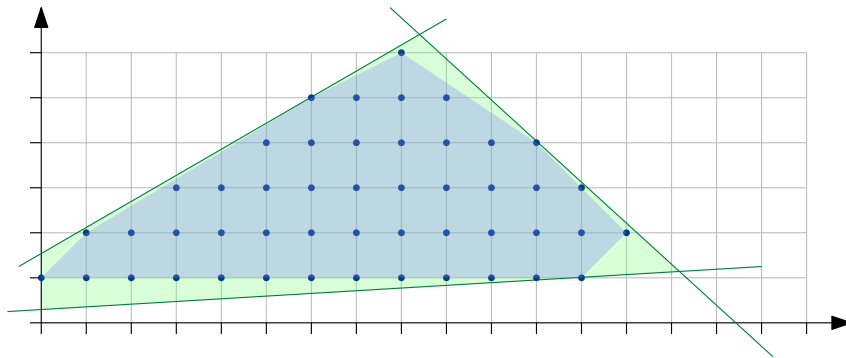


Figura 5.1: In verde è rappresentato il poliedro definito dai vincoli, in blu l'involuppo convesso dei suoi punti interi

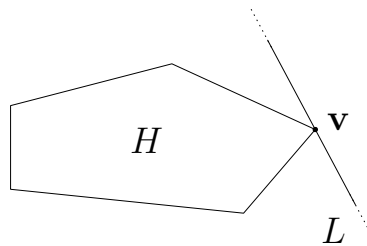
Proposizione 24 Sia $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ un'insieme di punti in \mathbb{R}^n , e sia H l'involuppo convesso di tali punti. Se \mathbf{v} è un vertice di H , allora $\mathbf{v} = \mathbf{u}_i$ per qualche i .

Dimostrazione: sia H l'involuppo convesso di $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$, e sia \mathbf{v} un vertice di H , esiste quindi un iper-piano

$$L = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{c}^T \mathbf{x} = \alpha\}$$

contenente \mathbf{v} tale per cui l'intero poliedro H è contenuto in una delle due metà definite da L

$$\begin{aligned} \mathbf{c}^T \mathbf{v} &= \alpha \\ \mathbf{c}^T \mathbf{x} &< \alpha, \forall \mathbf{x} \in H \setminus \{\mathbf{v}\} \end{aligned}$$



Si pone per assurdo che $\mathbf{u}_i \neq \mathbf{v}$, ossia che

$$\mathbf{v} = \sum_{i=0}^n \beta_i \mathbf{u}_i, \quad \sum_{i=1}^n \beta_i = 1$$

$$\exists j \text{ t.c. } 0 < \beta_j < 1$$

Sia j un fissato indice tale per cui $0 < \beta_j < 1$, si ha che

$$\mathbf{v} = \beta_j \mathbf{u}_j + \sum_{i \neq j} \beta_i \mathbf{u}_i$$

si consideri il seguente vettore

$$\sum_{i \neq j} \frac{\beta_i}{1 - \beta_j} \mathbf{u}_i$$

tale vettore è contenuto in H perché

$$\sum_{i \neq j} \beta_i = 1 - \beta_j \implies \sum_{i \neq j} \frac{\beta_i}{1 - \beta_j} = \frac{1}{1 - \beta_j} \sum_{i \neq j} \beta_i = 1$$

si riscrive

$$\mathbf{v} = \beta_j \mathbf{u}_j + (1 - \beta_j) \sum_{i \neq j} \frac{\beta_i}{1 - \beta_j} \mathbf{u}_i$$

quindi

$$\mathbf{c}^T \mathbf{v} = \mathbf{c}^T \left(\beta_j \mathbf{u}_j + (1 - \beta_j) \sum_{i \neq j} \frac{\beta_i}{1 - \beta_j} \mathbf{u}_i \right)$$

ricordando che $\mathbf{c}^T \mathbf{v} = \alpha$

$$\alpha = \mathbf{c}^T \left(\beta_j \mathbf{u}_j + (1 - \beta_j) \sum_{i \neq j} \frac{\beta_i}{1 - \beta_j} \mathbf{u}_i \right)$$

$$\alpha = \beta_j \mathbf{c}^T \mathbf{u}_j + (1 - \beta_j) \sum_{i \neq j} \frac{\beta_i}{1 - \beta_j} \mathbf{c}^T \mathbf{u}_i$$

essendo

$$\begin{aligned} \mathbf{c}^T \mathbf{u}_j &\leq \alpha \\ \mathbf{c}^T \mathbf{u}_i &\leq \alpha, \forall i \end{aligned}$$

si ha

$$\alpha \leq \beta_j \alpha + (1 - \beta_j) \sum_{i \neq j} \frac{\beta_i}{1 - \beta_j} \alpha$$

$$\alpha \leq \beta_j \alpha + (1 - \beta_j) \alpha = \alpha \implies$$

$$\alpha = \mathbf{c}^T \mathbf{v} = \beta_j \mathbf{c}^T \mathbf{u}_j + \mathbf{c}^T \left((1 - \beta_j) \sum_{i \neq j} \frac{\beta_i}{1 - \beta_j} \mathbf{u}_i \right) \leq \alpha$$

serve specificatamente che $\mathbf{c}^T \mathbf{u}_j = \alpha$, ma ciò è impossibile dato che solo il vertice \mathbf{v} soddisfa tale condizione, vi è quindi una contraddizione. ■

Tale proposizione ha delle implicazioni favorevoli, sia P un politopo, e P_I l'involuppo convesso dei punti interi di P , piuttosto che risolvere il programma intero

$$\begin{aligned} \max \mathbf{c}^T \mathbf{x} \\ \mathbf{x} \in P \\ \mathbf{x} \in \mathbb{Z}^n \end{aligned}$$

è possibile risolvere il programma lineare

$$\begin{aligned} \max \mathbf{c}^T \mathbf{x} \\ \mathbf{x} \in P_I \\ \mathbf{x} \in \mathbb{R}^n \end{aligned}$$

Dato che la soluzione sarà un vertice di P_I , che sappiamo essere un vettore di numeri interi. Tale strategia permette di risolvere un programma intero, l'involuppo convesso dei punti interi di un politopo è un'insieme generalmente più piccolo

$$\begin{aligned} P &= \{ \mathbf{x} : A\mathbf{x} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0} \} \\ P_I &= \left\{ \mathbf{x} : \begin{bmatrix} A \\ \bar{A} \end{bmatrix} \mathbf{x} \leq \begin{bmatrix} \mathbf{b} \\ \bar{\mathbf{b}} \end{bmatrix}, \mathbf{x} \geq \mathbf{0} \right\} \end{aligned}$$

L'involuppo convesso P_I è definito dai vincoli di P , con in aggiunta dei vincoli aggiuntivi $\bar{A}\mathbf{x} \leq \bar{\mathbf{b}}$. Dato un programma intero, com'è possibile trovare tali vincoli in modo da riformulare il problema come un programma lineare?



5.1.1 Il Teorema di Gomory

Si consideri il seguente programma intero

$$\begin{aligned} \max & 5x_1 + 8x_2 \\ & x_1 + x_2 \leq 6 \\ & 5x_1 + 9x_2 \leq 45 \\ & x_1, x_2 \geq 0, \quad x_1, x_2 \in \mathbb{Z} \end{aligned} \quad (5.2)$$

Si considera la sua versione rilassata in programma lineare (con aggiunta di variabili slack per renderlo in forma di equazione)

$$\begin{aligned} \max & 5x_1 + 8x_2 \\ & x_1 + x_2 + x_3 = 6 \\ & 5x_1 + 9x_2 + x_4 = 45 \\ & x_1, x_2, x_3, x_4 \geq 0 \end{aligned} \quad (5.3)$$

Si vuole risolvere il programma intero, si comincia risolvendo il programma lineare con il metodo del simplesso, arrivando alla base $\mathcal{B} = \{1, 2\}$ che fornisce il seguente tableau

$x_1 = 9/4 - 9/4x_3 - 1/4x_4$
$x_2 = 15/4 + 5/4x_3 - 1/4x_4$
$z = 41/4 - 5/4x_3 - 3/4x_4$

La BFS ottimale trovata è un vettore contenente numeri razionali, quindi non valida come soluzione del programma intero, si noti però che la seconda equazione del tableau, ossia

$$x_2 = 15/4 + 5/4x_3 - 1/4x_4$$

esprime un vincolo che deve essere soddisfatto da ogni soluzione ammissibile, anche quelle intere:

$$\begin{aligned} x_2 &= \frac{15}{4} + \frac{5}{4}x_3 - \frac{1}{4}x_4 \implies \\ x_2 - \frac{5}{4}x_3 + \frac{1}{4}x_4 &= \frac{15}{4} \implies \\ (1+0)x_2 + (-2+3/4)x_3 + (0+1/4)x_4 &= 3+3/4 \end{aligned}$$

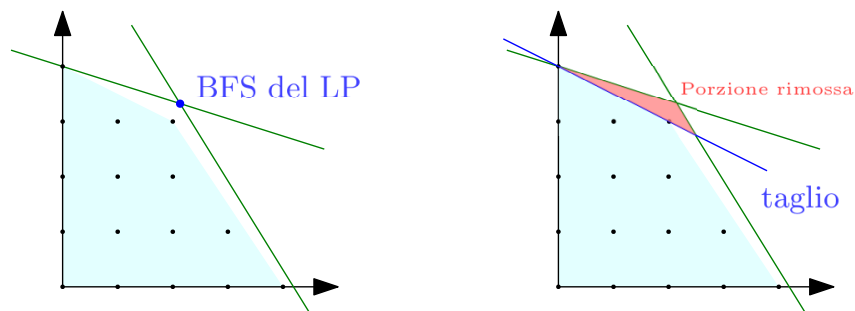
Il vincolo è stato espresso dividendo parti decimali ed intere dei coefficienti, se si assume che la soluzione è un vettore di numeri razionali, si ottiene il seguente vincolo, per qualche $k \in \mathbb{Z}^+$

$$0x_2 + \frac{3}{4}x_3 + \frac{1}{4}x_4 = k + \frac{3}{4}$$

Che si può riscrivere come

$$\begin{aligned} \frac{3}{4}x_3 + \frac{1}{4}x_4 &\geq \frac{3}{4} \\ -\frac{3}{4}x_3 - \frac{1}{4}x_4 &\leq -\frac{3}{4} \end{aligned}$$

Tale vincolo deve essere soddisfatto da ogni soluzione intera, si può quindi aggiungere al programma lineare, **tagliando** una porzione di politopo di soluzioni non intere.



5.2 Il Matching Perfetto di Peso Minimo

Questa sezione presenta un'applicazione della programmazione intera ad un noto problema sui grafi.

Definizione 30 Sia $G = (V, E)$ un grafo non diretto, un sotto-insieme $M \subseteq E(G)$ è detto un **matching** se tutti gli archi di M non hanno vertici in comune, ossia

$$\forall (u, v), (x, y) \in M, \quad \begin{array}{l} u \neq x \\ v \neq y \end{array}$$

un matching è detto **perfetto** se i suoi archi toccano tutti i vertici del grafo.

Dato un grafo G con pesi w sugli archi, si vuole trovare il matching perfetto (da ora in poi, denominato PM) di peso minimo, ossia, la cui somma dei pesi degli archi del matching sia minimale. Il problema deve essere modellato nei termini della programmazione intera, le variabili del problema rappresentano gli archi del grafo

$$\mathbf{x} = [x_1, x_2, \dots, x_{|E(G)|}]$$

vi è una variabile x_e per ogni arco $e \in E(G)$, e

$$x_e = \begin{cases} 1 & \text{se } e \text{ è nel PM} \\ 0 & \text{se } e \text{ non è nel PM} \end{cases}$$

La funzione obiettivo sarà $\mathbf{w}^T \mathbf{x}$, dove \mathbf{w} è il vettore dei pesi sugli archi. Bisogna ora formalizzare i vincoli, si consideri la funzione $\delta : V(G) \rightarrow \mathcal{P}(E(G))$ definita come segue

$$\delta(v) = \left\{ e \text{ tale che } \begin{array}{l} e = (v, x) \text{ per qualche } x \in V(G) \\ e = (x, v) \text{ per qualche } x \in V(G) \end{array} \vee \right\}$$

$\delta(v)$ è quindi l'insieme di tutti gli archi in G che coinvolgono il vertice v . In questo modo si può modellare il problema come segue

$$\begin{aligned} \min \quad & \mathbf{w}^T \mathbf{x} \\ \sum_{e \in \delta(v)} x_e &= 1 \quad \forall v \in V(G) \\ x_i &\in \{0, 1\} \quad \forall i \end{aligned} \tag{5.4}$$

Il vincolo $\sum_{e \in \delta(v)} x_e = 1$ impone che il nodo v deve essere toccato da un solo arco. Dato un grafo G , definiamo **vettore indicatore** un vettore \mathbf{x} che descrive un matching per il grafo (le cui componenti sono quindi in $\{0, 1\}$), chiaramente, i vertici del politopo del problema 5.4 sono vettori indicatori.

Definizione 31 il **politopo del matching perfetto** P_M è l'involuppo convesso dei vettori indicatori.

Il seguente teorema propone un risultato che permette di risolvere il problema del PM di peso minimo tramite la risoluzione di un programma lineare, rilassando i vincoli del programma intero 5.4, con l'assunzione che il grafo considerato sia *bipartito*.

5.2.1 PM per i Grafi Bipartiti

Proposizione 25 Un grafo è bipartito se e solo se non ha cicli di lunghezza (numero di vertici coinvolti) dispari.

Teorema 11 Sia G un grafo bipartito, il politopo del matching perfetto è il seguente insieme

$$P_M = \left\{ \mathbf{x} \in \mathbb{R}^{|E(G)|} : x_e \geq 0 \quad \forall e \in E(G), \quad \sum_{e \in \delta(v)} x_e = 1 \quad \forall v \in V(G) \right\}$$

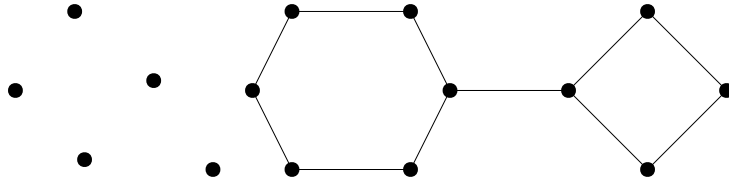
Praticamente, se il grafo è bipartito, non è necessario imporre il vincolo che $x_e \in \{0, 1\}$, i vertici di tale politopo saranno automaticamente vettori interi le cui componenti sono comprese in $\{0, 1\}$.

Dimostrazione: Sia $\bar{\mathbf{x}}$ un vertice di P_M , si pone per assurdo che, per qualche e , $\bar{x}_e \in (0, 1)$. Si definisce il **grafo di supporto** il grafo G_u tale per cui

$$\begin{aligned} V(G_u) &= V(G) \\ E(G_u) &= \{e \in E(G) : 0 < \bar{x}_e < 1\} \end{aligned}$$

Claim non esistono vertici in G_u di grado uguale ad 1. La dimostrazione è omessa.

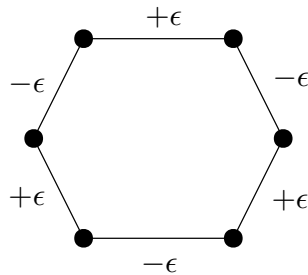
Quindi i vertici di G_u o sono isolati, oppure hanno grado almeno 2.



Si consideri una componente di G_u che ha almeno un'arco, essendo che tutti i vertici di tale componente hanno grado almeno due, in tale componente vi è un ciclo C , e questo è di lunghezza pari per la proposizione 25. Si consideri il seguente valore

$$\epsilon = \min_{e \in C} (\bar{x}_e, 1 - \bar{x}_e)$$

Si consideri il vettore \bar{x}^1 , ottenuto da \bar{x} , sommando e sottraendo in maniera alternata il valore ϵ dalle componenti relative agli archi del ciclo C .



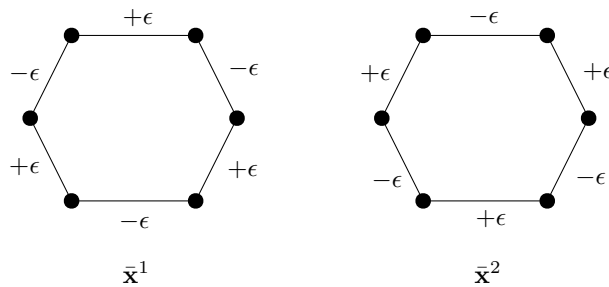
data la scelta di ϵ , si avrà che ogni $\bar{x}_e^1 > 0$ per ogni e , inoltre, per ogni vertice $v \notin C$

$$\sum_{e \in \delta(v)} \bar{x}_e^1 = 1$$

e per ogni vertice $v \in C$

$$\sum_{e \in \delta(v)} \bar{x}_e^1 = \sum_{e \in \delta(v)} \bar{x}_e$$

ciò è verificato dato che per ogni arco in cui si somma ϵ in C , ve ne è uno dalla quale si sottrae. In tal modo si è mostrato che $\bar{x}^1 \in P_M$. Si consideri poi un vettore \bar{x}^2 ottenuto in maniera analoga al precedente ma sommando e sottraendo ϵ dalle componenti degli archi in C in maniera opposta.



per gli analoghi motivi, $\bar{x}^2 \in P_M$. Si noti come

$$\bar{x} = \bar{x}^1 + \bar{x}^2$$

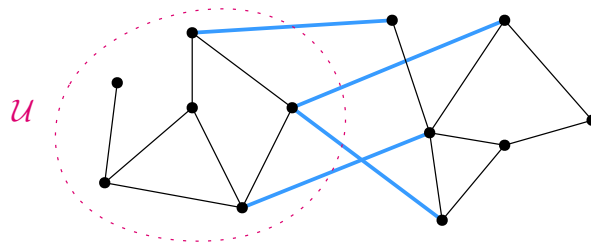
si era assunto che \bar{x} fosse un vertice, ciò porta ad una contraddizione e completa la dimostrazione. ■

Tale teorema dimostra che è sufficiente risolvere un programma lineare per trovare il PM di peso minimo per un grafo bipartito.

5.2.2 PM per i Grafi non Bipartiti

Se un grafo non è bipartito è comunque possibile definire analiticamente il politopo del matching perfetto, ma con alcune restrizioni.

Nel corso della dimostrazione, verrà denotato \mathcal{U} un qualsiasi sotto-insieme dei nodi del grafo, contenente un numero dispari di elementi, si definisce poi un'estensione della funzione δ precedentemente definita, se v è un vertice, $\delta(v)$ è l'insieme di tutti gli archi in G che coinvolgono il vertice v , se invece \mathcal{U} è un insieme di vertici, $\delta(\mathcal{U})$ è l'insieme degli archi che connettono i vertici in \mathcal{U} con i vertici in $V(G) \setminus \mathcal{U}$.



gli archi evidenziati sono quelli contenuti in $\delta(\mathcal{U})$

Teorema 12 (Edmonds) Sia $G = (V, E)$ un grafo, l'insieme

$$P^* = \left\{ x \in \mathbb{R}^{|E(G)|} : \begin{array}{l} \sum_{e \in \delta(v)} x_e = 1 \quad \forall v \in V(G), \quad x_e \geq 0, \\ \sum_{e \in \delta(\mathcal{U})} x_e \geq 1 \\ \forall \mathcal{U} \subseteq V(G), \quad |\mathcal{U}| \text{ dispari} \end{array} \right\} \quad (5.5)$$

è il politopo del matching perfetto di G .

Dimostrazione: Si vuole dimostrare che i vertici di P^* sono vettori indicatori del matching perfetto. Si assume che il teorema sia falso, sia G un contro-esempio per il teorema con il minimo numero possibili di vertici, l'assunzione è che esiste un vertice \bar{x} di P^* che ha una componente diversa da 0 o 1. Si consideri l'insieme F degli archi relativi a tali componenti

$$F = \{e \in E(G) : 0 < \bar{x}_e < 1\}$$

Sia G_F il grafo con gli stessi vertici di G , avente solo gli archi in F .

Caso 1: Per ogni sotto-insieme $\mathcal{U} \subseteq V(G)$ con un numero dispari di elementi, si ha che

$$\sum_{e \in \delta(\mathcal{U})} \bar{x}_e > 1$$

sia $n = |V(G)|$ il numero dei vertici, si considera il seguente fattore

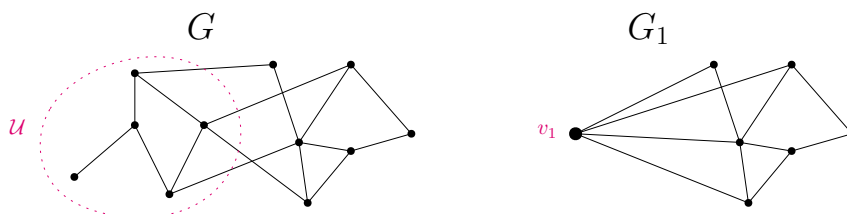
$$\varepsilon = \min_{\substack{e \in F \\ \mathcal{U} \subseteq V(G) \\ |\mathcal{U}| \text{ dispari}}} \left(\frac{\bar{x}_e}{2}, \left(\frac{1}{n} \sum_{e \in \delta(\mathcal{U})} \bar{x}_e \right) - \frac{1}{n} \right) \quad (5.6)$$

Se v è un vertice in G_F e non è isolato, allora il suo grado è almeno 2. G_F non ha cicli di lunghezza pari (dimostrazione omessa). G_F non ha cicli di lunghezza dispari tali per cui, ogni vertice del ciclo ha grado 2. **TODO continuare**

Caso 2: Esiste $\mathcal{U} \subseteq V(G)$ con un numero dispari di elementi e $|\mathcal{U}| \geq 3$ tale che

$$\sum_{e \in \delta(\mathcal{U})} \bar{x}_e = 1$$

Sia fisso tale sotto-insieme \mathcal{U} , si considera il grafo G_1 ottenuto comprimendo i vertici di \mathcal{U} in un'unico vertice denotato v_1 .



Sia \bar{x}_1 una restrizione di \bar{x} sui soli archi di G_1