Appunti Ricerca Operativa

Luca Seggiani
2024

1 Lezione del 23-09-24

1.1 Introduzione

1.1.1 Programma del corso

Il corso di ricerca operativa si divide in 4 parti:

1. Modello di Programmazione Lineare;
2. Programmazione Lineare su reti, ergo programmazione lineare su grafi;
3. Programmazione Lineare intera, ergo programmazione lineare col vincolo $x \in \mathbb{Z}^n$;
4. Programmazione Non Lineare.

Le prime 3 parti hanno come prerequisiti l'algebra lineare: in particolare operazioni matriciali, prodotti scalari, sistemi lineari, teorema di Rouché-Capelli. La quarta parte richiede invece conoscenze di Analisi II.

1.1.2 Un problema di programmazione lineare

La ricerca operativa si occupa di risolvere problemi di ottimizzazione con variabili decisionali e risorse limitate. Poniamo un problema di esempio:

Problema 1.1: Produzione

Una ditta produce due prodotti: **laminato A** e **laminato B**. Ogni prodotto deve passare attraverso diversi reparti: il reparto **materie prime**, il reparto **taglio**, il reparto **finiture A** e il reparto **finiture B**. Il guadagno è rispettivamente di 8.4 e 11.2 (unità di misura irrilevante) per ogni tipo di laminato.
Ora, nel reparto materie prime, il laminato A occupa 30, ore, e lo B 20 ore. Nel reparto taglio il laminato A occupa 10 ore e lo B 20 ore. Il laminato A occupa poi 20 ore nel reparto finiture A, mentre il laminato B occupa 30 ore nel reparto finiture B. I reparti hanno a disposizione, rispettivamente, 120, 80, 62 e 105 ore.
Possiamo porre queste informazioni in forma tabulare:

Reparto	**Capienza**	**Laminato A**	**Laminato B**
Materie prime	120	30	20
Taglio	80	10	20
Finiture A	62	20	/
Finiture B	105	/	30
Guadagno		8.4	11.2

Quello che ci interessa è chiaramente massimizzare il guadagno.

Decidiamo di modellizzare questa situazione con un modello matematico.

Il guadagno che abbiamo dai laminati rappresenta una **funzione obiettivo**, ovvero la funzione che vogliamo ottimizzare. Ottimizzare significa trovare il modo migliore di massimizzare o minimizzare i valori della funzione agendo sulle variabili decisionali.

La funzione obiettivo va ottimizzata rispettando determinati **vincoli**, che modellizzano il fatto che le risorse sono limitate. Una **soluzione ammissibile** è una qualsiasi soluzione che rispetta i vincoli del problema. Chiamiamo quindi **regione ammissibile** l'insieme di tutte le soluzioni ammissibili. All'interno della regione ammissibile c'è la soluzione che cerchiamo, ovvero la **soluzione ottima**.

Decidiamo quindi le **variabili decisionali**, ed esplicitiamo la funzione obiettivo e i vincoli.

In questo caso le variabili decisionali saranno le quantità di laminato A e B da produrre, che individuano un punto in $\mathbb{R}^2$ denominato (x_A, x_B). Decidere di usare la soluzione $(1, 1)$ significa decidere di produrre 1 unità di laminato A e 1 unità di laminato B, per un guadagno complessivo di $8.4 + 11.2 = 19.6$.

La funzione obiettivo sarà quindi:

$$f(x_A, x_B) = 8.4x_A + 11.2x_B, \quad f : \mathbb{R}^2 \to \mathbb{R}$$

lineare, e noi saremo interessati a:

$$\max(f(x_A, x_B))$$

rispettando i vincoli, ergo nella regione ammissibile. Per esprimere questi vincoli, cioè il tempo limitato all'interno di ogni reparto, introduciamo il sistema di disequazioni:

$$\begin{cases} 30x_A + 20x_B \leq 120 \\ 10x_A + 20x_B \leq 80 \\ 20x_A + 0x_B \leq 62 \\ 0x_A + 30x_B \leq 105 \\ -x_A \leq 0 \\ -x_B \leq 0 \end{cases}$$

dove notiamo le ultime due disequazioni indicano la positività di x_A e x_B, in forma $f(x_A, x_B) \leq b$. Questo sistema non indica altro che la regione ammissibile.

Possiamo riscrivere questo modello usando la notazione dell'algebra lineare. La funzione obiettiva e i vincoli diventano semplicemente:

$$\begin{cases} \max(c^T \cdot x) \\ A \cdot x \leq b \end{cases}$$

dove c rappresenta il vettore dei costi, A rappresenta la matrice dei costi a b il vettore dei vincoli. c è trasposto per indicare prodotto fra vettori.

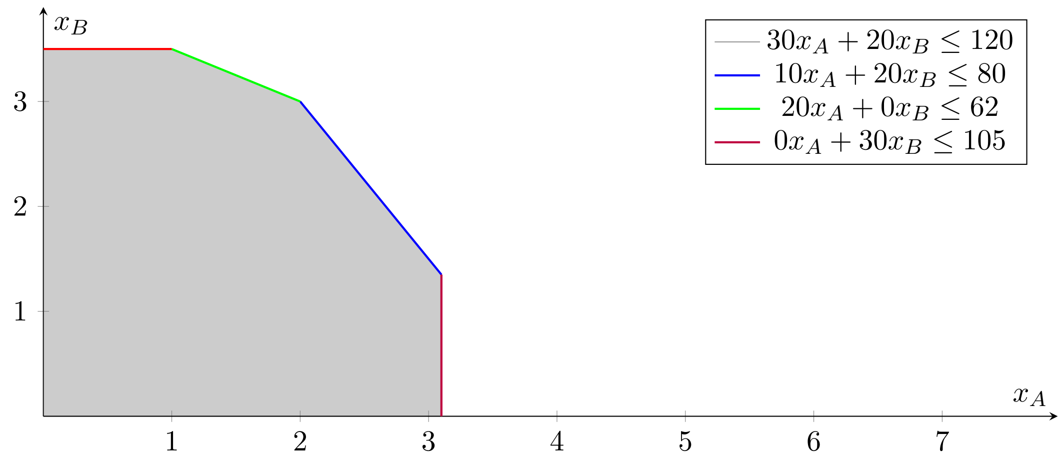
Possiamo scrivere A, b e c per esteso:

$$A : \begin{pmatrix} 30 & 20 \\ 10 & 20 \\ 20 & 0 \\ 0 & 30 \\ -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad b : \begin{pmatrix} 120 \\ 80 \\ 62 \\ 105 \\ 0 \\ 0 \end{pmatrix}, \quad c : \begin{pmatrix} 8.4 \\ 11.2 \end{pmatrix}$$

Notiamo come A e b hanno dimensione verticale $4 + 2 = 6$, dai 4 vincoli superiori e i 2 vincoli inferiori.

A questo punto, possiamo disegnare la regione ammissibile come l'intersezione dei semipiani individuati da ogni singola disuguaglianza.

Si riporta un grafico:



In diversi colori sono riportati i margini delle disequazioni, mentre in grigio è evidenziata la regione ammissibile. Qualsiasi punto all'interno della regione ammissibile vale come soluzione, e almeno uno di essi è soluzione ottimale.

Il modello finora descritto prende il nome di modello di programmazione lineare, e permette di formulare problemi di programmazione lineare (LP).

Definizione 1.1: Problema di programmazione lineare (1)

Un problema di programmazione lineare (LP) riguarda l'ottimizzazione di una funzione lineare in più variabili soggetta a vincoli di tipo $=$, $\leq$ e $\geq$, ovvero in forma:

$$\begin{cases} \min / \max(c^T \cdot x) \\ A_i x \leq b \\ B_j x \geq d \\ C_k x = e \end{cases}$$

"Programmazione" qui non ha alcun legame col concetto di programmazione informatica, ma si riferisce al fatto che il modello è effettivamente *programmabile*.

"Lineare" si riferisce alla linearità dei problemi che ci permette di risolvere (e quindi del modello).

2 Lezione del 24-09-24

2.1 Forma primale standard

Ciò che abbiamo formulato finora è un problema di programmazione lineare. Possiamo dire che la forma:

$$\begin{cases} \max(c^T \cdot x) \\ Ax \leq b \end{cases}$$

rappresenta un problema LP in forma **primale standard**, ricordando che c è il vettore dei coefficienti della funzione obiettivo, A la matrice dei coefficienti per ogni vincolo, e b il vettore dei termini noti per ogni vincolo.

Definizione 2.1: Forma primale standard

Un problema di programmazione lineare si dice in forma primale standard quando è espresso in forma:

$$\begin{cases} \max(c^T \cdot x) \\ Ax \leq b \end{cases}$$

Si adotta una forma primale standard in quanto si può trasformare ogni problema LP in una forma di questo tipo.

2.1.1 Normalizzazione di un problema LP

Un modo per portare un problema LP qualsiasi in forma primale standard è:

1. Si trasformano le disuguaglianze: $\geq \leftrightarrow \leq$
2. Si riscrivono le uguaglianze come coppie di diseguaglianze:

   $$f(x) = c \rightarrow \begin{cases} f(x) \leq c \\ f(x) \geq c \end{cases}$$

   da cui la (1):

   $$f(x) = c \rightarrow S \begin{cases} f(x) \leq c \\ -f(x) \leq -c \end{cases}$$

3. Se il problema richiede il minimo, si nota che $\max(f) = -\min(-f)$, e sopratutto:

   $$\bar{x} \in \text{argmax}(f) \Leftrightarrow \bar{x} \in \text{argmin}(-f)$$

   con $\text{argmax}(f)$ e $\text{argmin}(-f)$ rispettivamente gli insiemi dei punti di massimo e minimo. Questo significa che posso semplicemente cambiare di segno la funzione obiettivo per trovare da massimi minimi, e viceversa.

Notiamo inoltre che, nella forma primale standard, si ha:

$$x \in R^n, \quad A \in R^{n \times m}, \quad b \in R^m, \quad c \in R^n$$

2.2 Proprietà generali di un problema LP

La regione ammissibile di un problema PL si chiama **poliedro**. Si può dare agilmente una definizione algebrica di poliedro:

Definizione 2.2: Definizione algebrica di poliedro

Algebricamente, un poliedro è l'insieme delle soluzioni di un sistema di disequazioni lineari in $\mathbb{R}^n$ variabili:

$$P = \{x \in \mathbb{R}^n : Ax \leq b\}$$

Questa regione in un problema LP prende il nome di regione ammissibile.

Definizione 2.3: Definizione geometrica di poliedro

Geometricamente, un poliedro è l'intersezione di un numero finito di semispazi chiusi.

I semispazi chiaramente sono lineari, e in $\mathbb{R}^2$ rappresenterebbero semipiani. Chiusi significa che nelle disequazioni che descrivono i vincoli compargono solo $\leq$ e non $<$, ergo la regione ammissibile contiene la sua frontiera.

Possiamo dimostrare 4 proprietà dei poliedri:

1. Un'osservazione fondamentale è la seguente:

   Teorema 2.1: Soluzione ottimale di un problema LP

   La soluzione ottimale di un problema LP è contenuta nella frontiera della regione ammissibile.

   Questo si può ricavare dai teoremi di Fermat e di Weierstrass, e dalla convessità della regione ammissibile. Inanzitutto, si è stabilito che la soluzione ottimale non è altro che il massimo o minimo assoluto all'interno della regione ammissibile del problema. Il gradiente della funzione obiettiva è $\neq 0$ in ogni suo punto (funzione lineare a gradiente costante). Da Fermat, i massimi e minimi hanno sempre gradiente 0, ergo massimi o minimi locali (che esistono per Weierstrass) possono trovarsi solo sulla frontiera. A questo punto, possiamo imporre la convessità per asserire che quei punti di massimo o minimo sono anche globali.

2. Prendiamo in esempio il poliedro dato da:

$$\begin{cases} x_A > 0 \\ x_B > 0 \end{cases}$$

   o se vogliamo, in forma primale standard, dato dalle matrici A e b:

$$A : \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad b : \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

   questo poliedro non è limitato nella direzione positiva, ergo può arrivare a valori di x_A e x_B che tendono a $+\infty$. Da ciò si ha che può accadere che un problema LP ammetta soluzioni x tali che $x \to \pm\infty$, ovvero che il poliedro sia illimitato. In particolare, un poliedro limitato si dice **politopo**.

3. Notiamo poi che la soluzione di un problema LP può non essere unica. Questo accade ad esempio quando la soluzione sta su un segmento di frontiera: a quel punto tutti i punti del segmento sono soluzione. Da questo segue che:

   Teorema 2.2: Unicità della soluzione ottimale di un problema LP

   Se un problema LP ha almeno 2 soluzioni, allora ne ha infinite.

Ciò si può dimostrare come segue. Si riporta innanzitutto la notazione parametrica del segmento $\bar{zw}$, dati i due vettori di estremo z e w:

$$\lambda z + (1 - \lambda)w, \quad \lambda \in [0, 1]$$

A questo punto si pone che z e w sono entrambi soluzioni ottime, ergo:

$$\max(c^T \cdot x) = c^T z = c^T w = v$$

da cui si può dire che:

$$c^T(\lambda z + (1 - \lambda)w) = \lambda c^T z + (1 - \lambda)c^T w = \lambda v + (1 - \lambda)v = v$$

Ovvero ogni punto sul segmento porta la funzione obiettiva a massimo assoluto, quindi è soluzione ottimale.

4. Infine, notiamo che il poliedro della regione ammissible di un problema LP può essere vuoto, ergo $P = \emptyset$. In questo caso, si ha che $\max(c^T \cdot x) = -\infty$ e $\min(c^T \cdot x) = \infty$. Un poliedro vuoto significa che i vincoli stessi vanno modificati. Questo solitamente si fa risovendo una versione semplificata del problema LP.

Si può fare un'altro esempio per sottolineare l'importanza del punto di massimo (o minimo), e non quel massimo (o minimo). Finché nella funzione obiettivo i coefficienti compargono nello stesso rapporto (ergo finché si scelgono vettori c linearmente dipendenti), il punto di massimo (o minimo) non cambia, per via della linearità (e si presume omogeneità) della funzione obiettiva stessa. Sarà solo il massimo (o minimo) a variare di un rapporto pari a quello di cui variano i coefficienti.

2.3 Gradiente e linee di isocosto

Si può dimostrare il seguente teorema:

Teorema 2.3: Gradiente della funzione obiettivo

Il gradiente di una funzione obiettivo definita come $f(x) = c^T \cdot x$ sulla base di un qualche vettore c è in ogni punto il vettore c stesso.

Da questo gradiente si possono ricavare le cosiddette linee di isocosto (in dimensioni > 2 sarebbero superfici), cioè linee a valore costante della funzione obiettivo.

Definizione 2.4: Linea di isocosto

Si definisce linea di isocosto di una funzione obiettivo con vettore c una retta (o superficie):

$$f(x) = c^T \cdot x = k$$

per un qualsiasi k costante.

2.4 Cono di competenza

Dovrebbe essere chiaro adesso che i punti di soluzione ottima stanno tutti su un segmento o su un punto della frontiera. Nel caso si abbia un vettore gradiente perpendicolare ad un segmento della frontiera, quel segmento sarà soluzione ottima. In caso contrario, spostandoci a destra avremo l'estremo destro del segmento, e spostandoci a sinistra viceversa, finché non si diventerà perpendicolari a qualche altro segmento di frontiera.

Il cono (in R^2, angolo) di valori possibili del gradiente che rendono un punto ottimale prende il nome di **cono di competenza**.

Definizione 2.5: Cono di competenza

Il cono di competenza di un punto x^* è il cono, ovvero l'insieme di vettori gradiente, tale per cui il punto x^* conserva l'ottimalità sulla funzione obiettivo coi vincoli imposti.

Vedremo in seguito l'importanza di una nozione di cono per i problemi LP.

3 Lezione del 25-09-24

3.1 Assegnamento di costo minimo

Vediamo un problema:

Problema 3.1: Assegnamento

Quattro agenzie possono occuparsi di 4 progetti. Ogni agenzia presenta il costo stimato per la realizzazione di ogni progetto, in migliaia di euro. In forma tabulare, si riportano i valori:

	Agenzia 1	**Agenzia 2**	**Agenzia 3**	**Agenzia 4**
Progetto 1	20	17	16	14
Progetto 2	22	16	19	15
Progetto 3	21	17	15	23
Progetto 4	19	18	14	24

Vogliamo capire quale agenzia deve occuparsi di quale progetto per minimizzare i costi.

Con n agenzie e progetti ci sono $n!$ possibili combinazioni, ergo dobbiamo trovare un algoritmo più efficiente. Applicando il modello studiato finora, abbiamo la matrice dei costi c:

$$\begin{pmatrix} 20 & 17 & 16 & 14 \\ 22 & 16 & 19 & 15 \\ 21 & 17 & 15 & 23 \\ 19 & 18 & 14 & 24 \end{pmatrix}$$

che possiamo portare a:

$$c : (-18, +18 + ... + 24)$$

come linearizzazione lessicografica della tabella sopra riportata (notare che sarebbe un vettore colonna).

Adesso dobbiamo solo trovare un metodo per esplicitare i vincoli del problema:

- Ogni agenzia può occuparsi solo di un progetto;

- Ogni progetto richiede solo un'agenzia.

Possiamo rappresentare la corrispondenza fra elementi come un vettore, e quindi riportarne una matrice d'adiacenza. Assumendo di appaiare elementi con lo stesso numero, avremo:

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

La caratteristica di questa matrice, chiamiamola x, e che ogni elemento x_{ij} è:

$$x_{ij} = \begin{cases} 0 \\ 1 \end{cases}$$

Decidiamo di trattare la x come un vettore linearizzato lessicograficamente dalla matrice, proprio come avevamo fatto per il vettore costo. Per una matrice di adiacenza $n \times n$, di n elementi in ogni categoria, questo vettore ha dimensione n^2. Questo semplifica la notazione del problema, e sopratutto della matrice A, che sarebbe lasciando x matrice effettivamente un tensore.

Si dimostra quindi facilmente che i vincoli riportati prima possono quindi esprimersi come:

$$\begin{cases} x_{11} + x_{12} + x_{13} + x_{14} = 1 \\ x_{21} + x_{22} + x_{23} + x_{24} = 1 \\ x_{31} + x_{32} + x_{33} + x_{34} = 1 \\ x_{41} + x_{42} + x_{43} + x_{44} = 1 \end{cases}$$

per il primo punto, e:

$$\begin{cases} x_{11} + x_{21} + x_{31} + x_{41} = 1 \\ x_{12} + x_{22} + x_{32} + x_{42} = 1 \\ x_{13} + x_{23} + x_{33} + x_{43} = 1 \\ x_{14} + x_{24} + x_{34} + x_{44} = 1 \end{cases}$$

per il secondo. Imponendo la positività, si hanno quindi le matrici A e b:

$$A : \begin{pmatrix} 1 & 1 & 1 & 1 & 0... & & & & & ...0 \\ 0... & & ...0 & 1 & 1 & 1 & 1 & 0... & & ...0 \\ ... & & & & & & & & & \\ 0... & & & & & ...0 & 1 & 1 & 1 & 1 \\ ... & & & & & & & & & \\ -1 & 0... & & & & & & & & ...0 \\ ... & & & & & & & & & \\ 0... & & & & & & & & & -1 \end{pmatrix}, \quad b : (1, ..., 1, 0, ..., 0)$$

Si nota che il numero di vincoli necessari per n elementi è $2n + n^2$.

3.1.1 Assegnamento cooperativo e non cooperativo

A questo punto conviene fare una distinzione. Abbiamo definito finora il modello:

$$\begin{cases} \min c^T \cdot x \\ x_{11} + x_{12} + x_{13} + x_{14} = 1 \\ \dots \\ x_{41} + x_{42} + x_{43} + x_{44} = 1 \\ x_{11} + x_{21} + x_{31} + x_{41} = 1 \\ \dots \\ x_{14} + x_{24} + x_{34} + x_{44} = 1 \end{cases}$$

che così scritto non nega la possibilità di x con componenti reali. Nell'esempio ciò significa sono ammesse soluzioni dove più agenzie danno contributi reali ai progetti, che possiamo semanticamente interpretare come condividere il carico di lavoro, pur rispettando i vincoli imposti. Decidiamo che questo è corretto se si parla di un problema di **assegnamento cooperativo**. Visto che il problema posto non era di questo tipo, ma era di **assegnamento non cooperativo**, si introduce un'ulteriore vincolo:

$$x \in \mathbb{Z}^n$$

Adesso ogni azienda darà un contributo intero al suo progetto, ergo coi vincoli imposti prima, ogni azienda sarà unica nel dirigere un solo progetto.

Più formalmente, possiamo dire che il passaggio ad assegnamento cooperativo comporta un **rilassamento** dei vincoli del problema. Ovvero, in generale, se un problema non cooperativo ha minimo ottimale nc, il suo associato cooperativo avrà minimo ottimale c con:

$$c \leq nc$$

3.1.2 Forma primale standard

Portiamo quindi questo problema in una forma primale simile a quella vista per altri problemi LP, concesso il vincolo $x \in \mathbb{Z}^n$.

Finora avevamo usato le trasformazioni equivalenti per problemi LP:

1. Trasformazione delle disuguaglianze: $\geq \leftrightarrow \leq$
2. Trasformazione delle uguaglianze:

$$f(x) = c \;\rightarrow\; \begin{cases} f(x) \leq c \\ -f(x) \leq -c \end{cases}$$

3. Trasformazione minimo / massimo:

$$\max f = -\min f \text{ e sopratutto } \bar{x} \in \text{argmax}(f) \Leftrightarrow \bar{x} \in \text{argmin}(-f)$$

Possiamo applicare queste trasformazioni al modello, in particolare la (2), che porta il numero di vincoli a $4n + n^2$.

3.2 Introduzione di surplus

Vediamo un ulteriore tecnica per trasformare problemi LP: si può portare una disequazione del tipo:

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

in un uguaglianza introducendo una variabile ausiliaria s:

$$a_1x_1 + a_2x_2 + ... + a_nx_n + s = b$$

s prende il nome di **slack**, in italiano scarto, o *surplus*.

4 Lezione del 26-09-24

4.1 Geometria dei poliedri

Introduciamo progressivamente i tipi di **combinazione** che ci sono utili nello studio dei problemi di programmazione lineare.

4.1.1 Combinazioni lineari

Definizione 4.1: Combinazione lineare

Dati $x_1, x_2, ..., x_k \in \mathbb{R}^n$ punti, y si dice **combinazione lineare** di $x_1, x_2, ..., x_k$ se:

$$\exists \lambda_i \quad (i = 1, ..., k) \quad \text{t.c.} \quad y = \sum_{i=1}^{k} \lambda_i x_i$$

Le combinazioni lineari sono utili per esprimere la funzione obiettiva sulla base dei vettori costo, ma non bastano a trovarne una soluzione ottimale.

4.1.2 Combinazioni convesse

Si introduce quindi il concetto di:

Definizione 4.2: Combinazione convessa

Dati $x_1, x_2, ..., x_k \in \mathbb{R}^n$ punti, y si dice **combinazione convessa** di $x_1, x_2, ..., x_k$ se:

$$\exists \lambda_i \in [0, 1] \quad (i = 1, ..., k), \quad \sum_{i=1}^{k} \lambda_i = 1 \quad \text{t.c.} \quad y = \sum_{i=1}^{k} \lambda_i x_i$$

Possiamo dare un esempio di cos'è la combinazione convessa di due punti in $\mathbb{R}^2$. Posti x_1 e x_2, si ha:

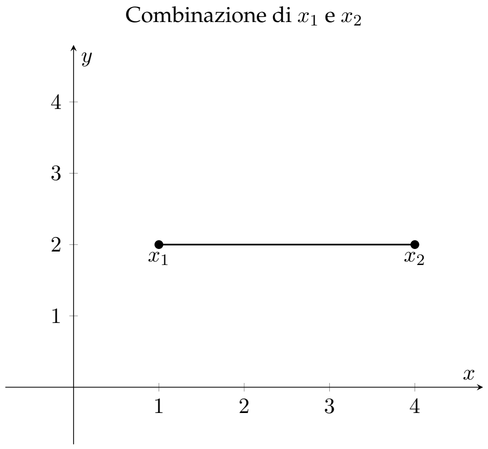
$$\lambda_1 + \lambda_2 = 1 \Rightarrow \lambda_2 = (1 - \lambda_1), \quad y = \lambda x_1 + (1 - \lambda) x_2, \quad \lambda \in [0, 1]$$

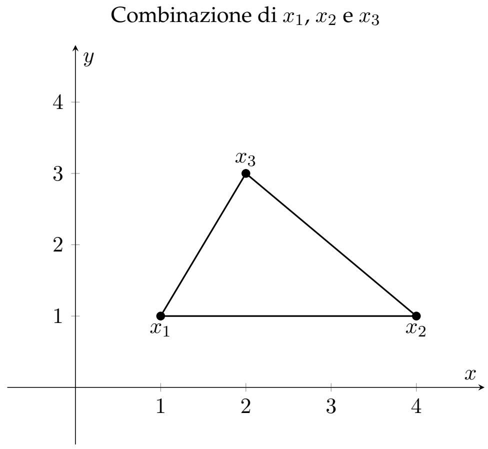
che riconosciamo essere l'equazione di un segmento $x_1\bar{}x_2$ (primo grafico).

Possiamo provare con tre punti: si avrà:

$$y = \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3, \quad \lambda_1 + \lambda_2 + \lambda_3 = 1, \quad \lambda_i \in [0, 1]$$

che si riconduce all'equazione del triangolo di vertici x_1, x_2, x_3 (secondo grafico).





Dai grafici si nota come una combinazione convessa descrive una parte di spazio, che si può definire:

Definizione 4.3: Involucro convesso

L'involucro convesso $\text{conv}(K)$ di un'insieme di punti $K = \{x_1, x_2, ..., x_n\}, \quad x_i \in \mathbb{R}^n$ è definito come il luogo di tutte le loro combinazioni convesse.

Si nota che l'involucro convesso negli esempi precedenti è effettivamente un poliedro convesso che contiene tutti i punti che lo formano. Si può infatti dire:

Teorema 4.1: Minimalità dell'involucro convesso

L'insieme $\text{conv}(K)$ di tutte le combinazioni convesse di n punti è l'insieme convesso minimale che li contiene tutti.

In R^n, possiamo esprimere un insieme convesso lineare come un poliedro convesso, e quindi dire che l'involucro convesso di n punti è il più piccolo (in termini di inclusione) poliedro convesso che li contiene tutti. Inoltre, un'insieme è convesso se e solo se corrisponde al suo involucro convesso.

Si noti che non è detto che a n punti corrisponda un poligono di n vertici. Potrebbe infatti accadere che uno dei punti (chiamiamolo x_j) sia già parte dell'involucro convesso, ergo $x_j \in \text{conv}(K)$.

Le combinazioni convesse ci permettono di descrivere parte delle regioni ammissibili (poliedri) dei problemi di programmazione lineare, ma restano ancora in sospeso problemi che ammettono regioni illimitate. Per descrivere tali regioni, si introduce un altro tipo di combinazione.

4.1.3 Combinazioni coniche

Definizione 4.4: Combinazione conica

Dati $x_1, x_2, ..., x_k \in \mathbb{R}^n$ punti, y si dice **combinazione conica** di $x_1, x_2, ..., x_k$ se:

$$\exists \lambda_i \geq 0 \quad (i = 1, ..., k) \quad \text{t.c.} \quad y = \sum_{i=1}^{k} \lambda_i x_i$$

La combinazione conica di più punti non è più il poliedro convesso che li contiene, ma il cono con vertice nell'origine, convesso o meno, che li contiene, definito come:

Definizione 4.5: Involucro conico

L'involucro conico $\text{cono}(K)$ di un'insieme di punti $K = \{x_1, x_2, ..., x_n\} \in \mathbb{R}^n$ è definito come il luogo di tutte le loro combinazioni coniche.

Questo cono si estende fino all'infinito ($\lambda_i \geq 0$) nelle direzioni dei vettori che lo formano. Il concetto è simile a quello di spazio somma, ma con la differenza che non si va ovunque nello span dei due vettori, ma si seguono le semirette che essi conducono.

Si può dire, analogamente alle combinazioni convesse, che il cono di n punti è il pi+ piccolo (in termini di inclusione) cono convesso che li contiene. Inoltre, un'insieme è un cono convesso se e solo se corrisponde al suo involucro conico.

4.2 Poliedri

Abbiamo definito un poliedro come la regione definita da un sistema di disequazioni lineari, o geometricamente come l'intersezione di un numero finito di semipiani chiusi in $\mathbb{R}^n$. Si dimostra che un poliedro, in quanto intersezione di insiemi convessi, è lui stesso convesso: potremo applicare la definizione di convessità per dire che, presi due punti $x_1, x_2 \in C$ nell'intersezione $C = C_1 \cap C_2$, si avrà che entrambi appartengono sia a C_1 che a C_2, ergo il segmento che li congiunge completamente in ciascuno di quei due insiemi è interamente contenuto in C.

Un poliedro che è anche cono si chiama cono poliedrico. Si dimostra che:

Teorema 4.2: Cono poliedrico

Se P è un cono poliedrico allora:

$$\exists Q \quad \text{t.c.} \quad P = \{x \in \mathbb{R}^n : Qx \leq 0\}$$

con Q matrice.

Senza dimostrazioni, questo è chiaro dal fatto che le disequazioni che compongono il poliedrico sono omogenee (hanno frontiere che passano dall'origine).

Si definiscono poi i **vertici** del poliedro:

Definizione 4.6: Vertice

Un vertice di un poliedro è un punto che non si può esprimere come combinazione convessa propria di altri punti del poliedro. Si indica l'insieme dei vettori di un poliedro P come $\text{vert}(P)$.

Notiamo che i vertici di un poliedro limitato corrispondono ai punti che formano la combinazione convessa equivalente al poliedro. Per poliedri illimitati, introduciamo invece:

Definizione 4.7: Direzione di recessione

Un vettore d è la direzione di recessione di un poliedro se:

$$x + \lambda d \in P \quad \forall x \in P, \quad \forall \lambda \geq 0$$

Si indica come $\mathrm{rec}(P)$ l'insieme delle direzioni di recessione di un poliedro.

Chiaramente, per ogni poliedro P, $0 \in \mathrm{rec}(P)$ e per i poliedri limitati, $\mathrm{rec}(P) = \{0\}$. Notiamo che le direzioni di recessione determinano i vettori del cono che coincide (almeno a distanze abbastanza grandi dall'origine) con i poliedri illimitati.

Inoltre, l'insieme $\mathrm{rec}(p)$ di un poliedro è in sé un cono poliedrico. In particolare:

Teorema 4.3: Cono di recessione

Il cono di recessione di un poliedro P è dato da:

$$P = \{x \in \mathbb{R}^n : Ax \leq b\} \Rightarrow \mathrm{rec}(P) = \{x \in \mathbb{R}^n : Ax \leq 0\}$$

Dimostrazione

Questo teorema vale perchè, per definizione, una direzione di recessione è $x + \lambda d$, che sostituita al sistema dà:

$$A(x + \lambda d) \leq b \Rightarrow Ax + \lambda Ad \leq b$$

Possiamo quindi sottrarre Ax da entrambi i lati:

$$\lambda Ad \leq b - Ax$$

Da ipotesi, $Ax \leq b$ e quindi $b - Ax \geq 0$. Chiamiamo questa quantità postiva k^+ e scriviamo:

$$\lambda Ad \leq k^+$$

Visto che $\lambda \to +\infty$, dovrà essere verò che $Ad \leq 0$, ergo con $d = x$, il teorema è dimostrato.

Un'altro teorema importante lega il cono ai suoi "raggi estremi", cioè quell'insieme finito di vettori la cui combinazione conica corrisponde all'intero cono:

Teorema 4.4: Raggi estremi

Un cono poliedrico $P = \{x \in \mathbb{R}^\ltimes : Ax \leq 0\}$ è l'involucro conico di un'insieme limitato dei suoi punti. Questi punti prendono il nome di *raggi estremi*.

Dimostrazione

Diamo una dimostrazione più o meno intuitiva di questo risultato. Indichiamo con $A_1, ..., A_m$ le righe della matrice A. Queste formeranno dei vettori che indicano la direzione su cui "agisce" la diseguaglianza. Si può semplicemente osservare che questi vettori sono perpendicolari agli estremi del cono. Prendiamo quindi il cono dei vettori, che formerà un'altro cono detto *duale* del cono di partenza:

$$\mathrm{cono}(A_1, ..., A_m) = \{x \in \mathbb{R}^n : Qx \leq 0\}$$

I vettori $Q_1, ..., Q_t$, presi come righe della matrice, sono perpendicolari ai vettori $A_1, ..., A_n$ di partenza, ergo paralleli ai raggi estremi del cono di partenza. Possiamo quindi dire che:

$$P = \text{cono}(Q_1, ..., Q_t)$$

4.2.1 Spazio di linealità

Definiamo infine lo spazio di linealità:

Definizione 4.8: Spazio di linealità

Lo spazio di linealità di un poliedro illimitato P è il più piccolo sottospazio contenuto interamente in P.

Ogni vettore di base di uno spazio di linealità è un vettore d tale che:

$$d \in \text{rec}(P), \quad -d \in \text{rec}(P)$$

ovvero un vettore che è contenuto sia positivo che negativo nelle direzioni di recessione del poliedro.

Questa distinzione è importante in quanto non si può dimostrare completamente il prossimo teorema su poliedri con spazio di linealità $\neq 0$.

4.2.2 Teorema di rappresentazione dei poliedri

Gli strumenti che abbiamo stabilito finora ci permettono di enunciare un'importante risultato, noto come **teorema di rappresentazione dei poliedri**, o teorema di Minkowski-Weyl.

Teorema 4.5: Rappresentazione dei poliedri

Dato un poliedro P definito come $P = \{x \in \mathbb{R}^n : Ax \leq b\}$, si ha:

$$\exists V = \{v_1, ..., v_k\} \in \text{vert}(P), \quad \exists E = e_1, ..., e_p \in \text{rec}(P) \quad \text{t.c.} \quad P = \text{conv}(V) + \text{cono}(E)$$

Questo significa che è possibile rappresentare qualsiasi poliedro attraverso i suoi vertici, e le direzioni in cui si estende all'infinito (ergo le sue direzioni di recessione).

4.2.3 H-rappresentazioni e V-rappresentazioni

Possiamo intendere il risultato precedente come segue: per ogni poliedro di un problema di programmazione LP, abbiamo due possibili rappresentazioni:

- **H-rappresentazione:** come intersezione di semispazi, ergo come insieme delle soluzioni di un sistema di disequazioni lineari.
- **V-rappresentazione:** come un'insieme di vertici e direzioni di regressione, quindi un involucro convesso e un cono di regressione.

Non diamo una dimostrazione del teorema ma possiamo, noti i concetti di H-rappresentazione e V-rappresentazione, dire che:

- Da un **H-rappresentazione** si può ricavare una **V-rappresentazione**: abbiamo un poliedro dato dall'H-rappresentazione $Ax \leq b$. Dividiamo questo poliedro in una parte limitata e una parte illimitata: la parte limitata sarà data dai vertici, mentre la parte illimitata sarà data dal cono di regressione:

  $$P = \text{conv}(v_1, ..., v_k) + rec(P)$$

  A questo punto possiamo esprimere, come dallo scorso lemma, come combinazione conica dei suoi raggi estremi:

  $$P = \text{conv}(v_1, ..., v_k) + \text{cono}(e_1, ..., e_p)$$

- Da un **V-rappresentazione** si può ricavare una **H-rappresentazione**: abbiamo un poliedro dato dalla V-rappresentazione $P = \text{conv}(v_1, ..., v_k) + \text{cone}(e_1, ..., e_p)$. Prendiamo separatamente gli involucri convessi e conici: avremo che per l'involucro convesso possiamo immediatamente riportare in una forma ad intersezione di semipiani (disequazioni):

  $$\text{conv}(v_1, ..., v_k) = \{x \in \mathbb{R}^n : A'x \leq b'\}$$

  mentre per l'involucro conico possiamo portare in una forma ad intersezione di disequazioni omogenee:

  $$\text{cone}(e_1, ..., e_p) = \{x \in \mathbb{R}^n : A''x \leq 0\}$$

  Per trovare l'H-rappresentazione, basterà combinare queste due rappresentazioni in una forma del tipo:

  $$P = \{x \in \mathbb{R}^n : A'x \leq b', A''x \leq 0\}$$

La somma in $P = \text{conv}(V) + \text{cono}(E)$ si riferisce alla somma vettoriale fra tutti i possibili punti di $\text{conv}(V)$ e $\text{conv}(E)$, come quella studiata sui sottospazi vettoriali (anche se nessuno dei due insiemi è un sottospazio vettoriale). Per un dato insieme $\text{conv}(V)$, quindi, l'aggiunta di $\text{cono}(E)$ rappresenta la "proiezione" di tale insieme nelle direzioni di recessione indicate dal cono.

Più propriamente, posto $\text{lineal}(P) = 0$, si ha:

Teorema 4.6: Rappresentazione dei poliedri non lineali

Dato un poliedro P definito come $P = \{x \in \mathbb{R}^n : Ax \leq b\}$, tale che $\text{lineal}(P)$, si ha:

$$P = \text{conv}(\text{vert}(P)) + \text{rec}(P)$$

la limitazione di linealità è necessaria in quanto un poliedro lineale potrebbe non essere rappresentato, nelle sue dimensioni infinite, dal semplice insieme dei suoi vettori. In verità, è possibile dimostrare che:

Teorema 4.7: Linealità di poliedri

Per ogni poliedro P non vuoto si ha:

$$\text{vert}(P) \neq \emptyset \Leftrightarrow \text{lineal}(P) = 0$$

ergo applicando lo scorso teorema potremmo provare a rappresentare un poliedro attraverso un'insieme di vettori vuoto.

Per i poliedri che otteniamo dai problemi di programmazione lineare, però, abbiamo i corollari:

- Un poliedro limitato è l'involucro convesso dei suoi vertici;
- Se il poliedro ha vincoli di positività sulle sue variabili, allora non è lineale, ergo si applica il teorema di rappresentazione. Questo è il tipo di poliedri a cui siamo abituati.

4.3 Teorema fondamentale della PL

Quanto riportare finora sulla geometria dei poliedri può essere usato per dimostrare il seguente teorema:

Teorema 4.8: Teorema fondamentale della PL

Sia dato un poliedro P rappresentato come:

$$P = \text{conv}(V) + \text{cono}(E), \quad V = \{v_1, ..., v_k\}, \quad E = \{e_1, ..., e_p\}$$

Se il problema $\mathcal{P}$ con regione ammissibile P ha valore ottimo finito, allora esiste $s \in \{1, ..., k\}$ tale che v_k è soluzione ottima di $\mathcal{P}$.

In sostanza, se un problema LP ha soluzione, essa si trova su uno dei vertici del poliedro della regione ammissibile.

Dimostrazione

Sia dato un problema LP $\mathcal{P}$ in forma primale standard, ergo posto come:

$$\begin{cases} \max c^T \cdot x \\ Ax \leq b \end{cases}$$

ergo con regione ammissibile rappresentata da un poliedro P.

Dal teorema della rappresentazione, possiamo esprimere il poliedro come:

$$P = \text{conv}(V) + \text{cono}(E), \quad V = \{v_1, ..., v_k\}, \quad E = \{e_1, ..., e_p\}$$

Combiniamo le due equazioni, ergo esprimiamo prima il punto $\bar{x}$ generico del poliedro applicando le definizioni di involucro convesso e conico:

$$\bar{x} \in P : P = \text{conv}(V) + \text{cono}(E), \quad \bar{x} = \sum_{i=1}^{k} \lambda_i v_i + \sum_{j=i}^{p} \mu_j e_j$$

ed esprimiamo quindi la funzione obiettivo come il prodotto scalare fra il vettore costo e il punto $\bar{x}$ del poliedro:

$$c^T \cdot \bar{x} = c^T \cdot \left(\sum_{i=1}^{k} \lambda_i v_i + \sum_{j=i}^{p} \mu_j e_j \right) = \sum_{i=1}^{k} \lambda_i c^T v_i + \sum_{j=i}^{p} \mu_j c^T e_j$$

A questo punto conviene chiarire su cosa significa che il problema ha valore ottimo finito. Il secondo termine è la combinazione conica della sommatoria dei vettori di recessione scalati dal vettore costo. Se almeno uno dei $c^T e_j > 0$, si avrà che portando $\mu_j \to +\infty$ la funzione avrà massimo $= \infty$. Geometricamente, questo significa che esiste una direzione illimitata del poliedro dove i vettori costo permettono alla funzione di crescere all'infinito.

Dunque sarà vero che $c^T e_j \leq 0 \quad \forall j \in \{1, ..., p\}$ se vogliamo che la funzione abbia valore ottimo finito.

Possiamo quindi usare questa ipotesi per dire:

$$c^T \cdot \bar{x} = \sum_{i=1}^{k} \lambda_i c^T v_i + \sum_{j=i}^{p} \mu_j c^T e_j \leq \sum_{i=1}^{k} \lambda_i c^T v_i \leq \sum_{i=1}^{k} \max_{1 \leq i \leq p} \left(\lambda_i c^T v_i \right)$$

$$= \left(\max_{1 \leq i \leq p} c^T v_i \right) \sum_{i=1}^{k} \lambda_i = \max_{1 \leq i \leq p} c^T v_i = c^T v_k$$

E quindi $\max_{x \in P} c^T \cdot x \leq c^T v_k$. A questo punto, visto che $\bar{x}$ è effettivamente un punto della regione ammissibile, sarà vero che:

$$c^T v_k \leq \max_{x \in P} c^T \cdot x$$

E dunque $\max_{x \in P} c^T \cdot x = c^T v_k$, come volevasi dimostrare.

5 Lezione del 30-09-24

5.1 Calcolo dei vertici

Troviamo adesso il modo di calcolare i vertici del poliedro di un problema LP in forma primale standard.

Avevamo già dato la definizione di vertice come punto non ottenibile come combinazione convessa degli altri punti del poliedro (definizione 4.6). Questa definizione è corretta ma poco utile per il calcolo procedurale. Dimostriamo quindi un teorema utile.

Innanzitutto, assumiamo che in generale, per un problema LP $\mathcal{P}$ con n variabili decisionali e m vincoli, si ha che $n < m$. A questo punto, possiamo dire:

Definizione 5.1: Soluzione di base primale

Sia dato un problema LP $\mathcal{P}$ in forma primale standard con $n < m$. Sia $B \subseteq \{1, ..., m\}$ un sottoinsieme di indici di riga tale che $\text{card}(B) = n$. A questo punto, sia A_B la sottomatrice di A con con righe indicate da B, e b_B il sottovettore colonna di B con righe indicate da B, con det $\det A_B \neq 0$. Allora la soluzione di:

$$A_B x = b_B$$

è detta soluzione di base primale di $\mathcal{P}$.

da qui possiamo dimostrare il teorema:

Teorema 5.1: Caratterizzazione dei vertici primali

Su un problema in forma primale standard, un punto x del poliedro P è un vertice di P se e solo se è una soluzione di base primale ammissibile, ovvero:

$$x \in \text{vert}(P) \Leftrightarrow x \text{ è soluzione di base primale}$$

Riflettiamo un'attimo su questo risultato: se la definizione di vertice era effettivamente sufficiente a dichiarare *quali* erano i vertici, un teorema di caratterizzazione come quello sopra riportato ci fornisce una procedura per calcolarli tutti a partire da P. In questo, definizione e teorema di caratterizzazione sono effettivamente intercambiabili, ovvero:

$$\text{definizione} \Leftrightarrow \text{teorema di caratterizzazione}$$

6 Lezione del 01-10-24

6.1 Soluzioni di base primali degeneri

Abbiamo dato un teorema di caratterizzazione dei vertici primali. Questo teorema si basava sulla nozione di **soluzione di base primale**. Possiamo fare una distinzione fra soluzione di base degeneri e non degeneri:

Definizione 6.1: Soluzione di base degenere

Quando una soluzione di base è soluzione di più combinazioni delle disequazioni del problema, essa si dice degenere.

Questa definizione è esatta ma non particolarmente utile. Sostanzialmente, ci dice soltanto che una soluzione degenere è **ridondante** su più combinazioni di disequazioni (cioè risolve $A_B x = b_B$ su più permutazioni degli $1, ..., m$ elementi in classi n in B). Si noti che ridondante non significa **eliminabile**: questa affermazione purtroppo è vera soltanto in R^2, dove effettivamente si può rimuovere una delle disequazioni ridondanti ed avere sempre lo stesso risultato.

Diamo quindi una caratterizzazione delle soluzioni di base primali degeneri appoggiandoci al teorema di caratterizzazione dei vertici, ergo al concetto di soluzione di base:

Teorema 6.1: Caratterizzazione delle soluzioni di base primali degeneri

Se una soluzione è di base, ergo scelto $B = \{1, ..., m\}$ con $\text{card}(B) = n$ è data da $A_B x = b_B$, possiamo dire che è pure degenere quando $\exists i \in N$ t.c. $A_i x = b_i$, con I = $\{1, ..., m\} - B$.

Quindi, una soluzione di base è degenere quando almeno una variabile di base si annulla per almeno una delle disequazioni non di base indicate dagli indici I, che sono tutti gli indici fra $\{1, ..., m\}$ non contenuti in B.

Sulla stessa linea di pensiero, possiamo dimostrare un'altro teorema, stavolta sul concetto piuttosto intuitivo di ammissibilità. Potremmo infatti dire che una soluzione di base ammissibile, cioè che rientra all'interno della regione ammissibile, è tale se:

Teorema 6.2: Caratterizzazione delle soluzioni di base primali ammissibili

Se una soluzione è di base, ergo scelto $B = \{1, ..., m\}$ con $\text{card}(B) = n$ è data da $A_B x = b_B$, possiamo dire che è ammissible quando $\forall i \in N$ si ha $A_i x \leq b_i$, con $I = \{1, ..., m\} - B$.

cioè banalmente rispetta tutte le disequazioni.

6.1.1 Considerazioni numeriche sui numeri di soluzioni base

Solitamente un problema con n variabili decisionali a $m \geq n$ vincoli. Posti questi vincoli, visto che per calcolare $\text{vert}(P)$ prendiamo effettivamente tutte le combinazioni degli m vincoli classe n variabili decisionali, possiamo usare il coefficiente binomiale per calcolare il numero massimo di potenziali vertici:

$$\text{card}(\text{vert}(P)) \sim \binom{m}{n} = \frac{m!}{n!(m-n)!}$$

In verità, i vertici sono solitamente meno, in quanto possiamo rimuovere le soluzioni non ammissibili. Inoltre, le soluzioni degeneri non contribuiscono al risultato, ergo anche quelle non sono rilevanti.

6.2 Riassunto delle trasformazioni equivalenti

Riassumiamo adesso le trasformazioni equivalenti che abbiamo individuato finora per le disequazioni di problemi LP:

- $\min(C^T \cdot x) \leftrightarrow \max(C^T \cdot x)$: trasformiamo problemi di massimo in problemi di minimo invertendo i segni;
- $Ax \geq b \leftrightarrow -Ax \leq -b$: invertiamo il verso della diseguaglianza moltiplicando per -1;
- $Ax = b \rightarrow Ax \leq b \wedge Ax \geq b$: convertiamo un'uguaglianza in una coppia di diseguaglianze;
- $Ax \leq b \rightarrow Ax + s = b$: convertiamo una diseguaglianza in un'uguaglianza introducendo una variabile di surplus. Si nota che la variabile di surplus può essere riconosciuta per essere rimossa, da:
  - $s > 0$;
  - Compare in un solo vincolo, che è di uguaglianza;
  - Ha coefficiente 0 nella funzione obiettivo, e 1 nell'equazione dove compare;
- $x \gtreqless 0 \rightarrow x = x^+ - x^-, \quad x^+ \geq 0, \quad x^- \geq 0$: aggiriamo il vincolo di positività introducendo parti positive e negative delle variabili decisionali, con rispettivi vincoli di positività.

Usiamo queste trasformazioni per portare i problemi LP in forme standard. Esistono molteplici forme standard, ma in questo corso ci riguardano: il formato *linprog*, usato dal pacchetto software *MATLAB*, le forme standard primale (già vista) e duale (che vedremo fra poco).

7 Lezione del 02-10-24

7.1 Trasporto

Poniamo il seguente problema:

Problema 7.1: Trasporto

Due centrali del latte di Firenze producono rispettivamente 50 e 60 mila litri di latte al giorno. Le centrali servono tre quartieri, che consumano rispettivamente 30, 30 e 20 mila litri di latte al giorno. Si conosce il costo necessario per portare un migliaio di litri di latte da ogni centrale a ogni quartiere, riportato nella seguente tabella:

	Novoli	**Statuto**	**Rifredi**
Centrale A	6	8	4
Centrale B	7	3	9

Vogliamo capire quanto latte deve spedire ogni centrale ad ogni quartiere.

Nota simpatica: secondo l'indagine INRAN-SCAI 2005-06, l'italiano medio consuma 0.115g di latte al giorno, che per un peso specifico di circa 1.040kg/L fanno 0.11L di latte al giorno. Al 2024, il comune di Firenze ha $364\,073$ abitanti, ergo dovrebbe avere bisogno di approssimativamente $40\,258\text{L}$ di latte al giorno. I fiorentini nell'esempio devono avere le ossa veramente forti!

Possiamo esprimere il problema dell'esempio come un problema LP. Abbiamo innanzitutto che i costi di trasporto formano una matrice:

$$C_{matr} = \begin{pmatrix} 6 & 8 & 4 \\ 7 & 3 & 9 \end{pmatrix}$$

che possiamo linearizzare, come avevamo fatto nei problemi di assegnamento di costo minimo, in un vettore costo:

$$C = (6, 8, 4, 7, 3, 9)$$

Questo vettore moltiplica il vettore delle variabili decisionali, che è la linearizzazione della matrice:

$$x_{matr} = \begin{pmatrix} x_{13} & x_{14} & x_{15} \\ x_{23} & x_{24} & x_{25} \end{pmatrix}$$

Questa matrice non rappresenta altro che quanto latte mandare ad ogni quartiere.

A questo punto, possiamo stabilire i vincoli. Innanzitutto, non si può avere più latte di quanto viene prodotto, ergo:

$$\begin{cases} x_{13} + x_{14} + x_{15} \leq 50 \\ x_{23} + x_{14} + x_{15} \leq 60 \end{cases}$$

inoltre, si vuole fornire ad ogni quartiere il fabbisogno richiesto, ergo:

$$\begin{cases} x_13 + x_{23} \geq 30 \\ x_14 + x_{24} \geq 30 \\ x_15 + x_{25} \geq 20 \end{cases}$$

Questo è un problema di programmazione lineare.

In generale, quindi, un problema di trasporto minimizza la funzione obiettivo data da una matrice di costo in $n \times m$ variabili, con m vincoli di riga sul vettore o_j dei limiti di produzione, e n vincoli di colonna sul vettore d_j della domanda, in forma:

$$\begin{cases} \min \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} = C^T \cdot x \\ \sum_{i=1}^{m} x_{ij} \geq d_j \quad \forall j = 1, ..., n \\ \sum_{j=1}^{n} x_{ij} \leq o_i \quad \forall i = 1, ..., m \\ x \geq 0 \end{cases}$$

Non ci sono soluzioni se la domanda supera l'offerta, cioè sé:

$$\sum_{j=1}^{m} d_j \geq \sum_{i=1}^{m} o_i$$

Mentre in caso di eccessi di produzione, potremmo trasformare le diseguaglianze in eguaglianze, e aggiugnere un carico "fittizio" con costo zero dove deviare il surplus di produzione.

Inoltre, come avevamo detto per i problemi di assegnamento di costo minimo, anche qui potremmo scegliere di distinguere fra trasporti divisibili (nel campo $x = \mathbb{R}^n$) e indivisibili (col vincolo aggiunto $x = \mathbb{Z}^n$).

7.2 Forma duale standard

Avevamo definito la forma primale standard:

$$\begin{cases} \max C^T \cdot x \\ Ax \leq b \end{cases}$$

Introduciamo adesso la forma duale standard:

Definizione 7.1: Forma duale standard

Un problema di programmazione lineare si dice in forma duale standard quando è espresso in forma:

$$\begin{cases} \min(c^T \cdot x) \\ Ax = b \\ x \geq 0 \end{cases}$$

7.2.1 Vertici del duale

Sulle forme duali è semplice il calcolo dei vertici. Possiamo infatti avere, come avevamo fatto sulla primale:

Definizione 7.2: Soluzione di base duale

Sia dato un problema LP $\mathcal{P}$ in forma duale standard. Sia $B \subseteq \{1, ..., n\}$ un sottoinsieme di indici di variabili decisionali tale che $\text{card}(B) = m$. Chiamiamo x_B l'insieme delle variabili decisionali individuate da B, e x_N l'insieme delle $n - m$ variabili decisionali rimanenti:

$$x = \{x_B, x_N\}$$

Impostiamo quindi tutte le x_N a 0: avremo un sistema di m variabili in m equazioni, quindi determinato. La soluzione di quel sistema è detta soluzione di base duale di $\mathcal{P}$.

Indichiamo spesso questo vertice come $(bA_B^{-1}, 0)$. Questa definizione porta ad una caratterizzazione dei vertici del tutto analoga a quella dichiarata sui problemi in forma primale standard:

Teorema 7.1: Caratterizzazione dei vertici duali

Su un problema in forma duale standard, un punto x del poliedro P è un vertice di P se e solo se è una soluzione di base duale ammissibile, ovvero:

$$x \in \text{vert}(P) \Leftrightarrow x \text{ è soluzione di base duale}$$

7.2.2 Soluzioni di base duali degeneri

Possiamo ricavare il concetto di soluzione degenere (e anche di soluzione ammissibile) sui vertici del poliedro del duale. Si ha:

Teorema 7.2: Caratterizzazione delle soluzioni di base duali degeneri

Se una soluzione è di base, ergo scelto $B = \{1, ..., n\}$ con $\text{card}(B) = m$ è data da $(bA_B^{-1}, 0)$, possiamo dire che è pure degenere quando $\exists i \in B$ tale che almeno una componente si annulla.

e riguardo l'ammissibilità:

Teorema 7.3: Caratterizzazione delle soluzioni di base duali ammissibili

Se una soluzione è di base, ergo scelto $B = \{1, ..., n\}$ con $\text{card}(B) = m$ è data da $(bA_B^{-1}, 0)$, possiamo dire che è ammissibile quando il vettore soluzione è ≥ 0.

8 Lezione del 03-10-24

8.1 Teoria della dualità

Introduciamo adesso uno dei concetti più importanti della programmazione lineare. Avevamo posto problemi LP in forma primale standard come:

$$\begin{cases} \min(c^T \cdot x) \\ Ax \leq b \end{cases}$$

Ottimizzare questo problema significa partire dal basso e avvicinarsi verso un punto di massimo. Potremmo scegliere di seguire il percorso opposto: cercare di estrapolare un limite superiore per la soluzione dai vincoli, e minimizzarlo.

Per fare ciò introduciamo m variabili, una per ogni disequazione, che denoteremo come $y_1, ..., y_m$. Moltiplichiamo ogni disequazione per la y_i corrispondente a destra e a sinistra. Su un semplice problema $n, m = 2$, questo darà una forma del tipo:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 \leq b_1 \\ a_{21}x_1 + a_{22}x_2 \leq b_2 \end{cases} \to \begin{cases} y_1 \cdot (a_{11}x_1 + a_{12}x_2) \leq b_1 y_1 \\ y_2 \cdot (a_{21}x_1 + a_{22}x_2) \leq b_2 y_2 \end{cases}$$

Se vincoliamo gli y_i in modo che ogni variabile decisionale x_i del sistema abbia un coefficiente del costo $\geq c_i$ corrispondente, otterremo una disequazione che ha a sinistra una situazione di valore uguale o addirittura migliore di quella data dalla funzione costo, e a destra un massimo (che era ciò che stavamo cercando, un limite superiore). Abbiamo quindi una serie di variabili vincolate:

$$\begin{cases} y_1 a_1 1 + y_2 a_2 1 \geq c_1 \\ y_1 a_1 2 + y_2 a_2 2 \geq c_2 \end{cases}$$

e una funzione da minimizzare:

$$\min(b_1 y_1 + b_2 y)$$

Cioè, ci siamo ricondotti ad un altro problema LP. Possiamo formalizzare questo risultato:

Definizione 8.1: Duale di un problema LP

Per un qualsiasi problema LP $\mathcal{P}$, detto primale, con $m \geq n$, possiamo definire il duale $\mathcal{D}$:

$$P : \begin{cases} \max(c^T \cdot x) \\ Ax \leq b \end{cases} \to D : \begin{cases} \min(b^T \cdot y) \\ A^T y = c \\ y \geq 0 \end{cases}$$

dove si nota $x \in \mathbb{R}^n$ e $y \in \mathbb{R}^m$.

Il duale viene posto in forma duale standard in quanto ciò che ci interessa è *stringere* il limite superiore fino al suo minimo, in un modo che fa combaciare perfettamente le variabili con il loro vettore costo, da cui le uguaglianze.

Si può dimostrare che l'operazione del calcolo del duale è involutoria: il duale del duale è nuovamente il primale, e così via.

8.1.1 Dualità debole

Visto che abbiamo costruito il duale per avere un limite superiore dei valori ottenuti dalla funzione obiettivo del primale, potremo dimostrare facilmente:

Teorema 8.1: Dualità debole

Se i poliedri P e il suo duale D non sono vuoti, allora:

$$c^T x \leq y^T b \quad \forall x \in P, \forall y \in D$$

Cioè il duale è sempre maggiore del primale.

8.1.2 Dualità forte

Idealmente, ciò che vorremmo è che primale e duale convergessero verso un punto comune, ergo l'ottimo di entrambi. Effettivamente, questo risultato è verificato:

Teorema 8.2: Dualità forte

Se i poliedri P e il suo duale D non sono vuoti, allora:

$$-\infty \leq \min_{y \in D} b^T y = \max_{x \in P} c^T x \leq +\infty$$

Il teorema della dualità forte afferma che, se entrambi i poliedri (primale e duale) sono non vuoti, allora condividono l'ottimo, e anzi, che due soluzioni nel primale e nel duale sono ottime solo se hanno lo stesso valore. Se invece solo il primale (solo il duale) è vuoto, si ha che entrambi condividono ottimo $-\infty$ (∞). Quando entrambi sono vuoti non si ha soluzione condivisa.

8.1.3 Scarti complementari

Si può dimostrare il seguente teorema:

Teorema 8.3: Scarti complementari

Se le soluzioni x e y dei problemi primale e duale $\mathcal{P}$ e $\mathcal{D}$ sono entrambe ottime, allora vale:

$$y^T(b - Ax) = 0$$

Questo si ricava dal fatto che, per la dualità forte, si ha che:

$$c^T x = y^T Ax = y^T b \Rightarrow y^T(b - Ax) = 0$$

Il significato del teorema è che, se una disequazione nel primale è *stretta*, allora la corrispondente variabile nel duale è $\neq 0$, e viceversa.

8.1.4 Soluzioni di base

Avevamo dato una definizione di soluzione di base per problemi LP in forma sia primale che duale. Possiamo dimostrare che non solo questa nozione esiste su entrambe le formule, ma è analoga su coppie primale / duale.

Avevamo posto che la formazione di una certa base $B \in \{1, ..., m\}$ per ricavare soluzioni di base. Per il primale, questo significa partizionare la matrice e i termini noti:

$$A = \begin{pmatrix} A_B \\ A_N \end{pmatrix}, \quad b = \begin{pmatrix} b_A \\ b_N \end{pmatrix}$$

mentre per il duale, significherà partizionare le variabili introdotte:

$$y = \begin{pmatrix} y_B \\ y_N \end{pmatrix}$$

noto il numero di $y_1, ..., y_m$ uguale a m.

Questo significa che possiamo trovare due soluzioni di base corrispondenti per un'unica base su primale e duale. Queste sono:

- Soluzione di base primale: $x = A_B^{-1} b_B$;
- Soluzione di base duale: $y_B^T = c^T A_B^- 1, \quad y_N = 0$;

Si dice che le soluzioni di base sono **complementari**.

Dimostrazione Vogliamo che $y^T(b - Ax)$ sia $= 0$ soddisfatte le condizioni di base. Applichiamo quindi la base:

$$y^T(b - Ax) = \left(y_B^T, y_N^T\right) \begin{pmatrix} b_B - A_B x \\ b_N - A_N x \end{pmatrix} = \left(c^T A_B^{-1}, 0\right) \begin{pmatrix} b_B - A_B A_B^{-1} b_B \\ b_N - A_N A_B^{-1} b_B \end{pmatrix}$$

$$= \left(c^T A_B^{-1}, 0\right) \begin{pmatrix} 0 \\ b_N - A_N A_B^{-1} b_B \end{pmatrix} = 0$$

Questo nome non è a caso, in quanto si può dimostrare le due soluzioni sono in scarti complementari. Da questo risultato, si ha che se entrambe le soluzioni sono ammissibili, cioé:

- La primale è ammissibile:
  $$\forall i \in N \text{ si ha } A_i x \leq b_i$$
  ergo i vincoli sono soddisfatti;
- La duale è ammissibile:
  $$y \geq 0$$

questo è condizione sufficiente perche la soluzione sia ottima, e dagli scorsi corollari, sia l'ottima sia del primale che del duale.

Formalizziamo quanto detto in un teorema:

Teorema 8.4: Condizioni di ottimalità di soluzione di base

Dato un vertice del primale, ottenuto da una certa base, si può costruire il complemento duale sulla stessa base. Se entrambi i vertici ottenuti sono ammissibili, allora sono uguali e ottimi dei rispettivi problemi.

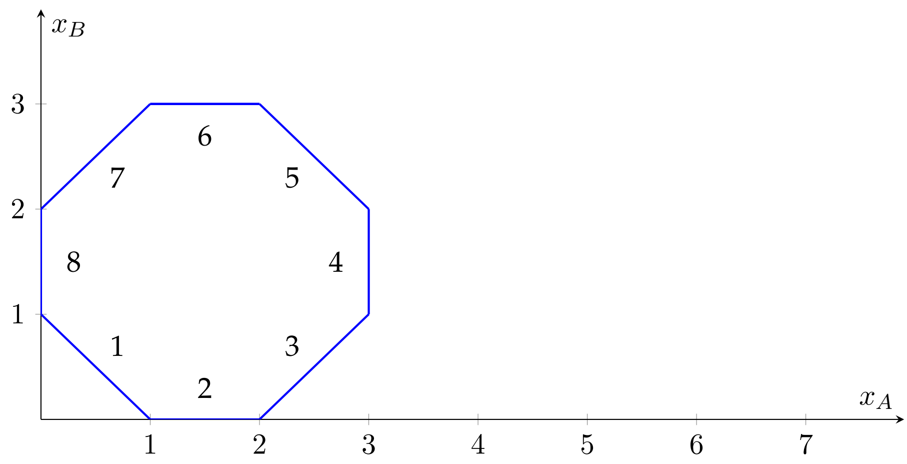
9 Lezione del 07-10-24

9.1 Algoritmo del simplesso primale

Supponiamo di avere un problema LP in formato primale standard con $n = 8$ vincoli, espresso come:

$$\begin{cases} \max(c^T \cdot x) \\ Ax \leq b \end{cases}$$

e con poliedro:



Scegliamo un vertice di partenza, per adesso ad arbitrio: diciamo $\bar{x} = (0, 1)$ (vedremo in seguito un'algoritmo particolare per ricavare un vertice, che ci permetterà anche di determinare se il poliedro è vuoto o meno). Ci chiediamo se questo vertice $\bar{x}$ è ottimo. Visto che è vertice, abbiamo che per una matrice A_B e un vettore b_B di base:

$$\bar{x} = A_B^{-1} b_B$$

e che possiamo costruire il complementare duale $\bar{y}$, impostando a zero le variabili fuori base e risolvendo il sistema:

$$\bar{y} = (c A_B^{-1}, 0)$$

e applicare il test di ottimalità, cioè vedere se:

$$c A_B^{-1} \geq 0$$

ergo $\bar{y} \in D$, quindi il complementare duale esiste e il vertice è ottimo. Se questa condizione risulta verificata, possiamo fermarci, in quanto abbiamo trovato la soluzione ottimale.

In caso contrario, avremo $\exists k \in B$ tale che $\bar{y}_k < 0$. Dovremo quindi spostarci verso un'altro vertice, magari *adiacente*, che dal punto di vista delle basi, significa cambiare un solo indice di base, conservando gli altri. Possiamo formalizzare questa affermazione definendo un **indice uscente** h ed un **indice entrante** k. Sostituire un indice di base significa effettuare il cambio di base:

$$B := B \setminus \{h\} \cup \{k\}$$

Resta la domanda di *quale* spigolo scegliere: in uno spazio vettoriale $\mathbb{R}^n$, ho a disposizione n spigoli che si staccano dallo stesso vertice. Ovviamente, vorrei scegliere uno spigolo che accresce la funzione obiettivo, e si può dimostrare che ne esiste almeno uno: altrimenti sarei già all'ottimo. Inoltre, avendo un metodo per scegliere sempre lo spigolo di crescita maggiore potrei dire 2 cose: l'algoritmo tende all'ottimo (il vertice da cui non si staccano spigoli che accrescono la funzione obiettivo), e termina in un numero finito di passi (prima o poi raggiungerà inevitabilmente un vertice che massimizza la funzione).

Prima però dobbiamo chiarire una questione: scegliere un nuovo spigolo significa trovare 2 indici base, uno da eliminare e uno da inserire. Si può dire che il primo indice, quello uscente, indica anche la direzione di spostamento: allentando un vincolo ci spostiamo sulla semiretta del prossimo. Allo stesso tempo, scegliere un indice da rimuovere non basta: dobbiamo scegliere quale introdurre, che geometricamente significa capire *quanto* ci possiamo spostare lungo la semiretta prima di uscire dalla regione di ammissibilità. Vediamo quindi questi due passaggi in ordine.

- **Indice uscente**
  Diciamo:
  $$W = \left(-A_B^{-1}\right)$$
  e prendiamo le colonne W^i corrispondenti agli indici di base scelti.

  Possiamo allora dire che l'equazione degli spigoli dati dalle disequazioni all'indice i sono:
  $$\bar{x} + \lambda W^i$$
  Mettiamo questa equazione nella funzione costo:
  $$c\left(\bar{x} + \lambda W^i\right) = c\bar{x} + \lambda c W^i$$
  Qui abbiamo $c\bar{x}$, che è il valore nel vertice, e un'altro termine scalato da λ. Ricordiamo poi che $cA_B^{-1} = \bar{y}_B$, e che $W = \left(-A_B^{-1}\right)$, ergo $cW^i = -\bar{y}_B$:
  $$c\bar{x} + \lambda c W^i = c\bar{x} - \lambda \bar{y}_B$$
  Vogliamo quindi "allentare" l'indice (e il corrispettivo vertice) che ci dà $\bar{y}_i < 0$, in quanto è quello che restituisce un $cW^i > 0$, e quindi un accrescimento della funzione. Definiamo allora questo indice:

  Definizione 9.1: Indice uscente primale

  Chiamiamo indice uscente h, da una certa soluzione della base B:
  $$h := \min\{i \in B \text{ t.c. } \bar{y}_i < 0\}$$

  Il $\min$ significa che in caso di più i negativi, si adotta la regola anticiclo (di Bland) di scegliere il primo. In caso di nessun i negativo, la complementare duale esiste e siamo sull'ottimo.

- **Indice entrante**
  Adesso cerchiamo per quali λ lo spigolo $\bar{x} + \lambda W^h$ resta ammissibile, ergo soddisfa:
  $$A_i\left(\bar{x} + \lambda W^h\right) \leq b_i, \quad i \in N$$
  Questo significa effettivamente vedere qual'è il primo vincolo che "stringiamo", o che incontriamo, spostandoci lungo la semiretta ottenuta allentando il vincolo dato dall'indice uscente.

  Possiamo dire:
  $$A_i\left(\bar{x} + \lambda W^h\right) = A_i\bar{x} + \lambda A_i W^h \leq b_i$$
  da cui si ricava (e si risolve) la disequazione di primo grado:
  $$\lambda A_i W^h \leq b_i - A_i\bar{x} \Rightarrow \lambda \leq \frac{b_i - A_i\bar{x}}{A_i W^h}$$
  Notiamo che se fosse $A_i W^h \leq 0, \ \forall i \in N$, avremmo che l'indice rappresenta una direzione di regressione, in quanto $\lambda \to +\infty$. Si ha quindi che il duale non ha

soluzione, e il primale $\to +\infty$. In caso contrario, noi vogliamo trovare il primo vincolo che si va a stringere, quindi dovremo calcolare tutti gli r_i:

$$r_i = \frac{b_i - A_i\bar{x}}{A_iW^h}, \quad i \in N, \quad A_iW^h > 0$$

e scegliere l'indice che dà $\vartheta = \min(r_i)$. Definiamo allora anche questo indice:

Definizione 9.2: Indice entrante primale

Chiamiamo indice entrante k, da una certa soluzione della base B e un certo indice uscente h:

$$k := \min\{i \in N \text{ t.c. } A_iW^h > 0, \quad \frac{b_i - A_i\bar{x}}{A_iW^h} = \vartheta\}$$

Anche qui, il $\min$ serve a selezionare il primo indice valido, ed è una regola anticiclo (di Bland). Notiamo due possibili situazioni:

– Si potrebbero avere più r_i uguali: questi rappresentano soluzioni di base degenere *in arrivo*, in quanto sono più modi di arrivare allo stesso vertice stringendo vincoli diversi;
– Si potrebbe avere un r_i nullo: questo significa che il vertice è sullo stesso vertice da dove siamo partiti, ergo rappresenta una soluzione di base degenere *in partenza*.

Come prima, le regole anticiclo di Bland assicurano anche che l'algoritmo non si blocchi a ciclare su queste soluzioni degeneri.

Abbiamo quindi tutti gli strumenti necessari alla formulazione dell'algoritmo del simplesso:

Algoritmo 1 del simplesso primale

Input: un problema LP in forma primale standard
Output: la soluzione ottima
Trova una base B che genera una soluzione di base primale ammissibile.
ciclo:
Calcola la soluzione di base primale $\bar{x} = A_b^{-1} b_B$ e la soluzione di base duale $\bar{y} = (cA_b^{-1}, 0)$
if $\bar{y_B} \geq 0$ **then**
Fermati, $\bar{x}$ è ottima per P e $\bar{y}$ è ottima per D k
else
Calcola l'indice uscente:

$$h := \min\{i \in B \text{ t.c. } \bar{y_i} < 0\}$$

poni $W := -A_B^{-1}$ e indica con W^h la h-esima colonna di W
end if
if $A_i W^h \leq 0 \quad \forall i \in N$ **then**
Fermati, $P \to +\infty$ e D non ha soluzione ottima.
else
Calcola:

$$\vartheta = \min\{\frac{b_i - A_i\bar{x}}{A_i W^h} \text{t.c.} \quad i \in N, \quad A_i W^h > 0\}$$

e trova l'indice entrante:

$$h := \min\{i \in N \text{ t.c. } A_i W^h > 0, \quad \frac{b_i - A_i\bar{x}}{A_i W^h} = \vartheta\}$$

end if
Aggiorna la base come:

$$B := B \setminus \{h\} \cup \{k\}$$

Torna a ciclo

10 Lezione del 08-10-24

Abbiamo trovato un'algoritmo per ricavare una soluzione ottimale partendo da una soluzione di base primale ammissibile. Nel fare ciò, si è dato per scontato che l'algoritmo di partenza ci avrebbe fornito una soluzione di base primale ammissibile, e che qualsiasi successivo passo del simplesso ci avrebbe restituito altre soluzioni di base primali ammissibili. In verita, le soluzioni di base trovate possono avere le seguenti combinazioni di ammissibilità su P primale e D duale:

- **Ammissibile su** P **e** D**:** in questo caso siamo all'ottimo dalla dualità forte;
- **Ammissibile su** P **ma non su** D**:** in questo caso siamo su un comune punto ammissibile non ottimo;
- **Non ammissibile su** P **o** D**:** in questo caso si scarta la soluzione;
- **Ammissibile su** D **ma non su** P**:** effettivamente, ancora non abbiamo un modo per gestire questa situazione.

Per questo motivo estendiamo l'algoritmo del simplesso al duale.

10.1 Algoritmo del simplesso duale

Intendiamo mantenere, ad ogni passo, la soluzione di base duale come ammissibile e controllare l'ammissibilità di quella primale. Se la soluzione di base primale è ammissibile (l'inverso di come avevamo visto per il simplesso primale), allora siamo sull'ottimo. Altrimenti, si cambia base, cercando di minimizzare la funzione obiettivo su una nuova soluzione di base duale.

Abbiamo quindi un vertice del poliedro duale, cioè una soluzione di base ammissibile del duale. Ci chiediamo se questo vertice $\bar{y}$ è ottimo. Visto che è vertice del duale, abbiamo che per una matrice A_B e un vettore costo c del primale:

$$\bar{y} = (cA_B^{-1}, 0), \quad \text{con } cA_B^{-1} \geq 0$$

dove ricordiamo questa notazione significa impostare le variabili non di base $\in N$ a 0 e risolvere il sistema in m variabili rimasto.

Adesso possiamo costruire il complementare primale $\bar{x}$, ponendo:

$$\bar{x} = A_B^{-1} b_b$$

e applicare il test di ammissibilità, cioè vedere se:

$$A_N(A_B^{-1} b_B) \leq b_N$$

ergo $\bar{x} \in P$, quindi il complementare duale esiste e il vertice è ottimo. Come prima, se questa condizione risulta verificata possiamo fermarci, in quanto abbiamo trovato la soluzione ottimale.

In caso contrario, avremo $\exists i \in N$ tale che $b_i - A_i(A_B^{-1} - b_B) < 0$. Questo equivale a ciò che avevamo trovato per il primale per quanto riguardava gli indici uscenti, con una sola differenza: per risolvere il duale, si trova **prima** l'**indice entrante**, e **poi** l'**indice uscente**. Vediamo quindi i passaggi in ordine:

- **Indice entrante**
  Abbiamo che, se il vertice non è ottimo, vale:

  $$\exists i \in N \text{ t.c. } b_i - A_i(A_B^{-1} - b_B) < 0$$

  Vogliamo che l'indice entrante sia quell che raggiunge questo valore negativo, quindi:

  Definizione 10.1: Indice entrante duale

  Chiamiamo indice entrante k, da una certa soluzione della base B:

  $$k := \min\{i \in N \text{ t.c. } b_i - A_i\bar{x} < 0\}$$

  Come sempre, $\min$ significa che in caso di più i negativi, si adotta la regola anticiclo (di Bland) di scegliere il primo. In caso di nessun i negativo, la complementare duale esiste e siamo sull'ottimo.

- **Indice uscente**
  Cerchiamo quindi l'indice uscente. Dovremo prima definire la matrice W come $W = -A_B^{-1}$, e calcolare il prodotto (perlopiù analogo al primale, ma si noti il segno della diseguaglianza capovolto):

  $$A_k W^i < 0$$

  Questo ci fornisce una regola, come nel simplesso primale, per l'esistenza di una soluzione: nel caso non sia verificato, si ha che il duale $\to -\infty$ e che il primale è vuoto. In caso contrario, si possono calcolare i rapporti, come:

  $$r_i = \frac{-\bar{y}_i}{A_k W^i}, \quad i \in B, A_k W^i < 0$$

  e scegliere l'indice che da $\vartheta = \min r_i$, cioè:

  Definizione 10.2: Indice uscente duale

  Chiamiamo indice uscente h, da una certa soluzione della base B e un certo indice entrante k:

  $$h := \min\{i \in B \text{ t.c. } A_k W^i < 0, \quad \frac{-\bar{y}_i}{A_k W^i} = \vartheta\}$$

  Anche qui, il $\min$ serve a selezionare il primo indice valido, ed è una regola anticiclo.

Abbiamo quindi revisionato tutti gli strumenti necessari alla formulazione dell'algoritmo del simplesso, stavolta duale:

Algoritmo 2 del simplesso duale

Input: un problema LP in forma primale standard
Output: la soluzione ottima
Trova una base B che genera una soluzione di base duale ammissibile.
ciclo:
Calcola la soluzione di base duale $\bar{y} = (cA_B^{-1}, 0)$ e la soluzione di base duale $\bar{x} = A_B^{-1} b_B$
if $A_N(A_B^{-1} b_B) \leq b_N$ **then**
Fermati, $\bar{y}$ è ottima per D e $\bar{x}$ è ottima per P
else
Calcola l'indice entrante:

$$k := \min\{i \in N \text{ t.c. } b_i - A_i \bar{x} < 0\}$$

poni $W := -A_B^{-1}$ e indica con W^i la i-esima colonna di W
end if
if $A_k W^i \geq 0 \quad \forall i \in B$ **then**
Fermati, $D \to -\infty$ e P non ha soluzione ottima.
else
Calcola:

$$\vartheta = \min\{\frac{-\bar{y}_i}{A_k W^i} \text{t.c.} \quad i \in B, \quad A_k W^i < 0\}$$

e trova l'indice uscente:

$$h := \min\{i \in B \text{ t.c. } A_k W^i < 0, \quad \frac{-\bar{y}_i}{A_k W^i} = \vartheta\}$$

end if
Aggiorna la base come:

$$B := B \setminus \{h\} \cup \{k\}$$

Torna a ciclo

11 Lezione del 09-10-24

11.1 Poliedro vuoto

Fino ad ora abbiamo rimandato la trattazione dell'algoritmo per determinare se un poliedro è vuoto. Diamo adesso quest'algoritmo, notando inoltre che, nel caso il poliedro non fosse vuoto, dovrebbe fornirci un possibile vertice di partenza per l'algoritmo del simplesso.

Partiamo da un poliedro in forma duale standard:

$$\begin{cases} Ax = b \\ x \geq 0 \end{cases}$$

dove m è il numero di variabili e n il numero di vincoli, e quindi $A : m \times n$.

Adottiamo questa forma in quanto siamo sicuri (a differenza della forma primale) che non presenti linealità. Costruiamo quindi quello che viene chiamato **duale ausilia-**

rio, DA:

$$\begin{cases} Ax = b \\ x \geq 0 \end{cases} \rightarrow \begin{cases} \min \sum_{i=1}^{n} \varepsilon_i \\ Ax + I\varepsilon = b \\ x_i \varepsilon \geq 0 \end{cases}$$

dove ε rappresenta un vettore di variabili ausiliarie di dimensione n, cioè una per ogni equazione. Possiamo rappresentare il sistema ottenuto anche attraverso due matrici a blocchi:

$$(A|I) \left(\frac{x}{\varepsilon} \right)$$

Chiamiamo quindi $v(\text{DA})$ la soluzione del duale ausiliario, e formuliamo il teorema:

Teorema 11.1:

Se $v(\text{DA}) > 0$, allora $D = \emptyset$, altrimenti, se $v(\text{DA}) = 0$, $D \neq \emptyset$. Equivale a dire:

$$v(\text{DA}) = 0 \Leftrightarrow D \neq \emptyset$$

Dobbiamo però capire come risolvere il duale ausiliario, magari senza usare questo teorema, in quanto questo si andrebbe a creare una catena infinita di duali ausiliari da risolvere... Notiamo quindi che il duale ausiliario ha sempre una base plausibile, che è quella data dagli indici delle ultime n variabili, quelle introdotte come ε_i.

Non solo, abbiamo anche che:

Teorema 11.2:

La soluzione ottima di (DA), se $D \neq \emptyset$, ci fornisce un vertice di D stesso.

Notiamo che, visto che il primo teorema chiedeva $v(\text{DA}) = 0$ perché $D \neq \emptyset$, allora si ha che nella soluzione ottima di DA nulla (quella che dimostra la non vuotezza del poliedro), le variabili ε_i sono nulle.

Il procedimento sarà quindi:

1. Prendere il problema duale;
2. Ricavare il duale ausiliario inserendo nei vincoli il vettore di n variabili ausiliarie ε;
3. Risolvere il duale ausiliario attraverso il simplesso duale, prendendo come passo iniziale la base data dagli ultimi n indici, cioè che comprende le variabili ausiliarie appena introdotte;
4. Fare $\geq n$ passi al simplesso, aspettandoci che i primi n passi rimuovano tutte le variabili ausiliarie.

Il vertice che ricaviamo dall'algoritmo prende il nome di **soluzione ammissibile di base**.

11.2 Ricavare le variazioni dai cambi di base

Abbiamo che svolgere un passo al simplesso (primale o duale) significa effettuare un cambio di base, con base entrante e base uscente. Possiamo calcolare valori definiti, che la soluzione di base sia ammissibile o meno, per la funzione obiettivo del primale e del duale per qualsiasi base.

Potremmo voler calcolare qual'è la variazione di valore di questa funzione dato un certo cambio di base. Ricordiamo quindi di aver introdotto la formula per spostarci lungo i vincoli di un certo λ sul primale:

$$cx(\lambda) = c\bar{x} - \lambda \bar{y}_h$$

e sul duale:

$$by(\lambda) = b\bar{y} + \lambda \left(b_h - A_k \bar{x}\right)$$

Queste due formule stabiliscono un legame fra il valore della funzione obiettivo e quello di determinate soluzioni $\bar{x}$ e $\bar{y}$ sottoposte a perturbazioni nella direzione dei vertici entranti (cioè allentando il vincolo h) di valore λ. Possiamo quindi usarle per calcolare quanto di chiedevamo all'inizio del paragrafo: la variazione del valore delle funzioni obiettivo del primale e del duale dopo il cambio di base, ricordando che il nostro λ sarà il rapporto r_i (in particolare lo avevamo chiamato ϑ) scelto per determinare il vincolo entrante (primale) o uscente (duale):

$$r_p = \frac{b_k - A_k \bar{x}}{A_k W^h}, \quad r_d = \frac{-\bar{y}_h}{A_k W^h}$$

11.3 Casi degeneri

Concludiamo infine la trattazione della PL notando il motivo dell'uso delle regole anticiclo di Bland nel calcolo degli indici uscenti ed entranti nell'algoritmo del simplesso. Prendiamo un problema di esempio:

$$\begin{cases} \max(x_1 + x_2) \\ x_1 \leq 1 \\ x_2 \leq 1 \\ x_1 + x_2 \leq 2 \\ x_i \geq 0 \end{cases}$$

Si ha che il vertice del poliedro $\bar{x} = (1, 1)$ è degenere: possiamo ottenerlo dalle basi $B = \{1, 2\}$, $B = \{1, 3\}$ e $B = \{2, 3\}$. In questo caso tutto funziona perché il vertice è anche ottimo, in caso contrario potremmo, se non si usassero le regole anticiclo di Bland, finire in un ciclo infinito.

Regole alternative a quella di Bland sono la scelta dell'indice sempre maggiore, o degli scarti o rapporti (apparentemente) ottimi, cioè più piccoli:

$$b_k - A_k \bar{x} = \min_{i \in N} \left(b_i - A_i \bar{x}\right)$$

Queste regole, sopratutto l'ultima, potrebbero sembrare equivalenti o migliori di quelle di Bland. Invece è importante ricordare che potrebbero causare cicli.

12 Lezione del 10-10-24

12.1 Introduzione alla programmazione lineare intera

Iniziamo adesso a studiare un tipo di problemi che finora avevamo menzionato, ma mai risolto. Si pone il seguente:

Problema 12.1: Caricamento

Un ladro è riuscito a scassinare un'importante caveau, e ha di fronte a sé una serie di gemme preziose. Tenendo conto del peso e del tipo delle gemme, ricava una tabella col valore e il peso di ogni gemma:

Valore	2	4	7	9	13	16
Peso	6	8	9	11	15	18

Amesso che il suo zaino non possa portare più di 30 kg, quali gemme dovrà scegliere per avere il maggiore profitto possibile?

Questo problema riprende il celebre *knapsack problem*, che risulta essere NP-completo. Attraverso la **programmazione lineare intera**, possiamo ricavare soluzioni, almeno approssimate.

Si ha che qualsiasi possibilità delle 2^n configurazioni di gemme che il ladro può portare sono rappresentate da un vettore:

$$x = (x_0, ..., x_n), \quad x_i \in \{0, 1\}$$

e che i il valore e il peso delle stesse sono vettori:

$$v = (2, 4, 7, 9, 13, 16)$$

$$p = (6, 8, 9, 11, 15, 18)$$

Possiamo quindi formulare il sistema:

$$\begin{cases} \max\left(2x_1 + 4x_2 + 7x_3 + 9x_4 + 13x_5 + 16x_6\right) \\ 6x_1 + 8x_2 + 9x_3 + 11x_4 + 15x_5 + 18x_6 \leq 30 \\ x \in \{0,1\}^n \end{cases} \rightarrow \begin{cases} \max\left(v^T x\right) \\ p^T x \leq P \\ x \in \{0,1\}^n \end{cases}$$

dove P è il vincolo di massimo dei pesi.

Questo problema ricalca la struttura di un problema LP, ma il vincolo finale, che limita i valori possibili delle componenti del vettore soluzione, lo rende IPL. Decidiamo di trasformare il problema in un LP corrispondente, che chiameremo **rilassamento continuo**, e risolvere quello. Possiamo fare ciò in due modi: il primo prevede di rimuovere semplicemente il vincolo, magari includendo la positività, o ancor meglio l'appartenenza $x \in [0, 1]$:

$$\begin{cases} \max\left(v^T x\right) \\ p^T x \leq P \\ 0 \leq x \leq 1 \end{cases}$$

Questo metodo è sicuramente funzionante, ma preferiamo calcolare il duale, a partire dal primale con imposta la positività:

$$\begin{cases} \max\,(2x_1+4x_2+7x_3+9x_4+13x_5+16x_6) \\ 6x_1+8x_2+9x_3+11x_4+15x_5+18x_6 \leq 30 \\ -x_1 \leq 0 \\ -x_2 \leq 0 \\ -x_3 \leq 0 \\ -x_4 \leq 0 \\ -x_5 \leq 0 \\ -x_6 \leq 0 \end{cases} \rightarrow \begin{cases} \min\,(30y_1) \\ 6y_1 - y_2 = 2 \\ 8y_1 - y_3 = 4 \\ 9y_1 - y_4 = 7 \\ 11y_1 - y_5 = 9 \\ 15y_1 - y_6 = 13 \\ 18y_1 - y_7 = 16 \\ y_i \geq 0 \end{cases}$$

da cui notiamo che le disequazioni in forma $a_i x + s_i = b_i$ si possono rendere come $a_i x \geq b_i$:

$$\begin{cases} \min\,(30y_1) \\ 6y_1 \geq 2 \\ 8y_1 \geq 4 \\ 9y_1 \geq 7 \\ 11y_1 \geq 9 \\ 15y_1 \geq 13 \\ 18y_1 \geq 16 \\ y_1 \geq 0 \end{cases} \equiv \begin{cases} \min\,(30y_1) \\ y_1 \geq \frac{1}{3} \\ y_1 \geq \frac{1}{2} \\ y_1 \geq \frac{7}{9} \\ y_1 \geq \frac{9}{11} \\ y_1 \geq \frac{13}{15} \\ y_1 \geq \frac{8}{9} \\ y_1 \geq 0 \end{cases}$$

Abbiamo che questo sistema è banale: il minimo di $30y_1$ sottoposto ai vincoli si avrà quando y_1 rispetta i vincoli, che sono tutti $\geq$, e rispettare qualsiasi vincolo significa rispettare il vincolo con b_i più grande, quindi in questo caso $\frac{8}{9}$ Si ha quindi che la soluzione ottima del duale è $\frac{8}{9}$, e possiamo trovare con la stessa facilità la soluzione del primale: abbiamo che la disequazione da cui si è ricavato la soluzione del duale ha indice $j = 6$. Si prende quindi la variabile con lo stesso indice nel primale, e si **satura**, cioè si porta al livello più alto possibile prima di violare il limite. Si nota che il valore ottimo trovato nel primale e nel duale (che è uguale dalla dualità forte) è $v = P\bar{y}_j = v_j \bar{x}_j$, con $\bar{x}$ e $\bar{y}$ le soluzioni ottime trovate rispettivamente al primale e al duale.

Nel nostro caso, $18x_6 \leq 30$, quindi $x_6 = \frac{30}{18}$, con valore $v = 16\frac{30}{18} = \frac{80}{3}$.

Questo approccio sarebbe quello che si applicherebbe intuitivamente senza sapere nulla di PL. Infatti i rapporti calcolati nel duale:

$$r_i = \frac{v_i}{p_i}$$

sono effettivamente i **rendimenti** di ogni elemento di peso e valore, cioè quanto valore portano in rapporto al loro peso. Chiaramente, vorremmo massimizzare gli elementi con rendimento maggiore, quindi saturiamo gli elementi con r_i massimo. Possiamo quindi definire il seguente algoritmo:

Algoritmo 3 caricamento per soluzioni ottime del rilassato continuo

Input: il rilassato continuo di un problema di caricamento
Output: la soluzione ottima $\bar{x}$
Considera i $r_i = \frac{v_i}{p_i}$ rendimenti di ogni elemento
Scegli l'indice j del rendimento massimo r_j
Satura la variabile j

Abbiamo che la soluzione trovata è un limite superiore per il problema di partenza, quello ILP: infatti, rilassando i vincoli, otteniamo una regione ammissibile più grande, e quindi aumentiamo il massimo. In verità anche l'arrotondamento in basso della soluzione v è un limite superiore: nel caso del problema abbiamo $\lfloor \frac{80}{3} \rfloor = 26$.

Un'idea banale per trovare il punto ottimo di questo problema ILP potrebbe essere quello di arrotondare il punto ottimo del rilassato continuo: nel nostro esempio, troviamo $\bar{x} = (0, 0, 0, 0, 0, 1)$, che dà massimo $v = 16$.

Notiamo un risultato fondamentale: se chiamiamo il valore arrotondato della soluzione del rilassato continuo v_s, e il valore ottenuto nell'approssimazione del *punto* di ottimo v_i, è vero che:

$$v_i \leq v_{ILP} \leq v_p$$

dove v_{ILP} è la soluzione del problema ILP che stiamo cercando. Per la precisione, possiamo definire l'errore:

$$\varepsilon = \frac{v_s - v_i}{v_i}$$

che sul problema in esame dà $\varepsilon \approx 60\%$, chiaramente poco utile. Sarà necessario adottare un qualche metodo per migliorare la stima data dall'arrotondamento del punto di ottimo del rilassato continuo.

Per adesso, generalizziamo quando trovato finora: bisogna distinguere fra v_i e v_s che si parli di problemi di minimizzazione o di massimizzazione. Si ha infatti che:

Teorema 12.1: Bound di problemi ILP

Per un dato problema ILP con soluzione v_{ILP}, si ha che:

$$v_i \leq v_{PLI} \leq v_s$$

- Se il problema è di **massimizzazione**, v_i è il valore calcolato soluzione ammissibile del rilassato continuo arrotondata **per difetto**, e v_s è la soluzione ottima del rilassato continuo arrotondata **per difetto**;
- Se il problema è di **minimizzazione**, v_i è la soluzione ottima del rilassato continuo arrotondata **per eccesso**, e v_i è il valore calcolato soluzione ammissibile del rilassato continuo arrotondata **per eccesso**;

12.1.1 Vincoli booleani

Possiamo chiamare il vincolo introdotto prima, $x \in \{0, 1\}^n$, **vincolo booleano** (in modo più o meno informale). Come abbiamo detto, questo vincolo può essere reso come $x_i \geq 0$, o ancor meglio $0 \geq x_i \geq 1$ quando si porta al rilassato continuo (abbiamo usato il primo insieme di vincoli per ricavare il duale del rilassato continuo).

Introduciamo quindi l'algoritmo, equivalente a quello presentato prima per :

Algoritmo 4 caricamento per soluzioni ottime del rilassato continuo con vincolo booleano

Input: il rilassato continuo di un problema con vincolo booleano
Output: la soluzione ottima $\bar{x}$
Considera i $r_i = \frac{v_i}{p_i}$ rendimenti di ogni elemento
Scegli l'indice j del rendimento massimo r_j
Satura la variabile j
Quando finisci lo spazio, satura con il bene rimanente (sic.)