Ricerca Operativa

Programmazione lineare

Prof. Paolo Brandimarte
Dip. di Scienze Matematiche – Politecnico di Torino
e-mail: `paolo.brandimarte@polito.it`
URL: `staff.polito.it/paolo.brandimarte`

Questa versione: 31 maggio 2023

Contenuto

Le slide seguenti sono tratte dal capitolo 7 di: P. Brandimarte, *Ottimizzazione per la Ricerca Operativa*, CLUT 2022.

- Geometria della programmazione lineare.
- L'algoritmo del simplesso primale.
- Dualità nella programmazione lineare.

Consideriamo un problema LP in forma standard:

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

dove $\mathbf{A} \in \mathbb{R}^{m,n}$, con $m < n$, ha rango massimo e $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{c} \in \mathbb{R}^n$. Se consideriamo la matrice tecnologica come un insieme di n colonne,

$$\mathbf{A} = \begin{bmatrix} \vdots & \vdots & & \vdots \\ \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_n \\ \vdots & \vdots & & \vdots \end{bmatrix}$$

dove $\mathbf{A}_j \in \mathbb{R}^m$, $j \in [n]$, possiamo interpretare il problema LP come la ricerca del modo più economico di esprimere $\mathbf{b}$ come combinazione conica delle colonne di $\mathbf{A}$:

$$\sum_{j=1}^{n} \mathbf{A}_j x_j = \mathbf{b}; \qquad x_j \geq 0,\ j \in [n].$$

Per il momento tralasciamo il requisito di non negatività e concentriamoci sulle combinazioni lineari delle colonne.

Ci possono essere infiniti modi di esprimere $\mathbf{b}$ come combinazione lineare delle colonne, ma è sufficiente considerare un sottoinsieme $\mathcal{B}$ di m colonne linearmente indipendenti, che costituisce una base per $\mathbb{R}^m$:

$$\sum_{j \in \mathcal{B}} \mathbf{A}_j x_j = \mathbf{b}.$$

Data una base $\mathcal{B}$, le **variabili basiche** $\{x_j \mid j \in \mathcal{B}\}$ che possono assumere un valore diverso da zero. Le rimanenti $n - m$ **variabili non basiche** sono lasciate a zero.

Una soluzione associata a una base, con al più m variabili che assumono valori non nulli, è detta **soluzione basica**. Se la soluzione basica soddisfa anche il requisito $x_j \geq 0$, $j \in \mathcal{B}$, essa è una **soluzione basica ammissibile**.

Esempio. Consideriamo il sistema di equazioni lineari:

$$\begin{bmatrix} -1 & 1 & 1 & -1 & 0 \\ 0 & 1 & 0 & 4 & 0 \\ 0 & 0 & 2 & 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}.$$

Una soluzione basica, corrispondente alla base formata dalle colonne $\mathbf{A}_1$, $\mathbf{A}_2$ e $\mathbf{A}_5$, è

$$x_1 = 2,\ x_2 = 3,\ x_3 = x_4 = 0,\ x_5 = 1.$$

Questa base è anche ammissibile. Al contrario, se consideriamo la base formata dalle colonne $\mathbf{A}_2, \mathbf{A}_3, \mathbf{A}_5$, troviamo la soluzione basica

$$x_1 = 0,\ x_2 = 3,\ x_3 = -2,\ x_4 = 0,\ x_5 = 5,$$

che però non è ammissibile, poiché $x_3 < 0$.

Data una base ammissibile, possiamo partizionare il vettore $\mathbf{x}$ nei due sottovettori $\mathbf{x}_B \in \mathbb{R}^m$ delle variabili basiche e $\mathbf{x}_N \in \mathbb{R}^{n-m}$ delle variabili non basiche. Allo stesso modo possiamo riordinare le colonne di $\mathbf{A}$ e riscrivere il sistema di equazioni lineari

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

come

$$\left[\mathbf{A}_B \mathbf{A}_N\right] \begin{bmatrix} \mathbf{x}_B \\ \mathbf{x}_N \end{bmatrix} = \mathbf{A}_B \mathbf{x}_B + \mathbf{A}_N \mathbf{x}_N = \mathbf{b}, \tag{1}$$

dove $\mathbf{A}_B \in \mathbb{R}^{m \times m}$ è invertibile e $\mathbf{A}_N \in \mathbb{R}^{m \times (n-m)}$. Possiamo quindi esprimere $\mathbf{x}$ come

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_B \\ \mathbf{x}_N \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{b}} \\ \mathbf{0}_{n-m} \end{bmatrix},$$

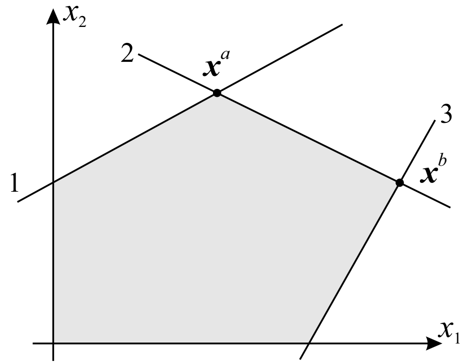
dove

$$\widehat{\mathbf{b}} = \mathbf{A}_B^{-1} \mathbf{b} \geq \mathbf{0}_m.$$

Teorema. Una soluzione basica ammissibile corrisponde a un punto estremo della regione di ammissibilità.

Consideriamo un poliedro in $\mathbb{R}^2$, caratterizzato dai vincoli

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 &\leq b_1 \\
a_{21}x_1 + a_{22}x_2 &\leq b_2 \\
a_{31}x_1 + a_{32}x_2 &\leq b_3 \\
x_1, x_2 &\geq 0.
\end{aligned}
$$



Per ricondurci al teorema, dobbiamo riscrivere i vincoli in forma standard mediante l'introduzione delle variabili di scarto (*slack*) non negative s_1, s_2, s_3:

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + s_1 &= b_1 \\
a_{21}x_1 + a_{22}x_2 + s_2 &= b_2 \\
a_{31}x_1 + a_{32}x_2 + s_3 &= b_3 \\
x_1, x_2, s_1, s_2, s_3 &\geq 0.
\end{aligned}
$$

Abbiamo quindi due variabili strutturali e tre di scarto. Ai due punti $\mathbf{x}^a$ e $\mathbf{x}^b$ corrispondono le soluzioni basiche ammissibili

$$
\mathbf{x}^a = \begin{bmatrix} x_1^a \\ x_2^a \\ 0 \\ 0 \\ s_3^a \end{bmatrix}, \qquad \mathbf{x}^b = \begin{bmatrix} x_1^b \\ x_2^b \\ s_1^b \\ 0 \\ 0 \end{bmatrix}.
$$

In tutti i vertici abbiamo tre variabili in base e due fuori. Queste possono anche essere variabili strutturali, che a loro volta rappresentano degli slack rispetto ai due assi coordinati (associati ai vincoli di non negatività).

I punti $\mathbf{x}^a$ e $\mathbf{x}^b$ sono due vertici adiacenti del poliedro, e le due basi ammissibili corrispondenti differiscono per un singolo elemento. Dato un vertice del poliedro, possiamo spostarci in un vertice adiacente portando una variabile fuori base e facendone entrare un'altra.

In generale potremmo avere casi diversi da quello che abbiamo appena considerato:

Vincoli ridondanti. Eventuali vincoli ridondanti vengono individuati ed eliminati da apposite procedure di *presolve*.

Poliedri non limitati. Sappiamo che per descrivere un poliedro illimitato occorre considerare le sue direzioni estreme. Dal punto di vista dell'ottimizzazione questo è utile solo quando la soluzione non è finita.

Poliedri vuoti. Si ottiene un poliedro vuoto quando i vincoli sono incompatibili. Il metodo del simplesso è in grado di rilevare il problema.

Il caso degenere. A ogni soluzione basica ammissibile corrisponde un punto estremo, e a ogni vertice corrisponde una soluzione basica ammissibile, ma non è detto che questa sia unica. Può infatti accadere che solo $k < m$ variabili basiche assumano un valore strettamente positivo.

Il primo algoritmo in grado di risolvere in modo efficiente problemi LP fu quello del simplesso, proposto nel 1947 da George Dantzig. Un qualsiasi algoritmo di soluzione deve essere in grado di trattare tutti i casi possibili:

1. Il problema ha un ottimo finito (non necessariamente unico).

2. La regione di ammissibilità è vuota e non esistono soluzioni ammissibili (problema *infeasible*.

3. Non esiste un ottimo finito e il valore della funzione obiettivo va a infinito, con un segno che dipende dal verso dell'ottimizzazione (problema *unbounded*).

Infine, dobbiamo ricordare che ogni algoritmo di ottimizzazione vincolata deve in qualche modo affrontare il problema di trovare una soluzione iniziale da cui partire.

Il metodo del simplesso primale si applica a un problema LP in forma standard

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

dove $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{m \times n}$ e $\mathbf{b} \in \mathbb{R}^m$, con $m < n$.

Supponiamo per semplicità che la matrice tecnologica $\mathbf{A}$ abbia rango massimo. Occorre muoversi da una base ammissibile all'altra, in modo da migliorare la funzione obiettivo fino a trovare una soluzione ottima.

Supponiamo che sia disponibile una soluzione basica ammissibile, corrispondente a una base ammissibile $\mathcal{B}$. Possiamo riscrivere il problema come

$$\begin{aligned} \text{min} \quad & \mathbf{c}_B^{\mathsf{T}}\mathbf{x}_B + \mathbf{c}_N^{\mathsf{T}}\mathbf{x}_N \\ \text{s.t.} \quad & \mathbf{A}_B\mathbf{x}_B + \mathbf{A}_N\mathbf{x}_N = \mathbf{b} \\ & \mathbf{x}_B \geq \mathbf{0}_m,\ \mathbf{x}_N \geq \mathbf{0}_{n-m}, \end{aligned}$$

dove $\mathbf{A}_B \in \mathbb{R}^{m\times m}$ è una matrice quadrata e invertibile, le cui m colonne corrispondono alle variabili in base; le rimanenti $n-m$ colonne sono raccolte nella matrice $\mathbf{A}_N \in \mathbb{R}^{m\times(n-m)}$. In modo simile, partizioniamo il vettore $\mathbf{c}$ in $\mathbf{c}_B \in \mathbb{R}^m$ e $\mathbf{c}_N \in \mathbb{R}^{n-m}$. Data una base ammissibile, la soluzione basica corrispondente è ottenuta ponendo a zero le variabili non basiche e risolvendo un sistema di equazioni lineari rispetto alle altre,

$$\mathbf{x}_B = \mathbf{A}_B^{-1}\mathbf{b} = \widehat{\mathbf{b}} \geq \mathbf{0}_m, \qquad \mathbf{x}_N = \mathbf{0}_{n-m},$$

e ha costo

$$\widehat{f} = \left[\mathbf{c}_B^{\mathsf{T}}\ \mathbf{c}_N^{\mathsf{T}}\right] \begin{bmatrix} \widehat{\mathbf{b}} \\ \mathbf{0}_{n-m} \end{bmatrix} = \mathbf{c}_B^{\mathsf{T}}\widehat{\mathbf{b}}. \tag{2}$$

Esprimiamo le variabili basiche in funzione delle variabili non basiche,

$$\mathbf{x}_B = \mathbf{A}_B^{-1}(\mathbf{b} - \mathbf{A}_N\mathbf{x}_N) = \widehat{\mathbf{b}} - \mathbf{A}_B^{-1}\mathbf{A}_N\mathbf{x}_N, \tag{3}$$

e riscriviamo la funzione obiettivo in funzione delle sole variabili fuori base:

$$\begin{aligned}\mathbf{c}^{\mathsf{T}}\mathbf{x} = \mathbf{c}_B^{\mathsf{T}}\mathbf{x}_B + \mathbf{c}_N^{\mathsf{T}}\mathbf{x}_N &= \mathbf{c}_B^{\mathsf{T}}\left(\widehat{\mathbf{b}} - \mathbf{A}_B^{-1}\mathbf{A}_N\mathbf{x}_N\right) + \mathbf{c}_N^{\mathsf{T}}\mathbf{x}_N \\ &= \mathbf{c}_B^{\mathsf{T}}\widehat{\mathbf{b}} + \left(\mathbf{c}_N^{\mathsf{T}} - \mathbf{c}_B^{\mathsf{T}}\mathbf{A}_B^{-1}\mathbf{A}_N\right)\mathbf{x}_N = \widehat{f} + \widehat{\mathbf{c}}_N^{\mathsf{T}}\mathbf{x}_N.\end{aligned}$$

Le quantità

$$\widehat{\mathbf{c}}_N^{\mathsf{T}} = \mathbf{c}_N^{\mathsf{T}} - \mathbf{c}_B^{\mathsf{T}}\mathbf{A}_B^{-1}\mathbf{A}_N \tag{4}$$

sono dette **costi ridotti** e misurano la variazione marginale del costo rispetto alle variabili fuori base, tenendo conto di come le variabili in base variano per non violare i vincoli di uguaglianza.

Se $\widehat{\mathbf{c}}_N \geq \mathbf{0}_{n-m}$, allora non è possibile ridurre il costo portando una nuova variabile in base. Se applichiamo la formula dell'equazione (4) alle variabili in base troviamo

$$\widehat{\mathbf{c}}_B^{\mathsf{T}} = \mathbf{c}_B^{\mathsf{T}} - \mathbf{c}_B^{\mathsf{T}}\mathbf{A}_B^{-1}\mathbf{A}_B = \mathbf{0}_m.$$

Possiamo quindi esprimere la seguente **condizione di ottimalità** per il simplesso primale:

$$\mathbf{c}^{\mathsf{T}} - \mathbf{c}_B^{\mathsf{T}}\mathbf{A}_B^{-1}\mathbf{A} \geq \mathbf{0}. \tag{5}$$

Se, al contrario, esiste $q \notin \mathcal{B}$ per cui $\widehat{c}_q < 0$, allora possiamo ridurre il costo portando x_q in base.

Una strategia naïve per scegliere q è trovare il costo ridotto minimo:

$$\widehat{c}_q = \min_{j \notin \mathcal{B}} \widehat{c}_j. \tag{6}$$

In realtà tale strategia, nota come regola di Dantzig, non è quella impiegata nelle implementazioni reali del metodo del simplesso.

- Il miglior tasso di riduzione del costo non tiene conto di quella che sarà la riduzione effettiva. Il passo viene implicitamente determinato dal fatto che le variabili sono tutte ristrette in segno.

- Se l'introduzione in base di una variabile fuori base ha l'effetto di ridurre il valore di una o più variabili della base corrente, questo limita il passo.

- Per un problema a grandi dimensioni non va sottovalutato lo sforzo necessario a individuare la variabile con il costo ridotto minimo.

- Un'altra difficoltà, nel caso di vertici degeneri, è la possibilità di creare uno ciclo che impedisce al metodo di convergere. Esistono strategie, che omettiamo di discutere, per evitare il problema, che comunque può rallentare i tempi di esecuzione.

Quando la variabile non basica x_q entra nella base, una variabile basica dovrebbe "lasciare" la base, preservando la condizione $\mathbf{Ax} = \mathbf{b}$. Data la base corrente, esprimiamo $\mathbf{b}$ e la colonna $\mathbf{A}_q$ che corrisponde alla variabile entrante come

$$\mathbf{b} = \sum_{i=1}^{m} x_{B(i)} \mathbf{A}_{B(i)} \tag{7}$$

$$\mathbf{A}_q = \sum_{i=1}^{m} d_i \mathbf{A}_{B(i)}, \tag{8}$$

dove $B(i)$ è l'indice della i-esima variabile basica ($i = 1, \ldots, m$), $\mathbf{A}_{B(i)}$ la colonna corrispondente, e d_i è una componente del vettore

$$\mathbf{d} = \mathbf{A}_B^{-1} \mathbf{A}_q. \tag{9}$$

Moltiplicando l'equazione (8) per un numero θ e sottraendola dalla (7), si ottiene

$$\mathbf{b} = \sum_{i=1}^{m} \left(x_{B(i)} - \theta d_i \right) \mathbf{A}_{B(i)} + \theta \mathbf{A}_q. \tag{10}$$

Dalla (10) si vede che $\theta \geq 0$ è il valore della variabile entrante nella nuova soluzione, e i valori delle variabili correnti nella base ne sono influenzate in un modo che dipende dal segno di d_i.

Se $d_i \leq 0$, $x_{B(i)}$ rimane non negativa al crescere di x_q. In questo caso non c'è un limite all'incremento della variabile entrante e il problema è unbounded.

Se invece c'è un indice i per cui $d_i > 0$, allora non possiamo fare crescere x_q illimitatamente, e il limite è dato dal valore di θ per cui una variabile nella base corrente scende a zero:

$$\theta = \frac{\widehat{b}_i}{d_i},$$

per $d_i > 0$. Per trovare il valore x_q assunto dalla variabile entrante e quale variabile esce dalla base, consideriamo il più piccolo tra i valori limite di θ:

$$x_q = \min_{\substack{i=1,\ldots,m \\ d_i>0}} \frac{\widehat{b}_i}{d_i} \tag{11}$$

Si introduce un vettore di variabili artificiali $\mathbf{z} \in \mathbb{R}^m$ e si scrivono i vincoli come

$$\begin{aligned} \mathbf{Ax} + \mathbf{z} &= \mathbf{b} \\ \mathbf{x}, \mathbf{z} &\geq \mathbf{0}. \end{aligned} \tag{12}$$

Assumiamo di avere riscritto i vincoli in modo che $\mathbf{b} \geq \mathbf{0}_m$.

Per questo sistema, è facile trovare una soluzione basica ammissibile:

$$\mathbf{x} = \mathbf{0}_n, \quad \mathbf{z} = \mathbf{b}.$$

Minimizziamo una funzione ausiliaria ϕ, detta forma di inammissibilità, e risolviamo il problema

$$\min \phi = \sum_{i=1}^{m} z_i \tag{13}$$

usando l'algoritmo del simplesso. Se il valore ottimo di (13) è $\phi^* = 0$, abbiamo trovato una base ammissibile per il problema originale.

Consideriamo il problema numerico di mix ottimo introdotto in precedenza:

$$\begin{aligned}
\max \quad & 45x_1 + 60x_2 \\
\text{s.t.} \quad & 15x_1 + 35x_2 \leq 2400 \\
& 25x_1 + 15x_2 \leq 2400 \\
& x_1 \geq 0, \quad 0 \leq x_2 \leq 50
\end{aligned} \tag{14}$$

Il problema viene riscritto in forma standard introducendo le tre variabili di slack $s_1, s_2, s_3 \geq 0$:

$$\begin{aligned}
\min \quad & -45x_1 - 60x_2 \\
\text{s.t.} \quad & 15x_1 + 35x_2 + s_1 = 2400 \\
& 25x_1 + 15x_2 + s_2 = 2400 \\
& x_2 + s_3 = 50 \\
& x_1, x_2, s_1, s_2, s_3 \geq 0
\end{aligned}$$

In forma matriciale abbiamo:

$$\mathbf{A} = \begin{bmatrix} 15 & 35 & 1 & 0 & 0 \\ 25 & 15 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} -45 \\ -60 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 2400 \\ 2400 \\ 50 \end{bmatrix}.$$

Passo 1

In questo caso è facile trovare una base ammissibile di partenza, $\{s_1, s_2, s_3\}$, a cui corrisponde $x_1 = x_2 = 0$. La matrice associata alla base è semplicemente $\mathbf{A}_B = \mathbf{I}_3$:

$$\mathbf{A}_N = \begin{bmatrix} 15 & 35 \\ 25 & 15 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{c}_N = \begin{bmatrix} -45 \\ -60 \end{bmatrix}, \quad \mathbf{c}_B = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \widehat{\mathbf{b}} = \begin{bmatrix} 2400 \\ 2400 \\ 50 \end{bmatrix}.$$

Inoltre, il valore corrente della funzione obiettivo è

$$\mathbf{c}_B^\mathsf{T} \widehat{\mathbf{b}} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2400 \\ 2400 \\ 50 \end{bmatrix} = 0.$$

Usando l'equazione (4), troviamo i costi ridotti

$$\widehat{\mathbf{c}}_N^\mathsf{T} = \mathbf{c}_N^\mathsf{T} - \mathbf{c}_B^\mathsf{T} \mathbf{A}_B^{-1} \mathbf{A}_N = \begin{bmatrix} -45 & -60 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 15 & 35 \\ 25 & 15 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -45 & -60 \end{bmatrix},$$

associati alle due variabili attualmente fuori base, x_1 e x_2.

Per migliorare la soluzione, portiamo in base la variabile x_2, a cui è associato il costo ridotto più negativo.

Dobbiamo ora individuare quale variabile esce dalla base. L'equazione (9) fornisce

$$\mathbf{d} = \mathbf{A}_B^{-1} \mathbf{a}_2 = \mathbf{I}_3 \begin{bmatrix} 35 \\ 15 \\ 1 \end{bmatrix}.$$

Tutti gli elementi del vettore $\mathbf{d}$ sono strettamente positivi e, applicando l'equazione (11), troviamo

$$\theta = \min\left\{\frac{2400}{35}, \frac{2400}{15}, \frac{50}{1}\right\} = \min\{68.5714, 160, 50\} = 50.$$

Tra le variabili attualmente in base, esce la terza variabile, che corrisponde allo slack s_3 del vincolo di mercato su x_2, che entra al suo limite superiore pari a 50.

Passo 2

La base corrente è $\{x_2, s_1, s_2\}$, a cui corrisponde il piano $x_1 = 0$ e $x_2 = 50$, di profitto 3000. Quindi abbiamo

$$\mathbf{A}_B = \begin{bmatrix} 35 & 1 & 0 \\ 15 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{A}_N = \begin{bmatrix} 15 & 0 \\ 25 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{c}_B = \begin{bmatrix} -60 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{c}_N = \begin{bmatrix} -45 \\ 0 \end{bmatrix}.$$

Inoltre

$$\widehat{\mathbf{b}} = \begin{bmatrix} 35 & 1 & 0 \\ 15 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 2400 \\ 2400 \\ 50 \end{bmatrix} = \begin{bmatrix} 50 \\ 650 \\ 1650 \end{bmatrix}.$$

Calcoliamo i costi ridotti

$$\widehat{\mathbf{c}}_N^{\mathsf{T}} = \mathbf{c}_N^{\mathsf{T}} - \mathbf{c}_B^{\mathsf{T}} \mathbf{A}_B^{-1} \mathbf{A}_N = \begin{bmatrix} -45 & 60 \end{bmatrix}.$$

Dobbiamo portare x_1 in base. L'equazione (9) fornisce in questo caso

$$\mathbf{d} = \mathbf{A}_B^{-1}\mathbf{a}_1 = \begin{bmatrix} 35 & 1 & 0 \\ 15 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 15 \\ 25 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 15 \\ 25 \end{bmatrix}$$

Dobbiamo applicare l'equazione (11) solo a due termini, e troviamo

$$\theta = \min\left\{\frac{650}{15}, \frac{1650}{25}\right\} = \min\{43.33, 66\} = 43.33.$$

Quindi, la variabile x_1 entra in base al valore 43.33, ed esce la seconda variabile nella base corrente, ovvero s_1.

Passo 3

La base corrente è $\{x_1, x_2, s_2\}$. Il mix corrente satura il mercato per il prodotto P2 e sfrutta al massimo la prima risorsa critica, lasciando slack sulla seconda. Il suo contributo al profitto è 4950 euro. A questo punto abbiamo

$$\mathbf{A}_B = \begin{bmatrix} 15 & 35 & 0 \\ 25 & 15 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{A}_N = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{c}_B = \begin{bmatrix} -45 \\ -60 \\ 0 \end{bmatrix}, \quad \mathbf{c}_N = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \widehat{\mathbf{b}} = \begin{bmatrix} 43.33 \\ 50 \\ 566.67 \end{bmatrix}.$$

I costi ridotti sono

$$\widehat{\mathbf{c}}_N^{\mathsf{T}} = \begin{bmatrix} 3 \\ -45 \end{bmatrix},$$

da cui concludiamo che dobbiamo (ri)portare s_3 in base. Procedendo in modo analogo a quanto fatto in precedenza troviamo

$$\mathbf{d} = \begin{bmatrix} -2.33 \\ 1 \\ 43.33 \end{bmatrix}, \qquad \theta = \min\left\{\frac{50}{1}, \frac{566.67}{43.33}\right\} = \min\{50, 13.0769\} = 13.0769.$$

Quindi, la variabile s_3 entra in base al valore 13.0769, ed esce la terza variabile nella base corrente, ovvero s_2.

Passo 4

Per la nuova base $\{x_1, x_2, s_3\}$ abbiamo

$$\mathbf{A}_B = \begin{bmatrix} 15 & 35 & 0 \\ 25 & 15 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \quad \mathbf{A}_N = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{x}_B = \mathbf{A}_B^{-1}\mathbf{b} = \begin{bmatrix} 73.8462 \\ 36.9231 \\ 13.0769 \end{bmatrix}.$$

I costi ridotti per s_1 e s_2 sono

$$\widehat{\mathbf{c}}_N^{\mathsf{T}} = \begin{bmatrix} 1.2692 & 1.0385 \end{bmatrix}. \tag{15}$$

Poiché entrambi i costi ridotti sono positivi, la base corrente è ottima. Il mix ottimo è $(73.8462, 36.9231)$, che contribuisce 5538.46 euro al profitto.

Applicando i concetti di dualità Lagrangiana alla programmazione lineare, abbiamo visto che il duale del problema in forma canonica,

$$(P_1) \qquad \begin{array}{ll} \text{min} & \mathbf{c}^\mathsf{T}\mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} \geq \mathbf{b}, \end{array}$$

è dato dal problema LP

$$(D_1) \qquad \begin{array}{ll} \text{max} & \mathbf{b}^\mathsf{T}\boldsymbol{\mu} \\ \text{s.t.} & \mathbf{A}^\mathsf{T}\boldsymbol{\mu} = \mathbf{c}, \\ & \boldsymbol{\mu} \geq \mathbf{0}. \end{array}$$

Applicando lo stesso tipo di ragionamento, è facile vedere che il duale del problema in forma standard,

$$(P_2) \qquad \begin{array}{ll} \text{min} & \mathbf{c}^\mathsf{T}\mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}, \end{array} \tag{16}$$

è

$$(D_2) \qquad \begin{array}{ll} \text{max} & \mathbf{b}^\mathsf{T}\boldsymbol{\mu} \\ \text{s.t.} & \mathbf{A}^\mathsf{T}\boldsymbol{\mu} \leq \mathbf{c}. \end{array} \tag{17}$$

Usando lo stesso tipo di ragionamento, possiamo costruire il duale di un problema LP in forma arbitraria. La tabella sintetizza le "regole" per la formazione di un duale, dove indichiamo con $\mathbf{A}_j$ la colonna j-esima della matrice $\mathbf{A}$, e con $\mathbf{a}_i^\mathsf{T}$ la riga i-esima.

Primale		Duale	
min	$\mathbf{c}^\mathsf{T}\mathbf{x}$	max	$\mathbf{b}^\mathsf{T}\boldsymbol{\mu}$
s.t.	$\mathbf{a}_i^\mathsf{T}\mathbf{x} \geq b_i$	s.t.	$\mu_i \geq 0$
	$\mathbf{a}_i^\mathsf{T}\mathbf{x} \leq b_i$		$\mu_i \leq 0$
	$\mathbf{a}_i^\mathsf{T}\mathbf{x} = b_i$		μ_i libera
	$x_j \geq 0$		$\mathbf{A}_j^\mathsf{T}\boldsymbol{\mu} \leq c_j$
	$x_j \leq 0$		$\mathbf{A}_j^\mathsf{T}\boldsymbol{\mu} \geq c_j$
	x_j libera		$\mathbf{A}_j^\mathsf{T}\boldsymbol{\mu} = c_j$

Esempio. Costruiamo il duale del problema di mix ottimo (14), scritto come

$$\begin{aligned} \min \quad & -45x_1 - 60x_2 \\ \text{s.t.} \quad & 15x_1 + 35x_2 \leq 2400 \\ & 25x_1 + 15x_2 \leq 2400 \\ & x_2 \leq 50 \\ & x_1, x_2 \geq 0. \end{aligned}$$

Associamo ai tre vincoli di disuguaglianza dei moltiplicatori q_1, q_2 e q_3, e formiamo il duale

$$\begin{aligned} \max \quad & 2400q_1 + 2400q_2 + 50q_3 \\ \text{s.t.} \quad & 15q_1 + 25q_2 \leq -45 \\ & 35q_1 + 15q_2 + q_3 \leq -60 \\ & q_1, q_2, q_3 \leq 0. \end{aligned}$$

Se introduciamo nuove variabili duali $\mu_i = -q_i$, $i = 1, 2, 3$, si trova un duale in forma più naturale:

$$\begin{aligned} \min \quad & 2400\mu_1 + 2400\mu_2 + 50\mu_3 \\ \text{s.t.} \quad & 15\mu_1 + 25\mu_2 \geq 45 \\ & 35\mu_1 + 15\mu_2 + \mu_3 \geq 60 \\ & \mu_1, \mu_2, \mu_3 \geq 0. \end{aligned}$$

Troviamo un valore di funzione obiettivo pari a 5538.46, che coincide con il profitto ottimo trovato risolvendo il primale (a meno del costo fisso pari a 5000 euro).

Inoltre la soluzione duale ottima è:

$$\mu_1^* = 1.2692, \qquad \mu_2^* = 1.0385, \qquad \mu_3^* = 0.$$

Le prime due variabili duali sono i prezzi ombra delle due risorse critiche, mentre la terza è nulla perché il vincolo sulla massima domanda del secondo prodotto non è attivo nel mix ottimo.

Teorema: dualità forte nel caso LP. Consideriamo la seguente coppia primale-duale:

$$\left(\begin{array}{ll} \text{min} & \mathbf{c}^{\mathsf{T}}\mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} \geq \mathbf{b} \end{array}\right) \qquad \Leftrightarrow \qquad \left(\begin{array}{ll} \text{max} & \mathbf{b}^{\mathsf{T}}\boldsymbol{\mu} \\ \text{s.t.} & \mathbf{A}^{\mathsf{T}}\boldsymbol{\mu} = \mathbf{c} \\ & \boldsymbol{\mu} \geq \mathbf{0}. \end{array}\right)$$

Se il primale ha una soluzione ottima finita $\mathbf{x}^*$, allora anche il duale ha una soluzione ottima finita $\boldsymbol{\mu}^*$, e vale la condizione

$$\mathbf{c}^{\mathsf{T}}\mathbf{x}^* = \mathbf{b}^{\mathsf{T}}\boldsymbol{\mu}^*. \tag{18}$$

Il teorema esclude la presenza di gap di dualità tra primale e duale, e quindi se uno dei due problemi ha ottimo finito, deve valere la condizione

$$\mathbf{b}^{\mathsf{T}}\boldsymbol{\mu}^* = \mathbf{c}^{\mathsf{T}}\mathbf{x}^*.$$

Tuttavia, sono possibili altri casi.

- Un problema primale di minimizzazione potrebbe essere illimitato inferiormente. In questo caso il problema duale non può avere alcuna soluzione ammissibile.

- Una situazione analoga vale se un problema di duale di massimizzazione è illimitato superiormente. Il primale non può essere ammissibile.

- Esistono poi anche casi in cui sia il primale che il duale non sono ammissibili.

Il simplesso duale

L'algoritmo del simplesso duale è un'alternativa al simplesso primale, ed è disponibile nelle librerie di solver commerciali.

L'idea è applicare l'algoritmo del simplesso al problema duale. La condizione degli scarti complementari permette un recupero agevole dei valori ottimi per le variabili primali.

Il simplesso duale è spesso più efficiente di quello primale. Questo è possibile perché la fase 1 del duale può essere più veloce di quella per il primale, o perché il simplesso duale può soffrire meno di problemi legati a vertici degeneri.

Una situazione in cui certamente è fondamentale applicare il simplesso duale è quando si ha già una soluzione ottima, e si aggiunge al primale un vincolo che la rende inammissibile.

Il caso praticamente più rilevante si verifica negli algoritmi di enumerazione implicita del tipo branch and bound/cut alla base dei solver per modelli MILP.

Consideriamo un modello LP in forma canonica

$$\begin{aligned} \min \quad & \mathbf{c}^\mathsf{T}\mathbf{x} \\ \text{s.t.} \quad & \mathbf{a}_i^\mathsf{T}\mathbf{x} \le b_i, \qquad i \in [m], \end{aligned}$$

e supponiamo che vi sia incertezza sui vettori $\mathbf{a}_i$. Una possibile forma di uncertainty set è un poliedro limitato, per cui abbiamo

$$\mathcal{P}_i = \{\mathbf{a}_i \mid \mathbf{C}_i\mathbf{a}_i \le \mathbf{d}_i\}, \qquad i \in [m],$$

dove $\mathbf{C}_i \in \mathbb{R}^{m_i \times n}$ e $\mathbf{d}_i \in \mathbb{R}^{m_i}$.

Vogliamo una soluzione $\mathbf{x}$ che sia ammissibile per tutti i valori possibili di $\mathbf{a}_i$, $i \in [m]$. Dovremmo quindi affrontare un problema del tipo

$$\begin{aligned} \min \quad & \mathbf{c}^\mathsf{T}\mathbf{x} \\ \text{s.t.} \quad & \mathbf{a}_i^\mathsf{T}\mathbf{x} \le b_i, \qquad i \in [m], \ \mathbf{a}_i \in \mathcal{P}_i. \end{aligned}$$

Si tratta di un problema di programmazione semi-infinita, con un numero finito di variabili decisionali, ma un numero infinito di vincoli.

Possiamo sfruttare la teoria della dualità per trasformarlo in un problema LP deterministico e trattabile. Come primo passo, riscriviamo il modello come

$$\begin{aligned} \min \quad & \mathbf{c}^\mathsf{T}\mathbf{x} \\ \text{s.t.} \quad & \max_{\mathbf{a}_i \in \mathcal{P}_i} \mathbf{a}_i^\mathsf{T}\mathbf{x} \le b_i, \qquad i \in [m]. \end{aligned} \tag{19}$$

Il secondo passo sta nel riconoscere che possiamo analizzare ogni vincolo separatamente. Infine, sfruttiamo la dualità per riscrivere il sottoproblema definito sulle variabili $\mathbf{a}_i$ come segue:

$$\begin{array}{rl} \max & \mathbf{x}^{\mathsf{T}}\mathbf{a}_i \\ \text{s.t.} & \mathbf{C}_i\mathbf{a}_i \le \mathbf{d}_i \end{array} \qquad \Rightarrow \qquad \begin{array}{rl} \min & \mathbf{d}_i^{\mathsf{T}}\mathbf{z}_i \\ \text{s.t.} & \mathbf{C}_i^{\mathsf{T}}\mathbf{z}_i = \mathbf{x} \\ & \mathbf{z}_i \ge \mathbf{0}, \end{array}$$

definito rispetto a variabili duali $\mathbf{z}_i \in \mathbb{R}^{m_i}$.

Data la relazione di dualità, è sufficiente trovare una soluzione *ammissibile* $\mathbf{z}_i$ per il duale, per cui l'obiettivo del duale non è maggiore di b_i. Questo garantisce che l'ottimo del problema primale rispetto ai coefficienti $\mathbf{a}_i$ non eccede il valore limite b_i.

Pertanto, la soluzione $\mathbf{x}$ è ammissibile nel senso robusto se esiste un vettore $\mathbf{z}_i$ tale che

$$\mathbf{d}_i^{\mathsf{T}}\mathbf{z}_i \le b_i, \qquad \mathbf{C}_i^{\mathsf{T}}\mathbf{z}_i = \mathbf{x}, \qquad \mathbf{z}_i \ge \mathbf{0}.$$

A questo punto, possiamo riformulare il problema LP robusto come

$$\begin{array}{rll} \min & \mathbf{c}^{\mathsf{T}}\mathbf{x} & \\ \text{s.t.} & \mathbf{d}_i^{\mathsf{T}}\mathbf{z}_i \le b_i, & i \in [m] \\ & \mathbf{C}_i^{\mathsf{T}}\mathbf{z}_i = \mathbf{x}, & i \in [m] \\ & \mathbf{z}_i \ge \mathbf{0}, & i \in [m], \end{array}$$

rispetto ai vettori di variabili $\mathbf{x}$ e $\mathbf{z}_i$, $i \in [m]$.

Nel capitolo 4 abbiamo osservato un legame tra opportunità di arbitraggio, in un modello di mercato finanziario con spazio degli stati discreto, e lemma di Farkas. Possiamo collegare tutto ciò alla dualità LP.

Consideriamo n asset, il cui prezzo adesso, al tempo $t = 0$, è pari a S_{i0}, $i \in [n]$. In futuro, al tempo $t = T$, i prezzi sono incerti. Abbiamo m scenari, per ognuno dei quali abbiamo i prezzi S_{iT}^j, $i \in [n]$, $j \in [m]$.

Indichiamo con h_i le unità di asset di tipo i detenute (*holding*), e cerchiamo un portafoglio di valore corrente minimo, sotto il vincolo che il portafoglio non possa avere valore negativo in futuro:

$$\min \quad \sum_{i \in [n]} S_{i0} h_i \tag{20}$$

$$\text{s.t.} \quad \sum_{i \in [n]} S_{iT}^j h_i \geq 0, \qquad j \in [m]. \tag{21}$$

Poniamoci le seguenti domande:

1. È possibile che il valore del portafoglio ottimo sia strettamente positivo?

2. È possibile che il valore del portafoglio ottimo sia strettamente negativo e finito?

La risposta alla prima domanda è negativa, perché il portafoglio nullo $h_i = 0$ è ammissibile. Anche la risposta alla seconda domanda è negativa, perché se esiste una portafoglio ammissibile per cui il valore corrente è negativo, possiamo moltiplicarlo per un fattore moltiplicativo arbitrario $\alpha > 0$ e ottenere una soluzione sempre ammissibile e con un profitto migliore.

Ci sono solo due possibilità: o il portafoglio ottimo ha valore zero, e quindi non esistono opportunità di arbitraggio, oppure esso ha valore illimitato, ed esiste una opportunità di arbitraggio.

In questo secondo caso, sappiamo dalla teoria della dualità che il duale del problema LP (20)–(21) non può essere ammissibile. Scriviamo il duale, associando una variabile duale $q_j \geq 0$ a ogni vincolo (21):

$$\begin{aligned} \max \quad & \sum_{j \in [m]} 0 \cdot q_j && (22) \\ \text{s.t.} \quad & \sum_{j \in [m]} S_{iT}^{j} q_j = S_{i0}, \qquad i \in [n] && (23) \\ & q_j \geq 0, \qquad j \in [m]. && (24) \end{aligned}$$

Se il duale ha valore finito, il suo valore non può essere diverso dall'ottimo del primale, che è pari a 0 se non ci sono opportunità di arbitraggio.

I vincoli (23) esprimono il prezzo corrente di ogni asset come combinazione conica dei prezzi futuri. Le variabili duali q_j costituiscono un funzionale di pricing non negativo, la cui esistenza esclude opportunità di arbitraggio. Se tale funzionale non esiste, allora il duale non è ammissibile e il primale ha ottimo non limitato.