

## Capitolo 6

# Programmazione lineare: proprietà generali

La programmazione lineare è indubbiamente l'argomento centrale dell'ottimizzazione. Fra i vari modelli di ricerca operativa la programmazione lineare è quello che è stato più ampiamente utilizzato. Non solo si applica a numerosi problemi reali che hanno di per sé una struttura lineare, ma è anche un indispensabile strumento di supporto, tecnico e concettuale, per modelli più complessi di tipo discreto, quali quelli di ottimizzazione combinatoria oppure di programmazione lineare intera. Se si considera che virtualmente quasi ogni problema che coinvolge decisioni di natura discreta può essere modellato con la programmazione lineare intera e questa, nella maggior parte dei casi, invoca la programmazione lineare per pervenire alla soluzione, ci si può render conto dell'importanza pratica e teorica della programmazione lineare.

Come si vedrà in questo capitolo l'insieme ammissibile di un problema di programmazione lineare è un poliedro, cioè un insieme convesso formato da un continuo di punti. Inoltre, siccome la funzione obiettivo è lineare, avviene che gli ottimi cadano sui vertici del poliedro, che sono invece in numero finito e costituiscono una struttura discreta di punti (vertici e spigoli hanno la struttura di grafo). È proprio questa duplice natura continua e discreta che rende la programmazione lineare così importante.

Anche se il padre della programmazione lineare viene comunemente (e giustamente) considerato George Dantzig che per primo ne ideò nel 1947 un algoritmo risolutivo, il cosiddetto *metodo del semplice* (si veda ad esempio Dantzig [1951] e la monografia Dantzig [1963]), tuttavia alcuni fra i concetti fondamentali della programmazione lineare datano molto più indietro nel tempo. Già Fourier [1827] aveva studiato come trovare soluzioni ammissibili di un sistema di disequazioni lineari e aveva fornito un metodo risolutivo che ha delle analogie con il metodo del semplice (si veda anche Kohler [1973]). Molti matematici studiarono le proprietà dei poliedri nel diciannovesimo secolo, in particolare la polarità e la dualità fra punti e piani. Il lemma di Farkas [1895] (vedi il teorema 4.76), antesignano delle proprietà di dualità, fu preceduto e seguito storicamente da simili ed indipendenti risultati (Gordan [1873], Minkowski [1896]). Un metodo di calcolo destinato a minimizzare gli errori d'osservazione con norma infinita dovuto a de la Vallée Poussin [1910] presenta lati simili al metodo del semplice. Infine i lavori più importanti che possono essere considerati i precursori delle ricerche di Dantzig sono quelli di von Neumann [1928,1937] in teoria dei giochi e modelli economici. Di fatto la teoria della dualità lineare è già contenuta in questi lavori. Inoltre una sorprendente anticipazione di molti risultati di programmazione lineare, incluso un metodo di calcolo, è contenuta nel lavoro del matematico russo Kantorovich [1939], ignoto per molto tempo ai ricercatori occidentali.

Tuttavia fu solo dopo i risultati di Dantzig che l'uso della programmazione lineare (il termine stesso fu coniato allora) divenne di dominio universale e segnò l'inizio di una serie

straordinaria di sviluppi matematici e di risultati algoritmici, che diedero corpo ad una nuova disciplina, la ricerca operativa. Va naturalmente sottolineato che questo sviluppo fu possibile grazie all'avvento dei calcolatori nati più o meno negli stessi anni della ricerca operativa.

Da allora, oltre al normale sviluppo e consolidamento della disciplina, la programmazione lineare ha vissuto anche due momenti particolarmente significativi. Khacian [1979] provò che la programmazione lineare è polinomiale e Karmarkar [1984] inventò un algoritmo polinomiale di natura completamente diversa dal metodo del simplesso, con garanzie di efficienza pratica tali da essergli competitivo. In questi ultimi anni l'accresciuta potenza di calcolo e il miglioramento degli algoritmi ha fatto sì che si possano risolvere problemi con decine di migliaia di vincoli e centinaia di migliaia di variabili.

## 6.1. Definizioni

Un problema di programmazione lineare è un qualsiasi problema di ottimizzazione in cui la funzione obiettivo sia lineare e l'insieme ammissibile sia assegnato tramite un sistema di disequazioni e/o equazioni affini. Molto frequentemente parte delle disequazioni è costituita da semplici vincoli di non negatività sulle variabili. Quindi in generale possiamo scrivere un problema di programmazione lineare come:

$$\begin{aligned} \min \quad & c_1 x_1 + c_2 x_2 \\ & A_{11} x_1 + A_{12} x_2 \geq b_1 \\ & A_{21} x_1 + A_{22} x_2 = b_2 \\ & x_1 \geq 0 \end{aligned}$$

Tuttavia ogni problema di programmazione lineare può essere ricondotto ad una delle due seguenti forme:

6.1 DEFINIZIONE. Forma canonica di un problema di programmazione lineare:

$$\begin{array}{ll} \min & cx \\ & Ax \geq b \\ & x \geq 0 \end{array} \quad \text{oppure} \quad \begin{array}{ll} \max & cx \\ & Ax \leq b \\ & x \geq 0 \end{array}$$

6.2 DEFINIZIONE. Forma standard di un problema di programmazione lineare:

$$\begin{aligned} \min \quad & cx \\ & Ax = b \\ & x \geq 0 \end{aligned}$$

La matrice  $A$  viene detta *matrice dei vincoli*. Per convenzione sia  $n$  il numero di colonne di  $A$  e  $m$  il numero delle righe. Per trasformare in forma canonica o standard un problema dato in forma generale, basta eseguire le seguenti operazioni:

a) se non è presente il vincolo  $x_i \geq 0$  si introducono due variabili  $x_i^+ \geq 0$  e  $x_i^- \geq 0$  al posto di  $x_i$  tali che

$$x_i = x_i^+ - x_i^-$$

b) un vincolo di disequaglianza

$$A_i x \geq b_i$$

si trasforma in un vincolo d'uguaglianza introducendo una variabile  $s_i \geq 0$  detta variabile di *scarto* o *ausiliaria* (in inglese *slack*) tale che

$$A_i x - s_i = b_i$$

c) un vincolo d'uguaglianza

$$A_i x = b_i$$

viene trasformato nei due vincoli di disequaglianza

$$\begin{aligned} A_i x &\geq b_i \\ -A_i x &\geq -b_i \end{aligned}$$

6.3 ESEMPIO. Il seguente problema

$$\begin{aligned} \min \quad & x_1 + 2x_2 - x_3 \\ & -x_1 + x_2 = 5 \\ & 2x_1 - x_2 - x_3 \leq 3 \\ & x_2 \geq 0 \quad x_3 \geq 0 \end{aligned}$$

viene trasformato nella seguente forma canonica

$$\begin{aligned} \min \quad & y_1 + 2y_2 - y_3 - y_4 \\ & -y_1 + y_2 + y_4 \geq 5 \\ & y_1 - y_2 - y_4 \geq -5 \\ & -2y_1 + y_2 + y_3 + 2y_4 \geq -3 \\ & y_1 \geq 0 \quad y_2 \geq 0 \quad y_3 \geq 0 \quad y_4 \geq 0 \end{aligned}$$

(dove  $x_1 = y_1 - y_4$ ,  $x_2 = y_2$ ,  $x_3 = y_3$ ) oppure nella seguente forma standard (medesimo legame fra  $x$  e  $y$ )

$$\begin{aligned} \min \quad & y_1 + 2y_2 - y_3 - y_4 \\ & -y_1 + y_2 + y_4 = 5 \\ & 2y_1 - y_2 - y_3 - 2y_4 + y_5 = 3 \\ & y_1 \geq 0 \quad y_2 \geq 0 \quad y_3 \geq 0 \quad y_4 \geq 0 \quad y_5 \geq 0 \end{aligned} \quad \blacksquare$$

L'insieme ammissibile di un problema di programmazione lineare è necessariamente un poliedro, dato che è determinato da un'intersezione finita di semispazi. Se il problema è in forma canonica ed esiste un  $\bar{x}$  tale che  $\bar{x} > 0$  e  $A\bar{x} > b$ , allora la dimensione del poliedro è  $n$ , cioè si tratta di un corpo convesso. Se invece il problema è in forma standard, la presenza di  $m$  equazioni abbassa la dimensione che non può essere più grande di  $n - m$  (è uguale a  $n - m$  se esiste  $\bar{x} > 0$  tale che  $A\bar{x} = b$ ).

Ci si può chiedere a questo punto se un problema di programmazione lineare potrebbe essere semplicemente definito come un problema caratterizzato da una funzione obiettivo lineare su un poliedro. Effettivamente, almeno per ciò che riguarda le proprietà della programmazione lineare, il fatto saliente è che  $F$  sia un poliedro e  $f$  sia lineare, poco importa come sia assegnato il poliedro. Non si può dire altrettanto per ciò che riguarda l'algoritmo

risolutivo. Infatti è il tipo di informazione che definisce il poliedro che dà luogo a problemi qualitativamente diversi dal punto di vista algoritmico.

Come si vedrà fra poco, il minimo di una funzione lineare su un poliedro  $F$  può essere cercato unicamente tra i vertici di  $F$ . Quindi se  $F$  è assegnato come lista di vertici la risoluzione si effettua in modo esaustivo con complessità lineare. Se invece  $F$  è assegnato tramite disequazioni (cioè si tratta di un problema di programmazione lineare) il problema ha complessità polinomiale, come vedremo in seguito. Ben più critico è il caso in cui, ad esempio  $F = \text{conv}\{x : Ax \geq b, x \geq 0, x \text{ intero}\}$  (e si tratta quindi di un problema di programmazione lineare intera) perché allora il problema diventa **NP**-difficile.

## 6.2. Geometria della programmazione lineare

Sia dato

$$v := \inf_{x \in K \subset \mathbb{R}^n} f(x)$$

con  $f(x) := cx$ , funzionale lineare e  $K$  insieme qualsiasi.

6.4 LEMMA. *Se  $f$  non è costante su  $K$  ed esiste ottimo  $\hat{x}$ , allora  $\hat{x} \in \partial K$ .*

DIMOSTRAZIONE. Sia  $x \in \overset{\circ}{K}$ . Siccome  $f$  non è costante su  $K$ , esiste  $y \in K$  tale che  $cy \neq cx$ . Se  $cy < cx$ ,  $x$  non può essere ottimo. Se invece  $cy > cx$  allora esiste  $\varepsilon > 0$  tale che  $x - \varepsilon(y - x) \in K$ , dato che  $y - x \in S$  con  $S$  sottospazio generatore di  $\text{aff } K$  e  $x \in \overset{\circ}{K}$ . Allora

$$c(x - \varepsilon(y - x)) = cx - \varepsilon(cy - cx) < cx$$

e  $x$  non può essere ottimo. ■

Quindi una prima proprietà delle funzioni obiettivo lineari è che, a meno di non essere costanti, trovano gli ottimi soltanto sulla frontiera dell'insieme ammissibile. Se  $K$  è un insieme convesso o, in particolare, un poliedro, si può dire ancora di più.

6.5 TEOREMA. *Sia  $K$  convesso e chiuso. Se esistono punti estremi di  $K$  ed esistono ottimi, allora almeno un ottimo è punto estremo di  $K$ .*

DIMOSTRAZIONE. Gli ottimi appartengono al piano di equazione  $cx = v$ . Sia  $H$  l'intersezione di  $K$  con il piano.  $H$  è chiuso e convesso ed è una faccia di  $K$ : infatti se  $cx$  è costante su  $K$  si ha  $H = K$ ; altrimenti in base al precedente lemma  $H \subset \partial K$  e quindi, in base al teorema 4.53,  $H$  è faccia propria di  $K$ . Siccome esistono punti estremi di  $K$ ,  $K$  non contiene rette. Quindi nemmeno  $H$  contiene rette, ovvero esistono punti estremi di  $H$ , ma questi sono anche punti estremi di  $K$ . ■

Se in particolare  $K$  è un poliedro, come nella programmazione lineare, abbiamo il seguente ovvio corollario.

6.6 COROLLARIO. *Sia  $P$  un poliedro. Se esistono vertici di  $P$  ed esistono ottimi, allora almeno un ottimo è vertice di  $P$ .* ■

Inoltre per i problemi di programmazione lineare dati in forma canonica o standard, si ha:

6.7 COROLLARIO. Sia  $P \neq \emptyset$  il poliedro ammissibile di un problema di programmazione lineare assegnato in forma canonica o standard. Se esistono ottimi (o alternativamente il problema non è illimitato) almeno uno di essi è vertice di  $P$ .

DIMOSTRAZIONE. Basta notare che  $P$  non può contenere rette a causa della presenza dei vincoli  $x \geq 0$ . ■

L'importanza di questi teoremi risiede nel fatto che se  $P$  è un poliedro, i problemi

$$\min_{x \in P} c x \quad \text{e} \quad \min_{x \in \text{ext } P} c x$$

sono equivalenti pur essendo il primo un problema continuo e il secondo un problema discreto. In effetti la risoluzione di un problema di programmazione lineare può essere fatta sia pensando al problema continuo (algoritmo dell'ellissoide, algoritmi ai punti interni) sia pensando a quello discreto (metodo del simplesso).

La ricerca dell'ottimo fra i vertici di  $P$  viene facilitata dalle seguenti considerazioni: sia  $x$  vertice di  $P$  e si definisca intorno di  $x$  l'insieme di vertici

$$N_x := \{y \in \text{ext } P : \text{conv}(x, y) \text{ è spigolo di } P\} \quad (6.1)$$

Dimosteremo ora che  $N_x$  è un intorno esatto e che quindi l'ottimo può essere raggiunto in un numero finito di passi da un algoritmo di ricerca locale. Serve dapprima dimostrare il seguente:

6.8 LEMMA. Siano  $K_1$  e  $K_2$  due insiemi chiusi e convessi tali che  $K_2 \subset K_1$ . Allora se  $G$  è una faccia di  $K_1$ ,  $G \cap K_2$  è una faccia di  $K_2$  di dimensione non superiore a quella di  $G$ .

DIMOSTRAZIONE. Sia  $x \in G \cap K_2$ ,  $z, y \in K_2$  e  $x$  combinazione convessa stretta di  $z$  e  $y$ . Da  $K_2 \subset K_1$  abbiamo  $z, y \in K_1$  e siccome  $x \in G$  e  $G$  è una faccia di  $K_1$ , deve valere  $z, y \in G$ , cioè  $z, y \in G \cap K_2$ . La tesi sulla dimensione è ovvia. ■

Vale allora il seguente teorema:

6.9 TEOREMA. Sia  $P$  un poliedro,  $f$  lineare. Se  $f$  è limitata inferiormente su  $P$  e

$$f(\hat{x}) \leq f(y) \quad \forall y \in N_{\hat{x}} \quad \text{con } \hat{x} \in \text{ext } P$$

allora  $\hat{x}$  è ottimo (alternativamente, se esiste ottimo di  $\inf_{x \in P} f(x)$ , allora  $N_{\hat{x}}$  è un intorno esatto).

DIMOSTRAZIONE. Siccome  $P$  è un poliedro,  $P = H + C$  dove  $H$  è un poliedro limitato e  $C$  è un cono poliedrale. In particolare

$$H = \left\{ \sum_{x^k \in \text{ext } P} \alpha_k x^k : \alpha_k \geq 0, \sum \alpha_k = 1 \right\}$$

$$C = \left\{ \sum_{d^k \in \text{dext } P} \beta_k d^k : \beta_k \geq 0 \right\}$$

Sia

$$D_{\hat{x}} = \left\{ \sum_{x^k \in \text{ext } P} \gamma_k (x^k - \hat{x}) : \gamma_k \geq 0 \right\}$$

dove  $\hat{x} \in \text{ext } P$ . Si noti che  $D_{\hat{x}}$  è un cono. È immediato che  $P \subset \hat{x} + D_{\hat{x}} + C$  (basta prendere  $\sum \gamma_k = 1$ ). Ogni  $y \in D_{\hat{x}} + C$  può essere espresso come

$$y = \sum_{x^k \in N_{\hat{x}}} \gamma_k (x^k - \hat{x}) + \sum_{d^k \in \text{dext } P} \beta_k d^k$$

poiché se  $x^k \notin N_{\hat{x}}$ ,  $\text{conv}(x^k, \hat{x})$  non è spigolo di  $P$  (e neppure vertice) per definizione di intorno e dal lemma discende che  $\hat{x} + \text{cone}(x^k - \hat{x})$  non è spigolo di  $\hat{x} + D_{\hat{x}} + C$  e quindi  $\text{cone}(x^k - \hat{x}) \not\subset \text{dext } D_{\hat{x}}$ .

Siccome  $f$  è limitata inferiormente su  $P$ ,  $f(d^k) \geq 0$ . Quindi  $\forall y \in \hat{x} + D_{\hat{x}} + C$  si ha

$$\begin{aligned} f(y) &= f(\hat{x}) + \sum_{x^k \in N_{\hat{x}}} \gamma_k (f(x^k) - f(\hat{x})) + \sum \beta_k f(d^k) \\ &\geq f(\hat{x}) + \sum_{x^k \in N_{\hat{x}}} \gamma_k (f(x^k) - f(\hat{x})) \quad (\text{perché } f(d^k) \geq 0) \\ &\geq f(\hat{x}) \quad (\text{perché } \hat{x} \text{ è ottimo locale discreto}) \end{aligned}$$

e siccome  $P \subset \hat{x} + D_{\hat{x}} + C$  si ha  $f(x) \geq f(\hat{x})$ ,  $\forall x \in P$ . ■

### 6.3. Dualità

Si consideri un problema di programmazione lineare posto nella seguente forma

$$\begin{aligned} v &= \min c x \\ A x &\geq b \\ x &\in X \quad (X \text{ poliedro}) \end{aligned}$$

e il relativo problema perturbato

$$\begin{aligned} v(y) &= \min c x \\ A x &\geq b - y \\ x &\in X \end{aligned}$$

Dimostriamo ora la seguente fondamentale proprietà della programmazione lineare:

**6.10 TEOREMA.** *In un problema di programmazione lineare che ammetta ottimo, vale sempre la dualità forte.*

**DIMOSTRAZIONE.** Data la linearità delle funzioni e la poliedralità di  $X$ , epi  $v$  è necessariamente un poliedro, con un numero finito di piani tangenti. Sono allora in numero finito anche i piani tangenti non verticali e quindi i valori di  $u$  per cui  $y_0 + u y = L(u)$  è piano tangente non verticale a epi  $v$ . Sia  $\hat{U}$  questo insieme finito. Quindi in base al corollario 4.106 e al teorema 5.31 si ha, per  $y \in \text{dom } v$

$$v(y) = \sup_{u \in \hat{U}} L(u) - u y = \max_{u \in \hat{U}} L(u) - u y$$

e in particolare

$$v(0) = \max_{u \in \hat{U}} L(u)$$

siccome inoltre  $\hat{U} \subset U \cap \mathbb{R}_+^m$  e  $v(0) \geq L(u)$ ,  $\forall u \in U$ ,  $u \geq 0$ , si ha anche

$$v(0) = \max_{u \in U} L(u) = d. \quad \blacksquare$$

Ci si può chiedere se lo scarto di dualità sia nullo, in senso esteso, anche quando un problema non ammetta ottimo, avendo cioè  $v = d = -\infty$  oppure  $v = d = +\infty$ . Ciò non è sempre vero. Infatti mentre  $v = -\infty$  implica  $d = -\infty$  e  $d = +\infty$  implica  $v = +\infty$  per la semplice proprietà di dualità debole, può accadere che entrambi i problemi primale e duale siano non ammissibili e quindi  $v = +\infty$ ,  $d = -\infty$  con scarto di dualità addirittura infinito! (si prenda ad esempio un problema in forma canonica con  $A = 0$ ,  $b = 1$ ,  $c = -1$ ).

È utile avvertire che non è del tutto corretto dire 'il' problema duale, in quanto il problema duale non è unico, dipendendo dalla particolare divisione dei vincoli in impliciti ed espliciti. Tuttavia il problema duale di maggior interesse è quello che si ottiene considerando il vincolo di non negatività sulle variabili come vincolo implicito, e quindi è a questo problema duale che ci si riferisce abitualmente. Si tenga comunque presente che tutti i problemi duali di un dato problema di programmazione lineare, valendo la dualità forte, sono equivalenti nel senso del teorema 5.42.

Nel caso della programmazione lineare il problema duale si determina in modo abbastanza semplice e, cosa notevole, è ancora un problema di programmazione lineare. Si consideri il problema in forma canonica:

$$\begin{aligned} \min \quad & c x \\ & A x \geq b \\ & x \geq 0 \end{aligned}$$

Allora

$$L(u) = \inf_{x \geq 0} c x + u(b - A x)$$

cioè

$$L(u) = u b + \inf_{x \geq 0} (c - u A) x$$

$L(u) > -\infty$  se e solo se  $c - u A \geq 0$ . Quindi

$$U = \{u : u A \leq c\}$$

e, se  $u \in U$ , si ha  $\inf_{x \geq 0} (c - u A) x = 0$ . Quindi

$$\begin{array}{ll} \sup_{u \in U} L(u) & \implies \max u b \\ u \in U & \implies u A \leq c \\ u \geq 0 & \implies u \geq 0 \end{array}$$

cioè ancora un problema di programmazione lineare in forma canonica.

Data la particolare simmetria fra problema duale e primale nella programmazione lineare, è possibile formulare condizioni di ottimalità in termini di semplice ammissibilità. Ad esempio  $(\hat{x}, \hat{u})$  sono ottimi se e solo se sono soluzioni ammissibili di

$$\begin{pmatrix} A & \mathbf{0} \\ \mathbf{0} & -A^T \\ -c & b^T \end{pmatrix} \begin{pmatrix} x \\ u^T \end{pmatrix} \geq \begin{pmatrix} b \\ -c^T \\ 0 \end{pmatrix}$$

$$x \geq 0 \quad u \geq 0$$

6.11 ESERCIZIO. Costruire un problema di programmazione lineare in forma canonica per il quale  $\hat{x}_j := 1, j := 1, \dots, 4$ , è ottimo primale e  $\hat{u}_i := 1, i := 1, \dots, 4$ , è ottimo duale. Indicare un metodo generale per ottenere un problema di programmazione lineare con soluzioni primali e duali assegnate. ■

Se il problema di programmazione lineare non è in forma canonica, non è necessario trasformarlo in forma canonica per ottenerne il duale. Infatti se il problema è il seguente

$$\begin{aligned} \min \quad & c^1 x^1 + c^2 x^2 \\ & A_{11} x^1 + A_{12} x^2 \geq b^1 \\ & A_{21} x^1 + A_{22} x^2 = b^2 \\ & x^1 \geq 0 \end{aligned}$$

si ottiene il seguente problema duale

$$\begin{aligned} \max \quad & u^1 b^1 + u^2 b^2 \\ & u^1 A_{11} + u^2 A_{21} \leq c^1 \\ & u^1 A_{12} + u^2 A_{22} = c^2 \\ & u^1 \geq 0 \end{aligned}$$

Come si vede la regola pratica per costruire il duale consiste in:

- se il primale è di tipo ‘min’ il duale è di tipo ‘max’ e viceversa;
- se il primale è di tipo ‘min’, trasformare eventuali vincoli di  $\leq$  in  $\geq$  (il contrario se il primale è di tipo ‘max’);
- ad ogni diseuguaglianza primale associare una variabile duale non negativa; ad ogni uguaglianza primale associare una variabile duale svincolata;
- ad ogni variabile primale non negativa associare un vincolo duale di diseuguaglianza  $\leq$  se il primale è di tipo ‘min’ ( $\geq$  se il primale è di tipo ‘max’) ottenuto usando i coefficienti della colonna rispettiva della matrice dei vincoli; ad ogni variabile primale svincolata associare un vincolo duale di uguaglianza ottenuto usando i coefficienti della colonna rispettiva;
- spostare i coefficienti di destra delle diseuguaglianze e/o uguaglianze primali nell’obiettivo duale e spostare i coefficienti dell’obiettivo primale nei coefficienti di destra delle diseuguaglianze e/o uguaglianze duali.

Spesso si deve costruire il duale di un problema definito tramite complicate relazioni di indici. In questi casi non è sempre immediato applicare le regole appena esposte. C’è tuttavia una tecnica semplice ma efficace che permette in questi casi di costruire il duale senza difficoltà. Questa tecnica fa uso di uno speciale operatore, inizialmente introdotto nel linguaggio APL, che, come suggerito in Graham, Knuth e Patashnik [1994] p. 24, si rivela di grande utilità in problemi di matematica discreta. L’operatore è definito come

$$[P] = \begin{cases} 0 & \text{se } P \text{ è falso} \\ 1 & \text{se } P \text{ è vero} \end{cases}$$

Ecco un esempio della sua efficacia in un’operazione di inversione di sommatorie:

$$\sum_{i=1}^n \sum_{j=1}^i a_{ij} = \sum_{i=1}^n \sum_{j=1}^n [j \leq i] a_{ij} = \sum_{j=1}^n \sum_{i=1}^n [j \leq i] a_{ij} = \sum_{j=1}^n \sum_{i=j}^n a_{ij}$$

ed ecco un esempio nella costruzione del duale:



6.12 ESEMPIO. Sia dato

$$\begin{aligned}
 \min \quad & \sum_{c \in C} \sum_{p \in P_c} L_p x_p \\
 & \sum_{p \in P_c} x_p = 1 \quad \forall c \in C \\
 & \sum_{\substack{p \in P_c \\ p \ni e}} x_p \leq y_e \quad \forall e \in E, \quad \forall c \in C \\
 & \sum_{e \in E} y_e = n - 1 \\
 & x_p \geq 0, y_e \geq 0
 \end{aligned} \tag{6.2}$$

Dove  $C$  è l'insieme di coppie di nodi di un grafo non orientato,  $P_c$  è l'insieme dei cammini fra i due nodi di  $c$ ,  $L_p$  è la lunghezza del cammino  $p$  (per lunghezze assegnate di archi),  $E$  è l'insieme degli archi del grafo,  $x_p$  è una variabile associata ad ogni cammino per ogni coppia di nodi e  $y_e$  è una variabile associata ad ogni arco  $e$ . Imponendo l'interezza alle variabili  $y_e$ , il problema modella il calcolo dell'albero di supporto in cui la somma delle distanze fra tutte le coppie di nodi è minima. Il primo gruppo di disequazioni impone che per ogni coppia di nodi sia scelto esattamente un cammino; il secondo gruppo attiva la variabile d'arco ogni qualvolta è scelto un cammino che adopera quel particolare arco e l'ultimo vincolo impone che siano attivati  $(n - 1)$  archi. Vogliamo far rilevare che il numero di colonne è straordinariamente elevato, ma ciononostante il problema si può affrontare. Vedremo più avanti che problemi simili possono essere risolti usando i vincoli duali, per cui è necessario poter costruire il duale. Si tratta effettivamente di un problema definito in modo complesso, ma casi simili sono abbastanza frequenti.

Per costruire il duale si proceda innanzitutto adottando simboli diversi per gli indici di colonna e per quelli di riga anche se si tratta degli stessi indici. Allora si indichi con  $D$  (e suoi elementi con  $d$ ) l'insieme  $C$  e con  $F$  (e suoi elementi con  $f$ ) l'insieme  $E$ , quando sono riferiti alle righe. Il problema è di minimizzazione e quindi dobbiamo riscrivere i vincoli di disequazione con il segno  $\geq$ . Pertanto (6.2) ridiventa:

$$\begin{aligned}
 \min \quad & \sum_{c \in C} \sum_{p \in P_c} L_p x_p \\
 & \sum_{p \in P_d} x_p = 1 \quad \forall d \in D \\
 - & \sum_{\substack{p \in P_d \\ p \ni f}} x_p + y_f \geq 0 \quad \forall f \in F, \quad \forall d \in D \\
 & \sum_{e \in E} y_e = n - 1 \\
 & x_p \geq 0, y_e \geq 0
 \end{aligned} \tag{6.3}$$

A questo punto vogliamo poter scrivere i vincoli in una forma omogenea in modo da poter

ottenere velocemente il duale. Per farlo usiamo l'operatore  $[\cdot]$  per cui (6.3) è equivalente a

$$\begin{aligned} \min \quad & \sum_{c \in C} \sum_{p \in P_c} L_p x_p \\ & \sum_{c \in C} \sum_{p \in P_c} [c = d] x_p = 1 \quad \forall d \in D \\ & - \sum_{c \in C} \sum_{p \in P_c} [f \in p] [c = d] x_p + \sum_{e \in E} [e = f] y_e \geq 0 \quad \forall f \in F, \quad \forall d \in D \\ & \sum_{e \in E} y_e = n - 1 \end{aligned}$$

Ora è facile scrivere il duale dove  $u_d$  sono le variabili duali associate al primo gruppo di vincoli,  $v_{df}$  quelle associate al secondo gruppo di vincoli e  $w$  la variabile associata all'ultimo vincolo. Solo le variabili  $v_{df}$  devono essere non negative. Il duale è

$$\begin{aligned} \max \quad & \sum_{d \in D} u_d + (n - 1) w \\ & \sum_{d \in D} [c = d] u_d - \sum_{d \in D} \sum_{f \in F} [f \in p] [c = d] v_{df} \leq L_p \quad \forall p \in P_c \quad \forall c \in C \\ & \sum_{d \in D} \sum_{f \in F} [e = f] v_{df} + w \leq 0 \quad \forall e \in E \\ & v_{df} \geq 0 \end{aligned} \tag{6.4}$$

A questo punto si possono 'togliere' alcuni operatori  $[\cdot]$  e ripristinare gli indici originari:

$$\begin{aligned} \max \quad & \sum_{c \in C} u_c + (n - 1) w \\ & u_c - \sum_{e \in E} [e \in p] v_{ce} \leq L_p \quad \forall p \in P_c \quad \forall c \in C \\ & \sum_{c \in C} v_{ce} + w \leq 0 \quad \forall e \in E \\ & v_{ce} \geq 0 \end{aligned} \tag{6.5}$$

Si noti che  $\sum_{e \in E} [e \in p] v_{ce}$  non è altro che la lunghezza del cammino  $p$  con lunghezze definite da  $v_{ce}$ . Indichiamo allora con  $L_p(v)$  la lunghezza del cammino  $p$  quando alle lunghezze originarie si sommi per ogni arco il costo  $v_{ce}$ . Otteniamo

$$\begin{aligned} \max \quad & \sum_{c \in C} u_c + (n - 1) w \\ & u_c \leq L_p(v) \quad \forall p \in P_c \quad \forall c \in C \\ & w \leq - \sum_{c \in C} v_{ce} \quad \forall e \in E \\ & v_{ce} \geq 0 \end{aligned}$$

Considerata la natura della funzione obiettivo nei termini di destra si possono sostituire le minime quantità. Sia allora  $\hat{L}_c(v)$  la lunghezza del cammino minimo fra i nodi della coppia

$c$  quando le lunghezze sono quelle originarie più i termini  $v_{ce}$  e sia  $K(v) := \max_c \sum_c v_{ce}$ . Alla fine si ha

$$\begin{aligned} \max \quad & \sum_{c \in C} u_c + (n-1)w \\ & u_c \leq \hat{L}_c(v) \quad \forall c \in C \\ & w \leq -K(v) \\ & v \geq 0 \end{aligned}$$

e si vede che il duale consiste nel trovare quei valori  $v \geq 0$  per cui  $\sum_c \hat{L}_c(v) - (n-1)K(v)$  è massimo. ■

6.13 ESERCIZIO. Si costruisca il duale di:

$$\begin{aligned} \max \quad & \sum_{s \in S} \sum_{d \in D(s)} r(s, d) x(s, d) \\ & \sum_{d \in D(s)} x(s, d) = \sum_{t \in S} \sum_{d \in D(t)} p(t, s, d) x(t, d) \quad \forall s \in S \\ & \sum_{s \in S} \sum_{d \in D(s)} x(s, d) = 1 \\ & x(s, d) \geq 0 \quad \forall d \in D(s), \quad \forall s \in S \end{aligned}$$

dove le variabili sono  $x(s, d)$ . Il problema massimizza il guadagno medio in un processo markoviano di decisione ad orizzonte infinito.  $S$  è l'insieme degli stati,  $D(s)$  è l'insieme delle decisioni disponibili nello stato  $s$ ,  $x(s, d)$  è la probabilità che il sistema sia nello stato  $s$  e venga presa la decisione  $d$ ,  $r(s, d)$  è il guadagno dovuto alla decisione  $d$  presa nello stato  $s$  e  $p(t, s, d)$  è la probabilità che avvenga una transizione dallo stato  $t$  allo stato  $s$  se viene adottata la decisione  $t$ . ■

6.14 ESERCIZIO. Si risolva il seguente problema (sfruttando il problema duale)

$$\begin{aligned} \max \quad & \sum_{j=1}^n c_j x_j \\ & \sum_{j=1}^i x_j \leq b_i \quad i := 1, \dots, n \\ & x_j \geq 0 \quad j := 1, \dots, n \end{aligned}$$

nell'ipotesi  $0 \leq b_1 \leq b_2 \leq \dots \leq b_n$  e  $0 \leq c_n \leq c_{n-1} \leq \dots \leq c_1$ . Si risolva nuovamente rimuovendo l'ipotesi. ■

## 6.4. Complementarità

Facendo riferimento alla forma canonica, se  $(\hat{x}, \hat{u})$  sono ottimi, la condizione di complementarità e la dualità forte implicano

$$\hat{u}(A\hat{x} - b) = 0 \quad (6.6)$$

Inoltre, siccome il duale del duale è il primale, le stesse condizioni applicate al duale e al duale del duale, danno

$$(\hat{u}A - c)\hat{x} = 0 \quad (6.7)$$

Le condizioni (6.6) e (6.7) non sono solo necessarie ma anche sufficienti, se  $\hat{x}$  e  $\hat{u}$  sono ammissibili.

6.15 TEOREMA.  $(\hat{x}, \hat{u})$  sono ottimi rispettivamente per il primale e il duale se e solo se

- sono ammissibili;
- $\hat{u}(A\hat{x} - b) = 0$ ;
- $(\hat{u}A - c)\hat{x} = 0$ .

DIMOSTRAZIONE. La necessità è già stata dimostrata. Per la sufficienza si noti che b) e c) implicano  $c\hat{x} = \hat{u}b$ , che, insieme con a) implica l'ottimalità. ■

Pertanto il senso delle relazioni di complementarità è che, in ottimalità, se  $\hat{x}_j > 0$  allora deve valere  $\hat{u}A^j = c_j$  e se  $\hat{u}A^j < c_j$  allora  $\hat{x}_j = 0$ . Analogamente se  $\hat{u}_i > 0$  allora  $A_i\hat{x} = b_i$  e se  $A_i\hat{x} > b_i$  allora  $\hat{u}_i = 0$ . Non si escludono i casi in cui si abbia sia  $\hat{x}_j = 0$  che  $\hat{u}A^j = c_j$  oppure sia  $\hat{u}_i = 0$  che  $A_i\hat{x} = b_i$ . Questa circostanza è di particolare importanza per le conseguenze teoriche e pratiche che comporta e quindi serve una definizione:

6.16 DEFINIZIONE. Un ottimo  $(\hat{x}, \hat{u})$  per il quale esiste un indice  $j$  per cui  $\hat{x}_j = 0$  e  $\hat{u}A^j = c_j$  oppure esiste un indice  $i$  per cui  $\hat{u}_i = 0$  e  $A_i\hat{x} = b_i$  si dice non regolare, altrimenti si dice regolare. ■

Questa definizione richiede una precisazione se nella descrizione del poliedro primale o duale compaiono delle equazioni. La riscrittura di un vincolo di eguaglianza come coppia di disequaglianze per portarlo in forma canonica conduce inevitabilmente, come è facile verificare, ad avere vertici ottimi non regolari. Pertanto intendiamo la definizione 6.16 in senso più forte *limitatamente ai soli vincoli di disequaglianza*.

Facciamo notare che non esiste in letteratura un termine universalmente accettato per indicare la 'regolarità'. Qualche volta si trova il termine 'complementarità stretta' (per la regolarità), altre volte 'degenerazione' (per la non regolarità), ma non c'è uniformità in generale. Qui si è scelto appunto il termine 'regolarità'.

Prima di indagare più a fondo sulle conseguenze della non regolarità, vogliamo far vedere l'importanza delle relazioni di complementarità, che risiede soprattutto nel fatto che è facile verificare se una certa soluzione ammissibile è anche ottima. Si consideri il seguente esempio:

6.17 ESEMPIO.

$$\begin{aligned} \max \quad & 3x_1 + x_2 - 2x_3 \\ & x_1 + x_2 - 4x_3 \leq 5 \\ & x_2 + 2x_3 \leq 8 \\ & -x_1 + x_2 \leq 4 \\ & x_1 + x_3 \leq 10 \\ & x_1, x_2, x_3 \geq 0 \end{aligned}$$

e si voglia verificare se la soluzione ammissibile  $x = (9, 0, 1)$  è ottima. Innanzitutto si scriva il duale:

$$\begin{aligned} \min \quad & 5u_1 + 8u_2 + 4u_3 + 10u_4 \\ & u_1 - u_3 + u_4 \geq 3 \\ & u_1 + u_2 + u_3 \geq 1 \\ & -4u_1 + 2u_2 + u_4 \geq -2 \\ & u_1, u_2, u_3, u_4 \geq 0 \end{aligned}$$

Poi si calcoli quali disequaglianze del problema primale sono verificate in modo stretto. Sostituendo si trova che sono la seconda e la terza disequaglianza. Per la complementarità questo implica che  $u_2 = u_3 = 0$ . Inoltre  $x_1 > 0$  e  $x_3 > 0$  implicano  $u_1 - u_3 + u_4 = 3$ , e

$-4u_1 + 2u_2 + u_4 = -2$ , cioè  $u_1 + u_4 = 3$ , e  $-4u_1 + u_4 = -2$ . Risolvendo il sistema si ottiene  $u_1 = 1$  e  $u_4 = 2$ . Bisogna verificare se questa soluzione è ammissibile per la seconda diseuguaglianza duale (che finora non è stata ancora considerata). Inoltre essendo  $u_1 > 0$  e  $u_4 > 0$  bisogna verificare se la prima e la quarta diseuguaglianza primale sono verificate come eguaglianze (verifica già fatta implicitamente). Essendo tutte le verifiche di complementarità soddisfatte possiamo concludere che  $\hat{x} = (9, 0, 1)$  è ottimo primale e che  $\hat{u} = (1, 0, 0, 2)$  è ottimo duale. Si noti che la soluzione è non regolare.

Supponiamo ora di esaminare la soluzione  $x = (6, 0, 4)$ . Sostituendo si trova che la prima e la terza diseuguaglianza sono soddisfatte strettamente, quindi  $u_1 = u_3 = 0$ . Imponendo questa condizione e  $u_1 - u_3 + u_4 = 3$  e  $-4u_1 + 2u_2 + u_4 = -2$  (da  $x_1 > 0$  e  $x_3 > 0$ ), si ottiene  $u_4 = 3$  e  $2u_2 = -5$ . Ma  $u_2$  deve essere non negativo e quindi, non essendo verificata la complementarità il punto non può essere ottimo. ■

6.18 ESERCIZIO. Sia dato il seguente problema:

$$\begin{aligned} \min \quad & x_1 + 3x_2 - x_3 - x_4 \\ & x_1 + x_2 + 2x_3 - x_4 = 2 \\ & 2x_1 - x_2 - x_3 + x_4 = 2 \\ & x_i \geq 0 \end{aligned}$$

Lo si risolva costruendo dapprima il duale, risolvendo questo per via grafica e infine ricavando l'ottimo primale dalle relazioni di complementarità. ■

6.19 ESERCIZIO. Calcolare gli ottimi primali e duali sapendo che in ottimalità si ha  $\hat{x}_2 = \hat{u}_2 = \hat{u}_3 = 0$ :

$$\begin{aligned} \min \quad & 2x_1 + 4x_2 + 3x_3 \\ & x_1 + x_2 \geq 10 \\ & 2x_1 - x_2 + 2x_3 \geq 25 \\ & x_2 - x_3 \leq 5 \\ & -x_1 + 3x_3 \geq 20 \\ & x_i \geq 0 \end{aligned}$$

Ora vogliamo vedere in dettaglio quali siano le conseguenze della non regolarità. Si supponga che il problema sia in forma canonica e lo si riscriva aggiungendo variabili di scarto nel primale e nel duale (per semplicità e senza perdita di generalità supponiamo che non vi siano vincoli di eguaglianza e che i poliedri primale e duale siano corpi convessi). I vincoli sono quindi:

$$\begin{aligned} Ax - Is = b & & uA + rI = c \\ x \geq 0 \quad s \geq 0 & & u \geq 0 \quad r \geq 0 \end{aligned} \quad (6.8)$$

Quindi la complementarità si può anche scrivere come  $\hat{x}_j \hat{r}_j = 0$  e  $\hat{u}_i \hat{s}_i = 0$ . Sia  $J$  l'insieme degli indici delle colonne di  $A$  e  $I$  quello delle righe. Si definiscano ora i seguenti insiemi di indici:  $Z_x = \{j \in J : \hat{x}_j = 0\}$ ,  $Z_r = \{j \in J : \hat{r}_j = 0\}$ ,  $Z_u = \{i \in I : \hat{u}_i = 0\}$ ,  $Z_s = \{i \in I : \hat{s}_i = 0\}$ . La complementarità implica  $Z_x \cup Z_r = J$  e  $Z_u \cup Z_s = I$ . Supponiamo che vi sia regolarità, cioè  $Z_x \cap Z_r = \emptyset$  e  $Z_u \cap Z_s = \emptyset$ . Possiamo rappresentare la situazione

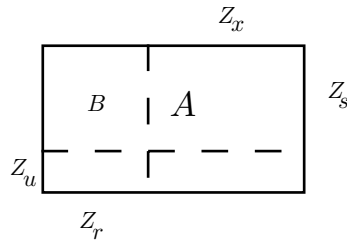


FIGURA 6.1

schematicamente come in figura 6.1 dove si è indicata una partizione di  $A$  secondo gli insiemi di indici.

Quindi le soluzioni diverse da zero (indichiamo con il simbolo  $B$  i vettori relativi alla matrice  $B$  di figura 6.1) devono soddisfare i rispettivi sistemi  $B \hat{x}_B = b_B$  e  $\hat{u}_B B = c_B$ . Dimostreremo fra poco che in condizioni di regolarità e di soluzione di vertice (esistono anche ottimi non vertici) la matrice  $B$  deve essere quadrata.

Un vertice di un poliedro in  $\mathbb{R}^n$  deve soddisfare come eguaglianza almeno  $n$  dei vincoli (diseguaglianze, equazioni o vincoli di non negatività) che definiscono il poliedro. Può succedere però che un vertice soddisfi come eguaglianza un numero di vincoli superiore alla dimensione dello spazio. Ad esempio la 'cima' di una piramide (di quelle egizie per intenderci) si trova nell'intersezione di quattro piani. Questa circostanza è particolarmente importante e merita una definizione.

**6.20 DEFINIZIONE.** *Un vertice di un poliedro in  $\mathbb{R}^n$  che sia determinato dall'intersezione di  $m > n$  piani viene detto degenerare.* ■

Vogliamo ora far vedere che per un ottimo la degenerazione, cioè la sovradeterminazione di un vertice, e la non regolarità, secondo la definizione 6.16, sono solo due facce diverse dello stesso fenomeno.

**6.21 ESERCIZIO.** Verificare che l'ottimo del problema 'trovare il punto più alto della piramide' è non regolare (oltre che ovviamente degenerare). ■

Facendo riferimento alla formulazione (6.8) vale il seguente risultato:

**6.22 TEOREMA.** *Siano  $\hat{x}$  e  $\hat{u}$  vertici ottimi dei rispettivi problemi. Allora  $\hat{x}$  e  $\hat{u}$  sono non degeneri, nei rispettivi spazi, se e solo se  $(\hat{x}, \hat{u})$  è regolare.*

**DIMOSTRAZIONE.** Si supponga che le soluzioni siano vertici regolari nel senso della Definizione 6.16. Per definizione di complementarità regolare si ha  $|Z_x| + |Z_r| = n$  e  $|Z_u| + |Z_s| = m$ , da cui  $|Z_x| + |Z_r| + |Z_u| + |Z_s| = n + m$ . Siccome  $(\hat{x}, \hat{s})$  è vertice, deve essere individuato da almeno  $m + n$  equazioni. Di queste,  $m$  sono già incluse nel vincolo  $Ax - Is = b$ . Le altre si ricavano imponendo uguali a zero o le  $x_j$  o le  $s_i$ . Quindi  $|Z_x| + |Z_s| \geq n$ . Analogamente si ricava per le variabili duali  $|Z_u| + |Z_r| \geq m$ . Quindi  $|Z_x| + |Z_r| + |Z_u| + |Z_s| \geq n + m$ . Ma la relazione precedente implica che la diseguaglianza è in realtà un'uguaglianza. Quindi  $|Z_x| + |Z_s| = n$  e  $|Z_u| + |Z_r| = m$ . Confrontando le relazioni si trova anche che  $|Z_s| = |Z_r|$  cioè  $B$  è quadrata. In modo analogo si può dimostrare l'implicazione inversa. ■

Quindi la non regolarità in soluzioni di vertice significa che, o per le variabili primali, o per quelle duali, il vertice è individuato da 'troppe' equazioni. La non regolarità è inoltre strettamente connessa con l'unicità degli ottimi.

6.23 TEOREMA. *Un vertice ottimo  $(\hat{x}, \hat{u})$  è ottimo unico se e solo se è regolare.*

DIMOSTRAZIONE. Si scriva il problema come in (6.8) e si indichi, per comodità di notazione,  $y := (x, s)$  e  $w := (u, r)$ . Si noti che  $y^1 \neq y^2$  implica  $x^1 \neq x^2$  e similmente per  $w$  e  $u$ . Supponiamo vi siano due ottimi distinti, ad esempio primali,  $y^1 \neq y^2$  e  $y^1$  sia vertice. Allora tutti i punti del segmento  $H$  congiungente  $y^1$  e  $y^2$  hanno lo stesso valore di funzione obiettivo. Inoltre se due componenti corrispondenti di  $y^1$  e  $y^2$  sono uguali, in particolare nulle, tutti i punti di  $H$  hanno il medesimo valore per quella componente. Siccome  $y^1 \neq y^2$ , almeno due componenti corrispondenti hanno valori diversi. Si noti che nella formulazione (6.8) la frontiera (nella topologia relativa) dell'insieme ammissibile è generata dai vincoli di non negatività. Essendo  $y^1$  vertice il punto  $\alpha y^1 + (1 - \alpha) y^2$  è non ammissibile per  $\alpha > 1$  e quindi  $y^1$  deve avere almeno una coordinata nulla fra le componenti diverse di  $y^1$  e  $y^2$ . Sia  $k$  questa componente, cioè  $y_k^1 = 0$  e  $y_k^2 > 0$ . Siccome  $y^2$  è ottimo, la complementarità implica  $w_k = 0$ . Ma  $y_k^1 = 0$  e  $w_k = 0$  significa non regolarità e quindi tutti i vertici ottimi sono non regolari.

Viceversa consideriamo il problema nella forma canonica con  $(\hat{x}, \hat{u})$  ottimi e sia  $v(y)$  la funzione di perturbazione. Supponiamo vi sia non regolarità, cioè esiste  $k$  tale che  $\hat{u}_k = 0$  e  $A^k \hat{x} = b_k$ . Se la variabile duale ottima non è unica la tesi è dimostrata. Sia quindi la variabile duale ottima  $\hat{u}$  unica. Quindi  $v(y)$  è differenziabile nell'origine e per ogni  $h$  esiste  $\varepsilon > 0$  tale che  $v(\varepsilon h) = v(0) - \varepsilon \hat{u} h$ . Si scelga in particolare  $\bar{h}_i = 0$ ,  $i \neq k$  e  $\bar{h}_k = -1$ . Quindi  $v(\varepsilon \bar{h}) = v(0) - \varepsilon \hat{u}_k \bar{h}_k = v(0)$ . Quindi il problema di programmazione lineare con il vincolo  $A^k x \geq b_k$  modificato in  $A^k x \geq b_k + \varepsilon$  è ammissibile con lo stesso valore ottimo. La soluzione  $\hat{x}$  non è ammissibile per il nuovo vincolo, quindi l'ottimo del problema modificato è un'altra soluzione  $\hat{x}'$ , che ovviamente è ammissibile per il problema originario e quindi la soluzione ottima non è unica. ■

Combinando assieme i due precedenti teoremi abbiamo:

6.24 COROLLARIO. *Se un vertice ottimo primale è degenero ed è ottimo unico, allora l'ottimo duale non è unico. Se un vertice ottimo primale non è unico e non è degenero, allora ogni vertice ottimo duale è degenero.* ■

Il precedente corollario stabilisce una corrispondenza fra la degenerazione di un vertice ottimo (nel senso di essere determinato da un numero sovrabbondante di faccette) e la non unicità dell'ottimo duale. Possiamo chiederci se la degenerazione e la non unicità possano essere entrambe confinate nel problema primale lasciando l'ottimo duale unico e non degenero. Ciò non può avvenire. Infatti è sempre possibile perturbare il vettore  $b$  del problema primale in modo da far scomparire la degenerazione mantenendo però la non unicità (dovuta essenzialmente a  $c$  e  $A$ ). A questo punto, per il corollario precedente la soluzione duale deve essere degenero, fatto dipendente da  $c$  e  $A$  e non da  $b$ . E quindi anche la soluzione duale del problema non perturbato deve essere degenero.

Nel teorema 6.23 si fa riferimento all'unicità di una soluzione ottima in generale, e non all'unicità di un vertice ottimo. Può infatti succedere che vi sia un unico vertice ottimo e che questo sia non regolare:

6.25 ESEMPIO. Si consideri la coppia primale-duale:

$$\begin{array}{ll}
 \min & x_2 \\
 & x_2 - s_1 = 1 \\
 & x_1, x_2, s_1 \geq 0
 \end{array}
 \qquad
 \begin{array}{ll}
 \max & u_1 \\
 & r_1 = 0 \\
 & u_1 + r_2 = 1 \\
 & u_1, r_1, r_2 \geq 0
 \end{array}$$

L'insieme ammissibile primale contiene l'unico vertice  $\hat{y} := (0, 1, 0)$  e quello duale i vertici  $w := (0, 0, 1)$  e  $\hat{w} := (1, 0, 0)$  dei quali il secondo è l'ottimo duale. Essendo il duale limitato e non vuoto il primale ammette ottimo che non può che essere il vertice  $\hat{y}$ . Confrontando  $\hat{y}$  con  $\hat{w}$  si vede che la soluzione è non regolare e il vertice ottimo è unico. Naturalmente non è unico l'ottimo primale. Ogni punto del tipo  $(a, 1, 0)$  è ottimo primale, ma non è un vertice. ■

L'unicità del vertice ottimo dell'esempio precedente è dovuta ad un fatto anomalo. Le soluzioni ottime stanno su una direzione estrema dell'insieme ammissibile. Escludendo questa possibilità si ottiene la non unicità dei vertici ottimi. Non è difficile dimostrare il seguente risultato:

6.26 COROLLARIO. *Se i poliedri primali e duale sono limitati, il vertice ottimo  $(\hat{x}, \hat{u})$  è unico se e solo se non c'è degenerazione.* ■

Un risultato notevole e particolarmente importante per gli algoritmi ai punti interni è che ottimi regolari esistono sempre anche quando vi siano vertici degeneri nel poliedro. Ovviamente questi ottimi regolari non sono vertici. Questo risultato è noto come teorema di Goldman e Tucker [1956]. Per una dimostrazione alternativa si veda anche Wright [1997] p. 35.

6.27 TEOREMA. *In un problema di programmazione lineare esistono sempre ottimi regolari.*

DIMOSTRAZIONE. La dimostrazione segue un po' in falsariga la seconda parte della dimostrazione del teorema 6.23. Consideriamo la regolarità riferita alle variabili duali  $\hat{u}$  e ai vincoli primali  $A\hat{x} \geq b$ . Costruiamo dapprima una soluzione ottima duale come combinazione convessa stretta di ottimi duali. Una componente  $\hat{u}_k$  di questa soluzione è nulla solo se la medesima componente è nulla per ogni soluzione duale ottima. Dobbiamo costruire allora una soluzione primale  $\hat{x}$  tale che  $A^k \hat{x} > b_k$  se  $\hat{u}_k = 0$ .

Sia allora  $v(y)$  la funzione di perturbazione. Per ogni  $h$  esiste  $\varepsilon > 0$  tale che  $v(\varepsilon h) = v(0) + \varepsilon Dv(0, h) = v(0) + \varepsilon \sup_{u \in \hat{U}} (-u)h$ , dove  $\hat{U}$  è l'insieme degli ottimi duali. Si scelga in particolare  $\bar{h}_i = 0$ ,  $i \neq k$  e  $\bar{h}_k = -1$  con  $k$  tale che  $\hat{u}_k = 0$ . Siccome  $\hat{u}_k = 0$  per ogni  $\hat{u} \in \hat{U}$  si ha  $v(\varepsilon \bar{h}) = v(0)$ . Quindi il problema di programmazione lineare con il vincolo  $A^k x \geq b_k$  modificato in  $A^k x \geq b_k + \varepsilon$  è ammissibile con lo stesso valore ottimo e con ottimo una soluzione  $\hat{x}'$  ammissibile per il problema originario e tale che  $A^k \hat{x}' \geq b_k + \varepsilon > b_k$ . Ripetendo la costruzione per ogni indice  $k$  tale che  $\hat{u}_k = 0$  si trovano altrettante soluzioni primali ammissibili ognuna delle quali soddisfa strettamente il vincolo primale corrispondente. Una combinazione convessa stretta di queste soluzioni soddisfa strettamente ogni vincolo primale per cui  $\hat{u}_k = 0$  e quindi è regolare. Una simile costruzione si può ripetere per le variabili  $x$  ed i vincoli duali. ■

6.28 ESERCIZIO. Sia dato  $c = 0$  in (6.8). Assumendo l'ammissibilità primale ogni soluzione è ottima. Si verifichi la complementarità regolare per una generica soluzione ammissibile e si indichi la ripartizione di  $A$  (secondo la figura 6.1). La matrice  $B$  è quadrata? ■

La complementarità può anche essere usata per calcolare un ottimo esatto da un ottimo approssimato. La situazione si può così riassumere: data una soluzione approssimata  $(\bar{x}, \bar{s}, \bar{u}, \bar{r}) \geq 0$ , e fissata una tolleranza  $\varepsilon > 0$  si ponga  $\bar{x}_j := 0$  se  $\bar{x}_j \leq \varepsilon$ ,  $\bar{x}_j := \bar{x}_j$  se  $\bar{x}_j > \varepsilon$ , e analogamente per le altre variabili. Se per  $(\bar{x}, \bar{s}, \bar{u}, \bar{r})$  avviene  $Z_x \cup Z_r = J$ ,  $Z_u \cup Z_s = I$ ,  $|Z_x| + |Z_s| \geq n$  e  $|Z_u| + |Z_r| \geq m$  possiamo a ragione pensare di essere 'vicino' ad un vertice ottimo. Se c'è regolarità i valori diversi da zero possono essere calcolati direttamente risolvendo i sistemi  $B x_B = b_B$  e  $u_B B = c_B$  e se le soluzioni sono non negative abbiamo la garanzia che la soluzione è ottima.



6.29 ESEMPIO. Sia data la coppia primale-duale:

$$\begin{array}{ll}
 \min & x_1 + x_2 + x_3 \\
 & -3x_1 + 2x_2 \geq 1 \\
 & 2x_1 - 3x_2 + x_3 \geq 1 \\
 & x_1, x_2, x_3 \geq 0
 \end{array}
 \qquad
 \begin{array}{ll}
 \max & u_1 + u_2 \\
 & -3u_1 + 2u_2 \leq 1 \\
 & 2u_1 - 3u_2 \leq 1 \\
 & u_2 \leq 1 \\
 & u_1, u_2 \geq 0
 \end{array}$$

e si conosca la seguente approssimazione dell'ottimo:

$$\begin{aligned}
 (\tilde{x}, \tilde{s}) &= ((0.002, 0.497, 2.49), (0.019, 0.032)) \\
 (\tilde{u}, \tilde{r}) &= ((1.97, 1.006), (4.89, 0.06, 0.004))
 \end{aligned}$$

Fissato  $\varepsilon = 0.1$  si consideri

$$\begin{aligned}
 (\bar{x}, \bar{s}) &= ((0, 0.497, 2.49), (0, 0)) \\
 (\bar{u}, \bar{r}) &= ((1.97, 1.006), (4.89, 0, 0))
 \end{aligned}$$

da cui  $Z_x = \{1\}$ ,  $Z_r = \{2, 3\}$ ,  $Z_u = \emptyset$ ,  $Z_s = \{1, 2\}$  che soddisfano la complementarità regolare e le relazioni  $|Z_x| + |Z_s| \geq n$  e  $|Z_u| + |Z_r| \geq m$  (soluzioni di vertice). La matrice  $B$  è data da:

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

e i sistemi da risolvere sono:

$$\begin{pmatrix} 2 & 0 \\ -3 & 1 \end{pmatrix} \begin{pmatrix} x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad (u_1 \quad u_2) \begin{pmatrix} 2 & 0 \\ -3 & 1 \end{pmatrix} = (1 \quad 1)$$

che danno come risultato:

$$(\hat{x}_2, \hat{x}_3) = \left( \frac{1}{2}, \frac{5}{2} \right) \qquad (\hat{u}_1, \hat{u}_2) = (2, 1)$$

Questi valori sono non negativi e quindi la soluzione è ammissibile (secondo la formulazione (6.8) la non negatività è l'unico vincolo di cui dobbiamo verificare l'ammissibilità; quelli di eguaglianza sono automaticamente imposti). Di conseguenza sono ottimi. ■

Se non vi è regolarità allora  $Z_x \cap Z_r \neq \emptyset$  oppure  $Z_u \cap Z_s \neq \emptyset$ . La situazione si può indicare schematicamente come in figura 6.2.

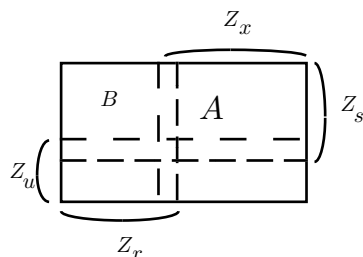


FIGURA 6.2

Anche in questo caso le soluzioni ottime possono essere ricercate limitandosi alla matrice  $B$ , non più necessariamente quadrata.  $B$  però può essere estesa ad una matrice quadrata aggiungendovi righe e/o colonne dagli insiemi  $Z_x \cap Z_r$  e/o  $Z_u \cap Z_s$ . Questo si può sempre fare. Vogliamo cioè far vedere che esiste  $k$  (ordine della matrice quadrata) tale che  $n - |Z_x| \leq k \leq |Z_r|$  e  $m - |Z_u| \leq k \leq |Z_s|$ . Ovviamente valgono sia  $|Z_r| + |Z_x| \geq n$  (complementarità) e  $|Z_x| + |Z_s| \geq n$  (individuazione di un vertice). Quindi  $n - |Z_x| \leq |Z_r|$  e  $n - |Z_x| \leq |Z_s|$ . Ponendo  $k = \min\{|Z_r|; |Z_s|\}$  si ha  $n - |Z_x| \leq k \leq |Z_r|$ . Similmente si ricava la seconda relazione.

6.30 ESEMPIO. Nell'esempio 6.17 si ha la situazione indicata in figura 6.3 con  $Z_x = \{2\}$ ,  $Z_r = \{1, 2, 3\}$ ,  $Z_u = \{2, 3\}$ ,  $Z_s = \{1, 4\}$ .

La matrice quadrata  $B$  è data dalle colonne 1 e 3 e dalle righe 1 e 4 e fornisce la soluzione precedentemente indicata  $\hat{x} = (9, 0, 1)$ ,  $\hat{u} = (1, 0, 0, 2)$ . Tuttavia si sarebbe potuto usare anche la matrice formata dalle colonne 1,2 e 3 e dalle righe 1 e 4. La soluzione duale rimane invariata ma la primale non è più univocamente determinata. Ogni soluzione del tipo  $x = (9 - a/5, a, 1 + a/5)$  soddisfa le relazioni di complementarità per valori sufficientemente piccoli di  $a$  ed è quindi ottima. Si tratta quindi di un caso di non unicità dell'ottimo. ■

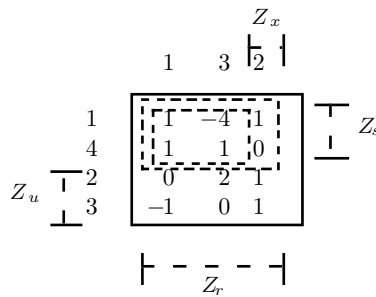


FIGURA 6.3

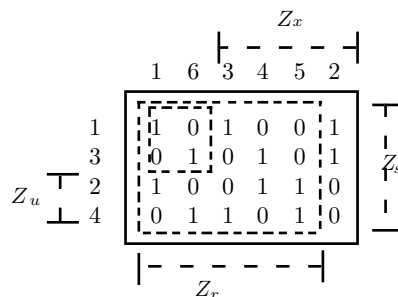


FIGURA 6.4

6.31 ESEMPIO. Sia data la coppia primale-duale:

$$\begin{array}{ll}
 \max & x_1 + x_3 + x_4 + x_6 \\
 & x_1 + x_2 + x_3 \leq 1 \\
 & x_1 + x_4 + x_5 \leq 1 \\
 & x_2 + x_4 + x_6 \leq 1 \\
 & x_3 + x_5 + x_6 \leq 1 \\
 & x_1, x_2, x_3, x_4, x_5, x_6 \geq 0 \\
 \min & u_1 + u_2 + u_3 + u_4 \\
 & u_1 + u_2 \geq 1 \\
 & u_1 + u_3 \geq 0 \\
 & u_1 + u_4 \geq 1 \\
 & u_2 + u_3 \geq 1 \\
 & u_2 + u_4 \geq 0 \\
 & u_3 + u_4 \geq 1 \\
 & u_1, u_2, u_3, u_4 \geq 0
 \end{array}$$

Si tratta di un problema di accoppiamento massimo pesato sul grafo completo di 4 nodi. L'ottimo non è unico. Si consideri ad esempio

$$\hat{x} = (1, 0, 0, 0, 0, 1) \qquad \hat{u} = (1, 0, 1, 0)$$

Si ottiene  $Z_x = \{2, 3, 4, 5\}$ ,  $Z_u = \{1, 3\}$ ,  $Z_s = \{1, 2, 3, 4\}$ ,  $Z_r = \{1, 3, 4, 5, 6\}$ . Come si vede c'è un elevato grado di degenerazione. La situazione è rappresentata in figura 6.4 dove

si vede la matrice con le colonne e le righe permutate come indicato. Ogni sottomatrice  $B$  inclusa fra quella più piccola (in tratteggio) e quella più grande (sempre in tratteggio) soddisfa i sistemi  $Bx_B = b_B$  e  $u_B B = c_B$ . ■

Le relazioni di complementarità suggeriscono una loro estensione a casi generici, cioè non soltanto a quelli in cui i dati provengono da una coppia primale-duale. Si riscrivano i vincoli (6.8) di complementarità come:

$$Mz - q = t \quad (z, t) \geq 0 \quad zt = 0 \quad (6.9)$$

dove

$$M := \begin{pmatrix} \mathbf{0} & -A^T \\ A & \mathbf{0} \end{pmatrix} \quad z := \begin{pmatrix} x \\ u^T \end{pmatrix} \quad q := \begin{pmatrix} -c^T \\ b \end{pmatrix} \quad t := \begin{pmatrix} r^T \\ s \end{pmatrix} \quad (6.10)$$

Il problema di trovare soluzioni ammissibili di (6.9) per una generica matrice simmetrica  $M$  (anziché il caso particolare di (6.10)) prende il nome di *problema della complementarità lineare* e si tratta di un problema **NP**-completo. Infatti esistono soluzioni ammissibili del seguente problema di knapsack

$$\sum_{i=1}^n a_i x_i = b \quad x_i = \{0, 1\}$$

se e solo se il seguente problema di complementarità lineare è ammissibile (verifica lasciata come esercizio)

$$M = \begin{pmatrix} -I_n & \mathbf{1}^T & -\mathbf{1}^T & 0 \\ \mathbf{1} & -1 & -1 & -1 \\ -\mathbf{1} & -1 & -1 & -1 \\ 0 & -1 & -1 & -1 \end{pmatrix} \quad q^T = (-a_1, \dots, -a_n, b, -b, 0)$$

Se  $M$  è positiva semidefinita il problema prende il nome di *complementarità lineare monotona* e si tratta di un problema polinomiale risolvibile con algoritmi ai punti interni. Se, infine,  $M$  è positiva semidefinita e del tipo in (6.10) il problema è ovviamente equivalente alla programmazione lineare.

## 6.5. Complessità computazionale della programmazione lineare

Consideriamo dapprima i seguenti due problemi: assegnato un poliedro tramite un insieme di disequazioni, quanto può essere grande la stringa che codifica i suoi vertici? E viceversa: assegnato un poliedro tramite un insieme di vertici, quanto può essere grande la stringa che codifica i coefficienti dei piani definenti le sue faccette? È chiaro che si tratta dello stesso problema, una volta espresso per un poliedro e l'altra per il suo polare. Definiamo allora:

**6.32 DEFINIZIONE.** Si dice che il poliedro  $P$  ha complessità di faccetta al più  $\varphi$  se la codifica di ogni disequaglianza definente una faccetta di  $P$  è minore o uguale a  $\varphi$ . ■

6.33 DEFINIZIONE. Si dice che il poliedro  $P$  ha complessità di vertice al più  $\nu$  se la codifica di ogni vertice di  $P$  è minore o uguale a  $\nu$ . ■

Per semplicità di trattazione e senza perdita di generalità, consideriamo il caso in cui i numeri definenti le istanze sono interi. Istanze definite con numeri razionali possono essere ricondotte a numeri interi con un cambiamento di scala pari al minimo comune multiplo dei denominatori dell'istanza. Deve essere chiaro tuttavia che, se  $P$  è definito da piani a coefficienti interi, non sono poi interi in generale i suoi vertici e viceversa, se  $P$  è assegnato tramite vertici interi, non sono interi in generale i coefficienti delle sue faccette.

Se  $P$  è assegnato come insieme di disequaglianze, sia  $M$  il massimo valore assoluto di tutti i numeri che definiscono  $P$  e sia  $L$  la lunghezza della stringa che codifica  $M$ , quindi  $L = \lfloor \log M \rfloor + 1$  (escluso il segno). Quindi  $\varphi := (n + 1)L$  è una limitazione superiore alla complessità di faccetta di  $P$ . Siccome interessa l'andamento asintotico delle relazioni possiamo tralasciare gli operatori  $\lceil \cdot \rceil$  o  $\lfloor \cdot \rfloor$  e le costanti additive nella valutazione di  $L$  in funzione di  $M$  e viceversa, per cui adotteremo la più semplice relazione  $L = \log M$ . Vogliamo ora dimostrare che la complessità di vertice è polinomiale in quella di faccetta e viceversa.

6.34 TEOREMA.  $\nu \in O(\varphi^3)$  e  $\varphi \in O(\nu^3)$ .

DIMOSTRAZIONE. Dimostriamo la prima limitazione, perché la seconda è analoga nello spazio polare. Un vertice si trova risolvendo un sistema lineare di  $n$  equazioni. Sia  $A$  la matrice di questo sistema lineare. Ogni coordinata di un vertice è quindi un numero razionale dato dal rapporto di due determinanti uno dei quali è  $\det A$ . È noto che  $|\det A|$  è il volume del parallelepipedo generato dalle righe di  $A$  tramite loro combinazioni lineari con coefficienti compresi fra 0 e 1. Quindi

$$|\det A| \leq \prod_i \sqrt{\sum_j a_{ij}^2} \leq (\sqrt{n} M)^n =: K \quad (6.11)$$

cioè

$$\log |\det A| \leq n \log M + \frac{n}{2} \log n \in O(n(L + \log n))$$

Quindi se  $M \geq n$  si ha  $\log |\det A| \leq (3/2)n \log M = (3/2)nL \leq (3/2)\varphi$ , mentre se  $M \leq n$  si ha

$$\log |\det A| \leq (3/2)n \log n \leq (3/2)n^2 \leq (3/2)n^2 L^2 \leq (3/2)\varphi^2$$

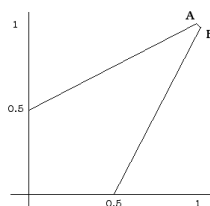
e allora il denominatore di ogni coordinata del vertice è un numero intero la cui stringa è codificata con al più  $(3/2)\varphi^2$  simboli. Siccome si ottiene la stessa limitazione per il numeratore, ogni coordinata del vertice è un razionale codificato con al più  $3\varphi^2$  simboli, e di conseguenza il vertice con  $3n\varphi^2 \leq 3nL\varphi^2 \leq 3\varphi^3$ . Quindi se la complessità di faccetta è al più  $\varphi$ , la complessità di vertice è al più  $3\varphi^3$ . ■

Questo significa che, se il sistema di disequaglianze che definisce un poliedro è costituito da coefficienti di poche cifre, anche i vertici si rappresentano con un numero di cifre limitato. Inoltre vertici distinti non possono essere arbitrariamente vicini. Infatti, se le coordinate corrispondenti di due vertici distinti sono esprimibili con i razionali  $p_1/q_1$  e  $p_2/q_2$ , allora la differenza è  $|p_1 q_2 - p_2 q_1|/(q_1 q_2) \geq 1/(q_1 q_2)$ , se le coordinate sono diverse. Siccome  $q_1, q_2 \leq K$  abbiamo il seguente risultato:

6.35 TEOREMA. La minima distanza in norma  $\|\cdot\|_\infty$  di due vertici distinti di un poliedro  $P$ , assegnato come insieme di disequaglianze con  $M$  massimo valore assoluto dei numeri definenti l'istanza, non può essere inferiore a  $K^{-2} = n^{-n} M^{-2n}$ . ■

6.36 ESEMPIO. Sia dato il seguente poliedro in  $\mathbb{R}^2$ :

$$\begin{aligned} -(a+2)x_1 + 2ax_2 &\leq a \\ 2ax_1 - (a+2)x_2 &\leq a \\ (a-1)x_1 + (a+1)x_2 &\leq 2a \\ x_1, x_2 &\geq 0 \end{aligned}$$



Sia  $a = 50$  (valore usato in figura). Quindi  $M = 100$  e  $K = (\sqrt{2} \cdot 100)^2 = 2 \cdot 10^4$ . La complessità di faccetta, in base 10, è al più  $\varphi = (2 + 1)L = 9$ . In base al teorema la complessità di vertice è al più  $3\varphi^3 = 2187$ . Quindi abbiamo la garanzia che ogni coordinata è esprimibile con meno di 1100 simboli. Si tratta di un valore molto alto anche perché la limitazione data dal teorema è molto ampia. Volendo usare la limitazione più stretta (6.11), per ogni coordinata di vertice otteniamo che sia numeratore che denominatore sono limitati da  $K$  e siccome  $\lceil \log K \rceil + 1 = 5$  ogni coordinata di vertice è un razionale in cui denominatore e numeratore non richiedono separatamente più di 5 simboli. I vertici A e B hanno infatti coordinate:

$$\begin{aligned} A &= \left( \frac{a(-1+3a)}{2+a+3a^2}, \frac{3a(1+a)}{2+a+3a^2} \right) = \left( \frac{3725}{3776}, \frac{3825}{3776} \right) \\ B &= \left( \frac{a(5+3a)}{-2+3a+3a^2}, \frac{a+3a^2}{-2+3a+3a^2} \right) = \left( \frac{3875}{3824}, \frac{3775}{3824} \right) \end{aligned}$$

Inoltre c'è la garanzia che due vertici sono distanti almeno  $K^{-2} = 2.5 \cdot 10^{-9}$ . La differenza fra A e B vale

$$\left( -\frac{24225}{902464}, \frac{23275}{902464} \right) \approx \{-0.026, 0.025\}$$

e la distanza (in norma  $\|\cdot\|_\infty$ ) è circa 0.026 (molto più grande di fatto della limitazione trovata). ■

Un'altra conseguenza importante del teorema precedente è che per il problema della programmazione lineare in forma ricognitiva: 'esiste  $x \in X : cx < w$  e  $Ax \geq b$ ?' esiste un certificato polinomiale costituito dalle coordinate di un vertice (se non esistessero vertici, l'affermazione rimane valida con un po' di 'aggiustamenti'; li lasciamo come esercizio al lettore). Quindi la programmazione lineare sta in **NP**.

Come conseguenza importante della validità della dualità forte abbiamo che il problema complementare della programmazione lineare in forma ricognitiva: 'è vero che non esiste  $x \in X$  tale che  $cx < w$  e  $Ax \geq b$ ?' sta in **NP**. Infatti basta fornire una variabile duale  $u$  ammissibile tale che  $L(u) \geq w$  e il teorema 6.10 garantisce l'esistenza di tale  $u$ . Quindi la programmazione lineare sta in **NP**  $\cap$  **co-NP**.

6.37 ESERCIZIO. Si è visto che può avvenire che problemi duali e primali siano entrambi non ammissibili. Quale tipo di certificato polinomiale si può utilizzare in questo caso per il complementare della programmazione lineare? ■

Fu provato solo nel 1979 da Khaciyan che la programmazione lineare appartiene a **P**. La dimostrazione si basava sui risultati di tipo generale ottenuti nella decade precedente da un gruppo di matematici sovietici, fra cui soprattutto N.Z. Shor, i quali, partendo dai metodi di quasi ascesa, avevano sviluppato delle tecniche che facevano uso di metriche dello spazio variabili (si veda ad esempio Shor [1970] oppure Yudin e Nemirovski [1976]). Khaciyan [1979] applicò queste tecniche alla programmazione lineare ottenendone un algoritmo polinomiale (ma non fortemente polinomiale) che ora illustreremo. Il metodo è noto come *algoritmo*

dell'ellissoide perché fa uso costante di ellissoidi che rappresentano la metrica variabile dello spazio.

Prima però di illustrare l'algoritmo dell'ellissoide è bene precisare che questo algoritmo, pur essendo polinomiale, non è praticamente utilizzabile, perché si tratta pur sempre di una complessità elevata e i calcoli richiesti sono instabili numericamente. Però la sua importanza teorica rimane fondamentale, non solo per ragioni storiche, ma anche per i motivi che saranno esposti nella prossima sezione. Un ellissoide è in generale rappresentabile nel seguente modo

$$E(c, Q) := \{x : (x - c) Q^{-1} (x - c) \leq 1\}$$

dove  $c$  è il centro, gli autovettori della matrice positiva definita  $Q$  definiscono le direzioni degli assi e le radici quadrate degli autovalori definiscono le lunghezze dei semiassi. Il volume di un ellissoide è dato da  $\text{vol}(E(c, Q)) = \sqrt{\det Q} \text{vol}(S_n)$  dove  $S_n$  è la sfera unitaria in  $\mathbb{R}^n$ . Il volume della sfera unitaria in  $\mathbb{R}^n$  è dato dalle seguenti formule:

$$\text{vol}(S_{2n}) = \frac{\pi^n}{n!} \quad \text{vol}(S_{2n-1}) = \frac{\pi^{n-1}}{n!} \frac{2^{2n}}{\binom{2n}{n}}$$

Dato un poliedro limitato  $P$  è essenziale, per poter usare l'algoritmo dell'ellissoide, avere a disposizione due ellissoidi, uno contenente  $P$  e l'altro contenuto in  $P$ , i cui volumi siano in rapporto limitato da un'opportuna funzione dei dati dell'istanza.

Un risultato teorico di Löwner e di John [1948] (si veda Danzer e al. [1963]) afferma che, dato un corpo convesso  $F$  limitato, esistono due ellissoidi  $E^+$  e  $E^-$  concentrici, con  $Q^+ = n^2 Q^-$  (cioè  $E^+$  si ottiene da  $E^-$  espandendo  $E^-$  di un fattore  $n$ ) e tali che  $E^- \subset F \subset E^+$ . Allora si ha  $\text{vol}(E^+) = n^n \text{vol}(E^-)$ . Purtroppo questi ellissoidi non sono calcolabili in tempo polinomiale. Calcoleremo ora in modo rudimentale, ma efficace ai fini della determinazione teorica della polinomialità del metodo dell'ellissoide, due ellissoidi, uno contenente e l'altro contenuto nel poliedro  $P := \{x \in \mathbb{R}^n : Ax \geq b\}$ .

Per quel che riguarda l'ellissoide esterno a  $P$  usiamo la limitazione (6.11). Ogni coordinata di ogni vertice è limitata superiormente da  $K = (\sqrt{n} M)^n$ . Quindi una sfera  $S^+$  di raggio  $R := \sqrt{n} K$  e centro nell'origine contiene i vertici di  $P$ .

Per quel che riguarda l'ellissoide interno a  $P$  consideriamo il seguente problema di programmazione lineare

$$\begin{aligned} \max \quad & r \\ & A^k (x + z^j r) \geq b_k \quad \forall j, \forall k \end{aligned} \quad (6.12)$$

dove  $z^j$ ,  $j = 1, \dots, 2^n$ , sono tutti i vettori con coordinate uguali a +1 oppure a -1. Il problema (6.12) calcola il vettore  $x$ , centro di un cubo di lato  $2r$ , in modo che i vertici del cubo siano in  $P$  e il lato sia massimo. Notiamo che la massima sfera inscritta si può calcolare più facilmente con il seguente problema:

$$\begin{aligned} \max \quad & r \\ & r \sqrt{A^k A^k} \leq A^k x - b_k \quad \forall k \end{aligned} \quad (6.13)$$

Tuttavia, ai fini della determinazione in generale di una limitazione inferiore per  $r$ , la presenza di numeri irrazionali in (6.13) crea qualche problema. Questo non succede nel problema (6.12). Il fatto che (6.12) presenti un numero esponenziale di vincoli non è rilevante perché non dobbiamo risolvere il problema, ma solo ricavare delle limitazioni inferiori.

Se  $P$  è un corpo convesso il valore ottimo  $\hat{r}$  in (6.12) è strettamente maggiore di zero e, in base alla limitazione (6.11) è non minore di  $K^{-1}$  (correttamente è non minore di

$(\sqrt{n+1}M)^{n+1}$  dato che in (6.12) sono presenti  $n+1$  variabili, tuttavia, ai fini di valutazioni asintotiche, possiamo ragionare come se ci fossero  $n$  variabili). La sfera  $S^-$ , di raggio  $r := K^{-1}$  e centrata nel centro del cubo, è inscritta nel cubo fornito da (6.12) e quindi è inscritta in  $P$ . Allora  $R/r = \sqrt{n}K^2$  e

$$\rho := \frac{\text{vol}(S^+)}{\text{vol}(S^-)} = \left(\frac{R}{r}\right)^n = (\sqrt{n}K^2)^n = (\sqrt{n}n^n M^{2n})^n = n^{n^2+n/2} M^{2n^2} \quad (6.14)$$

Si noti che  $\log \rho = (n^2 + n/2) \log n + 2n^2 \log M = (n^2 + n/2) \log n + 2n^2 L$ , è polinomiale nei dati dell'istanza.

Prima di descrivere l'algoritmo che risolve un problema di programmazione lineare, consideriamo il seguente sottoproblema: siano dati un ellissoide  $E := E(c, Q)$  ed un piano  $H(a, b) := \{x : ax = b\}$  che interseca l'ellissoide. Sia  $H^+(a, b) := \{x : ax \geq b\}$ , e sia anche  $c \notin H^+(a, b)$ . Si vuole trovare un nuovo ellissoide  $E' := E(c', Q') \supset E(c, Q) \cap H^+(a, b)$  di volume minimo. La formula che definisce un tale ellissoide è data da:

$$Q' := \alpha \left( Q - \frac{\beta}{aQa} Qa a Q \right) \quad c' := c + \frac{\gamma}{\sqrt{aQa}} Qa$$

dove  $\alpha, \beta, \gamma$  e successivamente  $\Delta$  sono dati da

$$\alpha := \frac{n^2(1-\Delta^2)}{n^2-1} \quad \beta := 1 - \frac{(1-\Delta)(n-1)}{(1+\Delta)(n+1)} \quad \gamma := \frac{1+\Delta n}{n+1} \quad \Delta := \frac{b-ac}{\sqrt{aQa}}$$

$\Delta$  rappresenta la distanza del piano dal centro misurata nella metrica dell'ellissoide. Quindi  $\Delta = 0$  se il piano passa per il centro,  $\Delta = 1$  se il piano è di supporto all'ellissoide. Se  $\Delta > 0$  il centro non è contenuto nel semispazio  $H(a, b)$  e viceversa vi è contenuto se  $\Delta < 0$ . Si usa parlare di 'tagli' profondi se  $\Delta$  si avvicina ad 1. Il rapporto fra i volumi dei due ellipsoidi è dato da:

$$\frac{\text{vol}(E')}{\text{vol}(E)} = \frac{n}{n+1} \left( \frac{n^2}{n^2-1} \right)^{\frac{n-1}{2}} (1-\Delta)(1-\Delta^2)^{\frac{n-1}{2}} \leq e^{-\frac{1}{2(n+1)}} e^{-\Delta-\Delta^2\frac{(n-1)}{2}} \leq e^{-\frac{1}{2(n+1)}} =: \mu < 1 \quad (6.15)$$

dove le limitazioni sono state ottenute sfruttando la relazione  $1 + \alpha \leq e^\alpha$ .

Questa tecnica può essere usata per determinare una soluzione ammissibile di un poliedro  $P$  definito da un sistema lineare di disequazioni, purché il poliedro sia un corpo convesso. L'algoritmo procede generando ellipsoidi nel seguente modo:

- il primo ellissoide  $E$  è dato dalla sfera  $S^+$  centrata nell'origine e di raggio  $R$  con  $R^2 = n^{n+1} M^{2n}$ . Quindi  $E := E(c, Q)$  con  $c := \mathbf{0}$  e  $Q := R^2 I$ ;
- si verifica se il centro  $c$  dell'ellissoide appartiene al poliedro. Se  $c$  è ammissibile l'algoritmo termina perché si è trovata una soluzione ammissibile;
- se  $c$  non è ammissibile, allora si sceglie una disequazione violata e questa definisce il semispazio  $H^+$ . Ovviamente  $P \subset H^+ \cap E$  e  $c \notin H^+$ . Da  $E$  e  $H^+$  si definisce in base alle formule un nuovo ellissoide  $E'$  per il quale è garantita la proprietà  $P \subset E'$ . Si pone  $E := E'$  e si itera.

A meno che  $P$  sia vuoto, si ha necessariamente  $\text{vol}(E) \geq \text{vol}(S^-)$ . In base alla limitazione (6.15), dopo  $k$  iterazioni si ha

$$\text{vol}(E) \leq \mu^k \text{vol}(S^+) = \mu^k \rho \text{vol}(S^-) \leq \mu^k \rho \text{vol}(E)$$

e quindi  $1 \leq \mu^k \rho$  da cui  $k(-\log \mu) \leq \log \rho$  e siccome  $-\log \mu = 1/(2n+2)$  abbiamo

$$k \leq 2(n+1) \log \rho = 2(n+1)((n^2+n/2) \log n + 2n^2 \log M) \in O(n^3(\log n + L))$$

che indica un numero polinomiale di iterazioni del metodo. Se, dopo  $2(n+1) \log \rho$  iterazioni non si è trovata una soluzione ammissibile allora  $P$  è vuoto. Normalmente avviene che in qualche iterazione  $\Delta > 1$  e questo indica non ammissibilità.

Per ogni iterazione si richiede un calcolo di complessità  $O(n^2)$  per aggiornare i parametri dell'ellissoide e un calcolo per determinare una disuguaglianza violata. Se tale calcolo è eseguito per semplice sostituzione di  $c$  in ogni disuguaglianza, la complessità computazionale è  $O(nm)$ . Quindi la complessità di trovare una soluzione ammissibile di  $Ax \geq b$  è data da  $O(mn^4(\log n + L))$  siccome  $m \geq n$  (altrimenti non ci sarebbero vertici).

Per quel che riguarda la programmazione lineare come problema di ottimizzazione, bisogna subito notare che non si può trasformare il problema di ottimizzazione in un problema di ammissibilità e risolverlo come si è appena visto. Infatti gli unici punti ammissibili in tali problemi sono gli ottimi e quasi mai gli ottimi formano un corpo convesso. Conviene invece procedere nel seguente modo (*slittamento dell'obiettivo - sliding objective function*): ogni qualvolta si trova una soluzione ammissibile  $\bar{x}$  si taglia  $P$  aggiungendo la disuguaglianza  $cx \leq c\bar{x}$  e usando questa come disuguaglianza violata. In questo modo  $P$  viene costantemente ridotto fino ad un poliedro molto piccolo, i cui elementi costituiscono tutti una buona approssimazione dell'ottimo.

È necessario a questo punto operare un 'aggiustamento' dei dati del problema in modo da evitare molteplicità degli ottimi. Si perturbi il vettore  $c$  della funzione obiettivo nel seguente vettore:

$$c' := \xi^n c + (1, \xi, \xi^2, \dots, \xi^{n-1})$$

dove  $\xi := 2K^2 + 1$ . La scelta di questo valore è dovuta al fatto che per ogni coppia di vertici  $\hat{x}$  e  $\tilde{x}$  la differenza  $\hat{x}_i - \tilde{x}_i$  è un razionale con numeratore inferiore a  $\xi$  e denominatore non superiore a  $K^2$ . Si può dimostrare (esercizio non banale che viene lasciato al lettore) che  $\min \{c'x : x \in P\}$  ha un unico ottimo che è anche ottimo di  $\min \{cx : x \in P\}$ . Per semplicità di notazione s'intende che, negli sviluppi successivi,  $c$  è il vettore perturbato.

Vogliamo ora determinare quante iterazioni sono necessarie per ottenere un'approssimazione  $\varepsilon$  del valore ottimo e successivamente valuteremo qual è il valore di  $\varepsilon$  che garantisce una 'buona' approssimazione dell'ottimo. Si supponga allora di voler trovare una soluzione ammissibile al problema:

$$\begin{aligned} cx &\leq c\hat{x} + \varepsilon \\ Ax &\geq b \end{aligned} \tag{6.16}$$

dove  $\hat{x}$  è un ottimo e  $\varepsilon$  è un'approssimazione prefissata. Sia  $P_\varepsilon$  il poliedro definito da (6.16). Si noti che  $P_\varepsilon \subset P_{\varepsilon'}$  se e solo se  $\varepsilon \leq \varepsilon'$ . Una soluzione ammissibile di (6.16) approssima l'ottimo con errore assoluto (sulla funzione obiettivo perturbata) al più  $\varepsilon$ . Il problema (6.16) è teorico, perché vi compare  $\hat{x}$ , ovviamente non noto. Viene introdotto per determinare una relazione fra  $\varepsilon$  e il volume di una sfera inscritta nel corpo convesso (6.16). A tale scopo riscriviamo il problema (6.12) come

$$\begin{aligned} \max \quad & r \\ & c(x + z^j r) \leq c\hat{x} + \varepsilon \quad \forall j \\ & A^k(x + z^j r) \geq b_k \quad \forall j, \forall k \end{aligned} \tag{6.17}$$

Sia  $\varepsilon := \sigma^{-1}$  dove  $\sigma$  è un intero. Quindi (6.17) diventa:

$$\begin{aligned} \max \quad & r \\ & \sigma c(x + z^j r) \leq \sigma c\hat{x} + 1 \quad \forall j \\ & A^k(x + z^j r) \geq b_k \quad \forall j, \forall k \end{aligned} \tag{6.18}$$



Sia  $\bar{M} = \sigma M \xi^n$  una limitazione superiore del numero più grande (in valore assoluto) nella matrice di (6.18), dove  $M$  è ridefinito includendo anche il vettore  $c$  (i valori originali). I termini di destra (fra cui  $\hat{x}$ ) non entrano nella valutazione perché il minimo valore che può assumere  $r$  è dato dall'inverso del determinante di una sottomatrice quadrata della matrice dei vincoli. Quindi  $r = (\sqrt{n} \bar{M})^{-n}$  e il rapporto (6.14) fra i volumi delle sfere circoscritta ed inscritta diventa

$$\rho := \frac{\text{vol}(S^+)}{\text{vol}(S^-)} = \left(\frac{R}{r}\right)^n = (\sqrt{n} n^n M^n \bar{M}^n)^n = n^{n^2+n/2} M^{n^2} \bar{M}^{n^2} \tag{6.19}$$

con  $\log \rho = (n^2 + n/2) \log n + n^2 (\log M + \log \bar{M}) = (n^2 + n/2) \log n + n^2 (2 \log M + \log \sigma + n \log \xi)$ . Quindi, ragionando come nel caso precedente, il numero di iterazioni  $k$  dell'algoritmo dell'ellissoide è limitato da

$$\bar{k} := 2(n + 1) \log \rho = 2(n + 1) ((n^2 + n/2) \log n + n^2 (2 \log M + \log \sigma + n \log \xi)) \tag{6.20}$$

e si ottiene facilmente, operando le varie sostituzioni

$$\bar{k} \in O(n^3 (\log \sigma + n \log n + nL)) \tag{6.21}$$

Si tratta ora di determinare quale valore di  $\varepsilon$  permette una 'buona' approssimazione del vertice ottimo. Anche se praticamente può essere consigliabile fermare l'iterazione quando alcune componenti del vettore  $(\bar{x}, \bar{s})$  sono abbastanza piccole, imporre il valore 0 per queste componenti e poi verificare la complementarità, ed eventualmente continuare ancora l'iterazione se la complementarità non potesse essere soddisfatta, dobbiamo però ora fare uso di una procedura che teoricamente garantisca un numero polinomiale di operazioni.

Quindi, per 'buona' approssimazione si intende una soluzione che possa essere arrotondata in modo polinomiale in un vertice ottimo. Affinché la procedura d'arrotondamento fornisca in modo inequivoco l'ottimo, non si può operare un semplice arrotondamento delle singole componenti di un vettore dato. L'obiettivo dell'arrotondamento è di ottenere un vettore di razionali con denominatore comune limitato e il più possibile vicino alla soluzione approssimata. A questo scopo si definisca il seguente problema:

**6.38 DEFINIZIONE.** (Problema dell'approssimazione diofantea simultanea - PADS) *Siano dati  $n$  razionali  $\bar{x}_i$ ,  $i := 1, \dots, n$ , un intero  $K$  ed un razionale  $0 < \eta < 1$ . Si determinino  $n$  interi  $p_1, \dots, p_n$  e un intero  $q \leq K$  tali che  $|q \bar{x}_i - p_i| \leq \eta$ , oppure si determini la non ammissibilità dell'istanza.* ■

Quindi si vuole trovare un vettore  $p/q$  di razionali sufficientemente vicino al vettore dato  $\bar{x}$  e con denominatore sufficientemente piccolo. Il PADS è collegato con la programmazione lineare in quanto  $\bar{x}$  è il valore ammissibile che si ottiene dall'algoritmo dell'ellissoide e  $p/q$  è il vertice ottimo, dato che i vertici sono razionali con denominatore limitato. Purtroppo il PADS è NP-difficile e quindi, se vogliamo arrotondare con un procedimento polinomiale, dobbiamo far ricorso al seguente risultato più debole (Nemhauser e Wolsey [1988] p. 200):

**6.39 TEOREMA.** *Esiste un algoritmo polinomiale che, per ogni istanza del PADS, fornisce esattamente una delle due risposte:*

- determina la non ammissibilità dell'istanza;
- trova  $n$  interi  $p_1, \dots, p_n$  e un intero  $q$  tali che

$$|q \bar{x}_i - p_i| \leq \theta \eta \qquad q \leq \theta K \tag{6.22}$$

con  $\theta = \sqrt{2^n (n + 1)}$ . ■

Questo risultato può sembrare in contrasto con l'asserita difficoltà del PADS. Però si tenga presente che non viene detto che l'algoritmo polinomiale fornisce la risposta di non ammissibilità per *ogni* istanza non ammissibile. Se così fosse l'algoritmo costituirebbe un certificato polinomiale per le istanze di tipo 'no' e la versione di decisione del PADS sarebbe in **co-NP**, cosa improbabile per un problema **NP-completo**. Quindi, anche se l'istanza non è ammissibile, l'algoritmo potrebbe fornire i valori  $p_i$  e  $q$  che soddisfano le più ampie limitazioni (6.22). Inoltre non si pensi di poter 'ingannare' il PADS prendendo in (6.22) dei valori di  $\eta$  e  $K$  scalati del fattore  $\theta$ . La nuova istanza del PDAS con valori  $\eta/\theta$  e  $K/\theta$  potrebbe non essere ammissibile e l'algoritmo potrebbe fornire proprio questa risposta.

La procedura di arrotondamento di  $\bar{x}$  verrà eseguita con l'algoritmo del teorema 6.39. Dobbiamo perciò fare in modo che i valori forniti da tale algoritmo corrispondano all'ottimo e solo all'ottimo, cioè che  $\hat{x} = p/q$  sia l'unica soluzione di (6.22) e allo stesso tempo sia ammissibile per il PADS. Per ottenere questo risultato dobbiamo prima trovare una limitazione per la distanza fra  $\bar{x}$  e  $\hat{x}$  a partire dal fatto che  $c\bar{x} \leq c\hat{x} + \varepsilon$ . Tutte le norme che verranno ora usate sono  $\|\cdot\|_\infty$ .

6.40 LEMMA. Se  $c\bar{x} \leq c\hat{x} + \varepsilon$ ,  $\varepsilon < K^{-2}$ , allora  $\|\bar{x} - \hat{x}\| \leq 2\varepsilon K^3$ .

DIMOSTRAZIONE. Come risultato intermedio vogliamo dimostrare che  $\bar{x}$  è combinazione convessa di un insieme di vertici di cui fa parte necessariamente il vertice ottimo  $\hat{x}$ . Sia  $\{\tilde{x}^1, \dots, \tilde{x}^k\}$  tale insieme di vertici e sia  $c\tilde{x}^1 = \min_i c\tilde{x}^i$ . Allora  $c\bar{x} = \sum_i \alpha_i c\tilde{x}^i \geq c\tilde{x}^1$ , e dall'ipotesi  $c\bar{x} \leq c\hat{x} + \varepsilon$  si ha  $c\tilde{x}^1 \leq c\hat{x} + \varepsilon$ . Se  $\tilde{x}^1$  è un vertice non ottimo si ha, dai risultati precedenti,  $\|\tilde{x}^1 - \hat{x}\| \geq K^{-2}$  e quindi  $c\tilde{x}^1 \geq c\hat{x} + K^{-2}$ . Per l'ipotesi su  $\varepsilon$  e  $K$ , questa disuguaglianza contraddice la precedente e quindi fra i vertici che generano  $\bar{x}$  è presente anche il vertice ottimo  $\hat{x}$ . Sia allora  $\bar{x} = \alpha_0 \hat{x} + \sum_i \alpha_i \tilde{x}^i$  con  $\alpha_0 > 0$ ,  $\alpha_i \geq 0$ ,  $\alpha_0 + \sum_i \alpha_i = 1$ . Quindi

$$c\bar{x} = \alpha_0 c\hat{x} + \sum_i \alpha_i c\tilde{x}^i \geq \alpha_0 c\hat{x} + \sum_i \alpha_i (c\hat{x} + K^{-2}) = c\hat{x} + (1 - \alpha_0) K^{-2}$$

Siccome  $c\bar{x} \leq c\hat{x} + \varepsilon$  si ottiene la limitazione  $(1 - \alpha_0) \leq \varepsilon K^2$ . Ora si ha

$$\hat{x} - \bar{x} = (1 - \alpha_0) \hat{x} - \sum_i \alpha_i \tilde{x}^i = (1 - \alpha_0) \left( \hat{x} - \sum_i \frac{\alpha_i}{1 - \alpha_0} \tilde{x}^i \right) =: (1 - \alpha_0) (\hat{x} - y)$$

dove  $y \in P$ , dato che è generato come combinazione convessa di vertici. La massima distanza (in norma  $\|\cdot\|_\infty$ ) fra due punti qualsiasi di  $P$  è  $2K$  e quindi abbiamo  $\|\hat{x} - \bar{x}\| \leq 2\varepsilon K^3$ . ■

Sia ora  $x'$  un qualsiasi vettore di razionali esprimibile come  $x'_i = p_i/q$  con  $q \leq \theta K$ . Anche  $\hat{x}$  è uno di questi vettori. Se  $x' \neq \hat{x}$  si ha  $\|x' - \hat{x}\| \geq \theta^{-2} K^{-2}$ . Supponiamo ora che  $x'$  sia fornito dall'algoritmo del teorema 6.39, cioè si abbia  $q \|\bar{x} - x'\| \leq \theta \eta$ . Dobbiamo preliminarmente assicurarci che l'istanza del PADS con valori  $\bar{x}$ ,  $K$  e  $\eta$  sia ammissibile in modo che l'algoritmo non possa fornirci la risposta di non ammissibilità. Se  $\|\bar{x} - \hat{x}\| \leq \eta K^{-1}$ ,  $\hat{x}$  è una tale soluzione ammissibile. Basta scegliere  $\varepsilon \leq \eta K^{-4}/2$  per garantire l'ammissibilità del PADS.

Si tratta ora di scegliere  $\eta$  in modo che  $x' = \hat{x}$ . A tale scopo si consideri la disuguaglianza triangolare  $\|\bar{x} - \hat{x}\| + \|\bar{x} - x'\| \geq \|\hat{x} - x'\|$ . Entrambi i termini di sinistra sono limitati superiormente da  $\theta \eta$  e quello di destra è limitato inferiormente da  $\theta^{-2} K^{-2}$ , se  $\hat{x} \neq x'$ . Basta perciò scegliere  $2\theta \eta < \theta^{-2} K^{-2}$  per rendere inconsistenti le disuguaglianze e quindi avere  $x' = \hat{x}$ . Allora  $\eta < \theta^{-3} K^{-2}/2$ , per cui bisogna scegliere

$$\varepsilon < \frac{1}{4} \theta^{-3} K^{-6} \implies \sigma > 4\theta^3 K^6 \quad (6.23)$$

Ricapitolando, scelto  $\eta < \theta^{-3} K^{-2}/2$ , si deve far iterare l'algoritmo per almeno  $\bar{k}$  iterazioni, con  $\bar{k}$  dato da (6.20). Sia  $\bar{x}$  la migliore soluzione ammissibile ottenuta dopo  $\bar{k}$  iterazioni. Quindi si ha  $\bar{x} \in P_{c\bar{x}-c\hat{x}}$ . Dopo  $\bar{k}$  iterazioni abbiamo la garanzia che il volume dell'ellissoide  $E$  contenente il poliedro  $P_{c\bar{x}-c\hat{x}}$  è minore del volume della sfera  $S$  inscritta in  $P_\varepsilon$  con  $\varepsilon = \eta K^{-4}/2$  e quindi si ha

$$\text{vol}(P_{c\bar{x}-c\hat{x}}) \leq \text{vol}(E) \leq \text{vol}(S) \leq \text{vol}(P_\varepsilon)$$

e quindi necessariamente (data la natura dei poliedri  $P_\alpha$ )  $P_{c\bar{x}-c\hat{x}} \subset P_\varepsilon$  e quindi ancora  $c\bar{x} - c\hat{x} \leq \varepsilon$ . A questo punto possiamo arrotondare la soluzione  $\bar{x}$  usando l'algoritmo approssimato del PADS con valori  $\eta$  e  $K$ .

Tenendo conto di (6.21), di (6.23) e del fatto che in ogni iterazione si richiede un lavoro  $O(nm)$ , si è ottenuto il seguente risultato:

**6.41 TEOREMA.** *L'algoritmo dell'ellissoide risolve un problema di programmazione lineare in tempo polinomiale con complessità  $O(mn^5(\log n + L))$ .* ■

Ci sono tuttavia ancora dei punti da chiarire, che sono stati volutamente lasciati da parte per non appesantire la trattazione, già di per sé alquanto complicata. Le formule di aggiornamento del centro dell'ellissoide richiedono il calcolo di una radice quadrata, ovvero fanno uso (teorico) di numeri irrazionali e, in ogni caso, non ci si è preoccupati che i numeri che descrivono i vari ellissoidi rimangano limitati polinomialmente. Bisogna invece porre un limite alle cifre necessarie a rappresentare i numeri, approssimare i numeri quando il limite viene violato e fare in modo che questo algoritmo modificato mantenga le caratteristiche del precedente in cui le operazioni sugli ellissoidi erano svolte con codifica aritmetica e misura uniforme. Questo tipo di aggiustamento dei calcoli è alquanto complesso e non viene presentato qui. Il lettore può trovarlo descritto in Lovász [1986] oppure in Grötschel e al. [1988].

**6.42 ESEMPIO.** Sia dato il problema:

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

Per valutare il raggio iniziale tale da contenere in modo garantito il poliedro notiamo che  $M = 3$ ,  $n = 2$  e quindi ogni coordinata di vertice è al massimo  $K = (\sqrt{n}M)^n = (\sqrt{2} \cdot 3)^2 = 18$ . Allora possiamo porre  $R^2 = nK^2 = 2 \cdot 18^2 = 648$  e quindi si scelgono come valori iniziali  $c = (0, 0)$  e  $Q = 648I$ . Si trovano i seguenti valori  $\theta = 3.46$ ,  $\eta = 3.71 \cdot 10^{-5}$ ,  $\sigma = 5.65 \cdot 10^9$ ,  $\xi = 649$  che comportano un numero massimo di 924 iterazioni. La funzione obiettivo viene perturbata in

$$\xi^n c + (1, \xi) = 649^2 (1, 1) + (1, 649) = (421202, 421850)$$

In figura 6.5 sono raffigurate le prime quattro iterazioni (figure in scala). Alla nona iterazione il centro dell'ellissoide è contenuto nel poliedro e quindi si passa ad usare come piano la funzione obiettivo. In figura 6.6 sono raffigurate questa iterazione e le tre successive

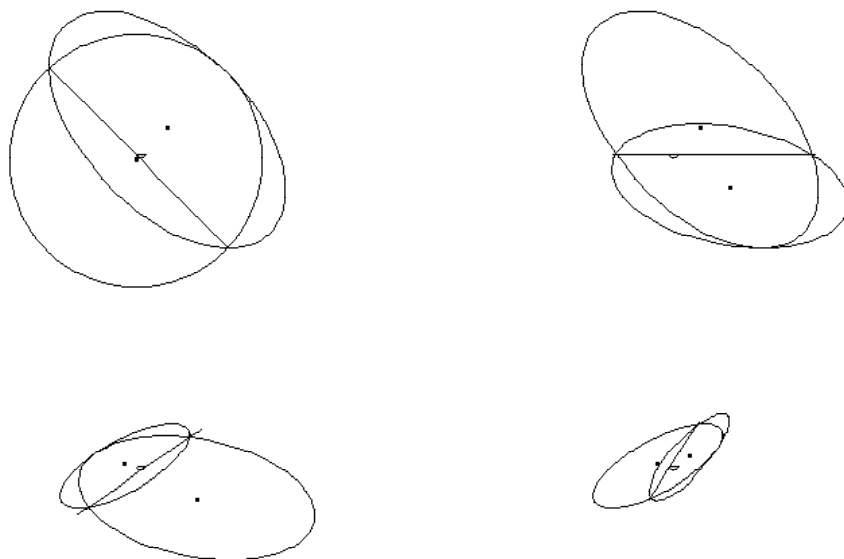


FIGURA 6.5

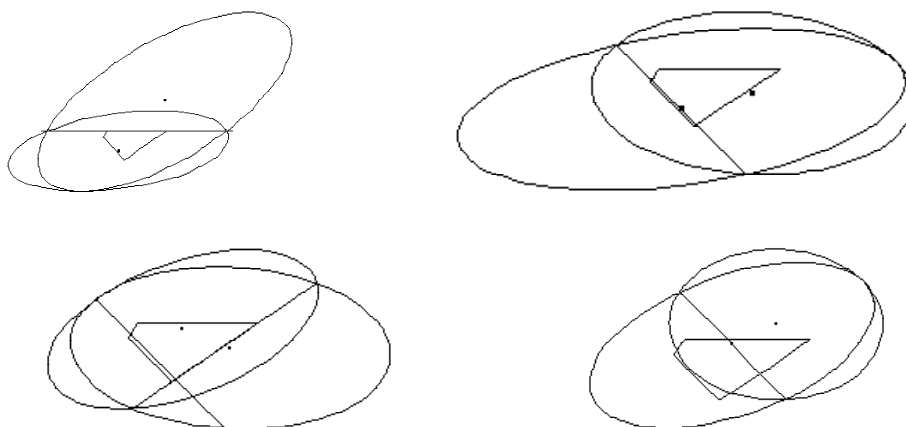


FIGURA 6.6

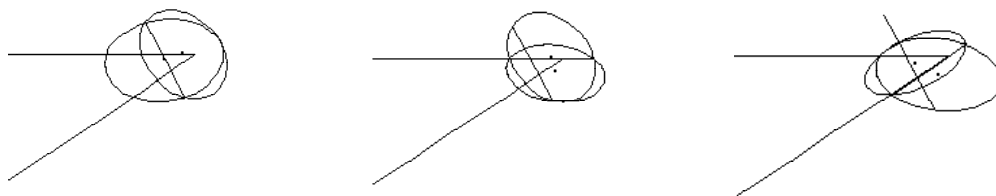


FIGURA 6.7

(figure non in scala). Si noti come parte del poliedro non sia più contenuta nell'ellissoide, in quanto l'ellissoide deve semplicemente contenere  $P_{c\bar{x}-c\hat{x}}$ . In figura 6.7 si vedono le iterazioni 20, 21 e 22 (figure non in scala). La soluzione che si ottiene alla 22-ma iterazione è  $x = (1.849, 0.972)$ , da cui  $s = (4.602, 0.219, 0.0273, 1.822)$ . Invece di proseguire con le iterazioni ed arrivare a dei valori che teoricamente possono essere arrotondati in modo univoco tramite il PADS, proviamo subito ad arrotondare a zero i valori piccoli, e quindi assumiamo  $Z_x = \emptyset$  e  $Z_s = \{2, 3\}$ . Proviamo a risolvere  $Bx_b = b_B$  con

$$B := \begin{pmatrix} 2 & -3 \\ 0 & 1 \end{pmatrix}$$

ottenendo  $x = (2, 1)$  ammissibile. Per verificare l'ottimalità bisogna calcolare la variabili duali da  $u_B B = c_b$  ottenendo  $(u_2, u_3) = (1/2, 5/2)$ . Trattandosi di una soluzione ammissibile è anche ottima, dato che le condizioni imposte corrispondono alla complementarità. ■

## 6.6. Il metodo dell'ellissoide in ottimizzazione combinatoria

Anche se il metodo dell'ellissoide non è raccomandabile come pratico metodo di calcolo, vi sono tuttavia delle conseguenze molto importanti per l'ottimizzazione combinatoria che furono notate per la prima volta da Grötschel e al. [1981, 1984]. Si noti che il valore  $m$ , numero di disequaglianze, che appare nella complessità computazionale dell'intero algoritmo è unicamente dovuto al fatto di dover esaminare le disequaglianze una alla volta per scoprire quella eventualmente violata. Quindi, come già accennato, se questa determinazione può avvenire per altra via, indipendentemente dal numero di disequaglianze, l'ammissibilità di un poliedro si determina in tempo polinomiale anche se lo stesso è definito da un numero esponenziale di disequaglianze.

Per capire l'importanza, soprattutto in ottimizzazione combinatoria, di questo risultato presentiamo ora quattro particolari problemi definiti su poliedri. Sia dunque  $P$  un poliedro in  $\mathbb{R}^n$  e definiamo i due problemi elementari di decisione:

6.43 DEFINIZIONE. (Problema dell'appartenenza) Dato  $\bar{x} \in \mathbb{R}^n$ ,  $\bar{x} \in P$ ? ■

6.44 DEFINIZIONE. (Problema della validità) Dato un semispazio chiuso  $H^+$ ,  $P \subset H^+$ ? ■

I due problemi di decisione possono venire riformulati pretendendo come soluzione del problema, qualcosa di più di un semplice 'no' quando la risposta è negativa. In altri termini si vuole un certificato che provi la risposta negativa.

6.45 DEFINIZIONE. (Problema della separazione) Dato  $\bar{x} \in \mathbb{R}^n$ , si decida se  $\bar{x} \in P$  oppure no, e se  $\bar{x} \notin P$  si fornisca un piano che separi  $\bar{x}$  da  $P$ . ■

6.46 DEFINIZIONE. (Problema della violazione) Dato un semispazio chiuso  $H^+$ , si decida se  $P \subset H^+$  oppure no, e se  $P \not\subset H^+$  si fornisca un punto  $y \in P$ ,  $y \notin H^+$ . ■

Una semplice riflessione porta a notare che il problema dell'appartenenza e quello della validità sono due formulazioni diverse dello stesso problema, la prima riferita al poliedro  $P$  e la seconda al poliedro polare  $P^*$ . Altrettanto si può dire per il problema della separazione e per quello della violazione. Algoritmicamente però i problemi possono essere fondamentalmente diversi a seconda di come è descritto il poliedro  $P$ .

Se  $P$  è descritto tramite un insieme di  $m$  disequazioni, il problema dell'appartenenza si risolve verificando l'ammissibilità di  $\bar{x}$  per ognuna delle disequazioni. Se almeno una disequazione è violata l'istanza è di tipo 'no' e tale disequazione fornisce automaticamente un piano separatore, per cui anche il problema della separazione è facilmente risolto. Il problema della validità è invece equivalente ad un problema di programmazione lineare. Sia  $H^+ := \{x : cx \geq \alpha\}$  e  $P := \{x : a^i x \geq b_i, i := 1, \dots, m\}$ . Allora la risposta è 'sì' se e solo se  $\min \{cx : x \in P\} \geq \alpha$ . Nel caso la risposta sia 'no', l'ottimo del problema di programmazione lineare fornisce automaticamente il certificato richiesto dal problema della violazione.

Se  $P$  è descritto come elenco di vertici  $\hat{x}^1, \dots, \hat{x}^m$ , allora la situazione si ribalta, come è intuibile, dato che si lavora nello spazio polare. Il problema della validità e della violazione sono immediati: basta verificare che tutti i vertici soddisfino  $H^+ := \{x : a^0 x \geq b_0\}$ , e se almeno uno non soddisfa la disequazione, tale vertice è proprio il certificato richiesto dal problema della violazione. Però per verificare l'ammissibilità di un punto  $\bar{x} \in \mathbb{R}^n$  si deve vedere se  $\bar{x}$  si può esprimere come combinazione convessa dei vertici di  $P$ , cioè se esiste una soluzione ammissibile di

$$\begin{aligned} \sum_i \alpha_i &= 1 \\ \sum_i \alpha_i \hat{x}^i &= \bar{x} \\ \alpha_i &\geq 0 \end{aligned} \quad (6.24)$$

problema di programmazione lineare risolvibile tramite la coppia primale-duale

$$\begin{aligned} \min \quad & z^+ + z^- \\ & \sum_i \alpha_i = 1 \\ & \sum_i \alpha_i \hat{x}^i + z^+ - z^- = \bar{x} \\ & \alpha_i, z^+, z^- \geq 0 \end{aligned} \quad \begin{aligned} \max \quad & u_0 + u \bar{x} \\ & u_0 + u \hat{x}^i \leq 0, \quad i := 1, \dots, m \\ & u \leq \mathbf{1} \\ & u \geq -\mathbf{1} \end{aligned} \quad (6.25)$$

dove il duale fornisce direttamente il piano separatore se (6.24) non è ammissibile.

Si supponga ora che il poliedro sia algoritmicamente definito da una procedura che risolve il problema della separazione e supponiamo inoltre che la complessità computazionale di tale procedura sia costante, per cui consideriamo la procedura come un 'oracolo' associato al poliedro che, ogni qualvolta viene presentato un punto  $\bar{x}$ , fornisce l'informazione richiesta.

Non si può tuttavia fare a meno di associare all'oracolo una misura di quanto complessa sia la 'descrizione' del poliedro. Se l'oracolo deve fornire un piano separatore, i valori numerici di tale piano non possono non dipendere dalla descrizione del poliedro. Quindi, per poter fare delle considerazioni utili, dobbiamo pensare che l'oracolo sia provvisto di una 'garanzia' che assicura l'utente che la descrizione del poliedro è limitata in una qualche misura. Tale garanzia non riguarda tanto il numero di faccette o di vertici, che potrebbero essere esponenziali, quanto piuttosto la complessità di faccetta o di vertice.

Un oracolo viene pertanto detto *oracolo di separazione razionalmente garantito* se esiste una limitazione  $\varphi$  sulla complessità di faccetta, oppure una limitazione  $\nu$  sulla complessità di vertice. Un tale oracolo richiede, per essere usato, la lettura della garanzia, per cui contribuisce con una lunghezza d'ingresso pari al valore  $\varphi$  oppure  $\nu$  (come se questi dati fossero codificati in unario).

Ci chiediamo se sia possibile risolvere in tempo polinomiale il problema della violazione una volta ammessa l'esistenza di un oracolo di separazione razionalmente garantito. Si è già detto che, se il semispazio è definito da  $H^+ := \{x : cx \geq \alpha\}$ , allora un'istanza del problema della violazione è di tipo 'sì' se e solo se  $\min \{cx : x \in P\} \geq \alpha$ . Si tratta quindi di un problema di programmazione lineare di tipo particolare in quanto l'istanza non viene descritta nel modo usuale, ma tramite l'oracolo di separazione razionalmente garantito. Se la garanzia è riferita alla complessità di vertice, il valore assoluto di ogni coordinata di ogni vertice di  $P$  è al più  $K = 10^\varphi$  (in base 10). Se invece la garanzia è riferita alla complessità di faccetta, il teorema 6.34 assicura che le coordinate dei vertici sono 'sufficientemente' limitate.

In entrambi i casi esiste una sfera circoscritta il cui logaritmo del raggio è polinomiale nei dati dell'istanza (cioè la lunghezza della garanzia espressa in unario – perché in questo caso si assume come valida la codifica unaria?). Se il poliedro è un corpo convesso, esiste una sfera inscritta di raggio sufficientemente 'grande' grazie alla garanzia. In caso contrario si può dimostrare che è possibile determinare l'insieme affine che contiene il poliedro (con un ragionamento che prevede un'approssimazione diofantea simultanea per determinare i coefficienti dell'insieme affine) e quindi restringere i calcoli all'insieme affine.

Ora, noti i valori dei raggi delle sfere inscritta e circoscritta a  $P$ , si può applicare il metodo dell'ellissoide con un numero di iterazioni polinomiale. Finché non si trova una soluzione ammissibile l'oracolo fornisce automaticamente anche un piano separatore che permette l'applicazione del metodo dell'ellissoide. Nel momento in cui si trova una soluzione ammissibile si usa lo slittamento dell'obiettivo. La complessità del metodo dipende polinomialmente da  $n$  e  $\varphi$ , come non è difficile dimostrare sfruttando i vari risultati.

Analogamente, il problema della separazione può essere risolto polinomialmente tramite un oracolo di violazione razionalmente garantito sfruttando il duale di (6.25). Il valore ottimo di (6.25) è maggiore di zero se e solo se  $\bar{x} \notin P$  e l'ottimo fornisce il piano separatore qualora  $\bar{x} \notin P$ .

Si consideri il poliedro  $\tilde{P} \subset \mathbb{R}^{n+1}$  delle variabili duali ammissibili in (6.25). Non è difficile dimostrare che  $\tilde{P}$  è un corpo convesso. La complessità di faccetta di  $\tilde{P}$  è la medesima della complessità di vertice di  $P$ . Pertanto si determinano facilmente i raggi utili delle sfere circoscritta a  $\tilde{P}$  ed inscritta in  $\tilde{P}$ . Per applicare l'algoritmo dell'ellissoide a  $\tilde{P}$  si noti che il centro dell'ellissoide è un vettore  $(c_0, c)$ . Se è violata una delle disequaglianze  $-1 \leq c_i \leq 1$ , la si usa per passare al successivo ellissoide. Altrimenti si consideri il semispazio  $H^+ := \{x : c_0 \leq -cx\}$  e si invochi l'oracolo di violazione. Se la risposta dell'oracolo è  $P \subset H^+$  nessuna delle disequaglianze che definiscono  $\tilde{P}$  è violata e si può verificare se  $c_0 > -c\bar{x}$ . In caso affermativo il problema della separazione è risolto e  $(c_0 + c\bar{x})/2 = cx$  è il piano separatore cercato. Altrimenti si usa il piano  $c_0 + c\bar{x} = 0$  per tagliare l'ellissoide. Se l'oracolo dà come risposta un punto  $y \in P$  tale che  $c_0 + cy > 0$ , si usa il piano  $c_0 + cy = 0$ . Considerazioni simili alle precedenti permettono di concludere che il numero di passi dell'algoritmo dell'ellissoide è polinomiale.

L'importanza di questi risultati in ottimizzazione combinatoria è dovuta al fatto che in molti casi l'insieme delle soluzioni ammissibili di un problema combinatorio viene associato ad un insieme  $X$  di vettori d'incidenza in un opportuno spazio euclideo. Se la funzione obiettivo è lineare  $\min \{cx : x \in X\}$  è equivalente a  $\min \{cx : x \in \text{conv}(X)\}$ . Tipicamente il poliedro  $\text{conv}(X)$  ha un numero esponenziale sia di faccette che di vertici, ma questo fatto non preclude la possibilità di un algoritmo polinomiale se è possibile costruire un algoritmo polinomiale che simuli l'oracolo di separazione. Per questa via è quindi possibile dimostrare la polinomialità di alcuni problemi.

Sul versante negativo, si supponga di studiare un problema NP-difficile e si supponga di avere a disposizione un poliedro  $P$ . Vogliamo sapere se  $P = \text{conv}(X)$ . Essendo il pro-

blema **NP**-difficile è molto probabile che un oracolo di separazione polinomiale non esista, altrimenti  $\mathbf{P} = \mathbf{NP}$  in base ai precedenti risultati. In alcuni casi si può identificare un poliedro  $P$  candidato a soddisfare l'uguaglianza  $P = \text{conv}(X)$ . Se avviene che, dato  $\bar{x}$  siamo in grado di decidere in tempo polinomiale se  $x \in P$  e, se  $x \notin P$ , fornire, sempre in tempo polinomiale, una disuguaglianza violata allora il problema **NP**-difficile sarebbe risolvibile in tempo polinomiale e quindi abbiamo valide ragioni per credere che  $P \neq \text{conv}(X)$ .

L'uso che si fa dell'algoritmo dell'ellissoide può ricordare i celebri 'Gedankenexperimente' della fisica, esperimenti virtuali ma rispettosi delle leggi della fisica e quindi validi nelle loro implicazioni teoriche. L'uso dell'algoritmo dell'ellissoide va visto come un 'Gedankenalgorithm', un algoritmo virtuale, che non necessita di alcuna implementazione pratica, ma le cui implicazioni teoriche sono assolutamente valide.

6.47 ESEMPIO. Si consideri il poliedro  $P$  dell'esempio 4.89 di cui ripetiamo la definizione:

$$\sum_{i \in J} x_i \geq \frac{|J|(|J| + 1)}{2} \quad \forall J \subset N, J \neq N$$

$$\sum_{i \in N} x_i = \frac{|N|(|N| + 1)}{2}$$

Ci chiediamo se il problema  $\min \{cx : x \in P\}$  è polinomiale. Ci basta sapere se il problema di separazione è polinomiale. Si noti che  $P$  non è un corpo convesso, però la difficoltà si aggira facilmente lavorando sull'insieme affine di  $P$ . Quindi supponiamo che tutti i punti che considereremo soddisfino l'uguaglianza  $\sum_{i=1}^n x_i = n(n+1)/2$ . Pertanto sia dato un punto  $y$ . Se vogliamo trovare una disuguaglianza violata, cioè trovare un sottoinsieme  $J$  tale che  $\sum_{i \in J} y_i < |J|(|J| + 1)/2$ , possiamo pensare di suddividere tutti i sottoinsiemi a seconda della cardinalità. Fra tutti i sottoinsiemi di cardinalità  $m$  basta considerare quello che contiene gli indici con i valori minori di  $y_i$ . Quindi si tratta di riordinare in modo crescente le componenti del vettore  $y$  (per comodità di notazione sia  $y_1, y_2, \dots, y_n$  tale ordine) e considerare gli  $n-1$  valori  $a_k := \sum_{i=1}^k y_i$  per  $k := 1, \dots, n-1$ . Se avviene  $a_k < k(k+1)/2$  per qualche  $k$  allora si è identificata una disuguaglianza violata. Altrimenti  $y$  è ammissibile. Quindi il problema della separazione è risolvibile in tempo polinomiale e il problema  $\min \{cx : x \in P\}$  è polinomiale. A dire il vero, un algoritmo polinomiale di complessità  $O(n \log n)$  è facilmente deducibile dalla struttura stessa del problema senza far ricorso all'algoritmo dell'ellissoide. ■

6.48 ESEMPIO. Si consideri il problema di determinare un albero di supporto in un grafo che minimizzi la somma delle distanze sull'albero fra tutte le coppie di nodi. È noto che si tratta di un problema **NP**-difficile. Una modellizzazione di tale problema è stata considerata nell'esempio 6.12. Ci chiediamo se i vertici del poliedro definito dal problema primale (6.2) possono essere frazionari.

Si consideri allora il duale (6.5). Anche se il poliedro di (6.5) ha un numero esponenziale di disuguaglianze il problema di separazione è polinomiale perché si può effettuare risolvendo  $n(n-1)/2$  problemi di cammino minimo fra tutte le coppie di nodi con costi di arco non negativi. Allora il problema (6.5) è polinomiale e quindi sono da attendersi vertici frazionari in (6.2) altrimenti il problema in questione sarebbe polinomiale. ■