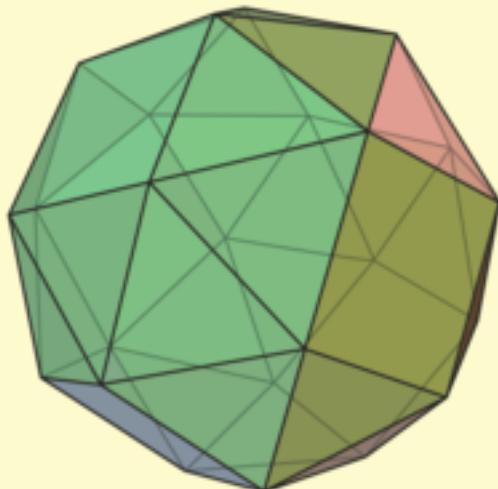


FABIO SCHOEN

Fondamenti di Ricerca Operativa



Fabio Schoen

Fondamenti di
Ricerca Operativa

www.lulu.com
id: 12502182

Fabio Schoen
Dipartimento di Sistemi e Informatica
Università degli Studi di Firenze
via di Santa Marta, 3
50139 Firenze (Italy)
E-mail: fabio.schoen@unifi.it
Web: <http://gol.dsi.unifi.it>

ISBN: 978-1-105-49496-3
© 2012 Fabio Schoen

Indice

Indice	iii
1 Introduzione	1
I Fondamenti di Ottimizzazione Lineare	3
2 Introduzione	5
2.1 Esempi elementari	5
Modelli di miscelazione	5
Il modello della dieta	11
Il modello dell'abbinamento	17
3 Problemi di ottimizzazione	21
3.1 Introduzione e definizioni principali	21
3.2 Modelli lineari	23
Forma standard di un problema di programmazione lineare	26
Trasformazione di programmi lineari in forma standard	28
3.3 Equivalenza tra problemi di ottimizzazione	36
4 Teoria della Programmazione Lineare	49

INDICE

4.1	Basi e soluzioni di base nella Programmazione Lineare	49
4.2	Il teorema fondamentale della programmazione lineare	53
5	Geometria della programmazione lineare	65
5.1	Soluzione grafica di problemi di <i>PL</i>	77
6	Il metodo del simplex	83
6.1	Formulazione matriciale	83
	Ottimalità di una soluzione ammissibile di base	84
	Scelta di una nuova soluzione ammissibile di base - operazione di pivot	90
	Determinazione della variabile uscente di base	93
	Scelta della variabile entrante in base	100
	Scelta del vettore uscente dalla base	102
	Soluzioni di base e vertici degeneri	106
	Finitezza del metodo del simplex	108
6.2	Presentazione dell'algoritmo del simplex mediante dizionario	113
6.3	Inizializzazione del metodo del simplex	120
	Il metodo due fasi	120
	Il metodo "big M"	128
6.4	Problemi con infiniti ottimi	130
6.5	Efficienza del metodo del simplex	139
II	Teoria della dualità	145
7	Introduzione alla dualità	147
7.1	Introduzione	147
7.2	Rilassamento e duale lagrangiano	149
	Introduzione alla dualità per la generazione di limiti inferiori	154
8	Dualità lineare	159

8.1	Duale di un generico problema di <i>PL</i>	159
8.2	Teoremi di dualità	165
	Un'applicazione della dualità alla programmazione lineare	178
8.3	Informazione duale nel dizionario del simplesso . . .	179
8.4	Interpretazione della dualità	184
	Duale del problema della dieta	184
	Duale del problema di miscelazione ottimale	188
	Duale del problema di "mix" di produzione	193
8.5	Dualità e teoria dei giochi	197
9	Il metodo del simplesso duale	203
9.1	Il metodo del simplesso duale: formulazione matriciale	205
9.2	Il metodo del simplesso duale: sviluppo mediante dizionario	211
10	Analisi di sensitività	215
10.1	Introduzione	215
10.2	Sensitività sul termine noto	218
10.3	Sensitività sul vettore dei costi	227
	Variazione del costo di una variabile non di base . .	227
	Variazione del costo di una variabile di base . . .	228
10.4	Aggiunta di una nuova variabile	232
10.5	Sensitività sulla matrice dei coefficienti	235
	Modifica di una colonna non di base	235
	Modifica di una colonna di base	235
10.6	Aggiunta di un nuovo vincolo	236
10.7	Lettura dell'output dei codici di programmazione li- neare	238
	Problema della dieta	238

INDICE

III Ottimizzazione di flussi su reti	249
11 Introduzione alla teoria dei grafi ed alle reti di flusso	251
11.1 Terminologia e definizioni principali	251
11.2 Alberi	260
11.3 Flussi su reti – reti di flusso	265
Basi e soluzioni di base nei problemi di flusso	267
12 L'algoritmo del simplex su reti	277
12.1 Duale del problema di flusso su reti a costo minimo .	279
12.2 Determinazione di una soluzione di base	281
12.3 Determinazione della soluzione complementare del duale	289
12.4 Calcolo dei coefficienti di costo ridotto	294
12.5 Determinazione dell'arco uscente di base	295
12.6 Ricalcolo della soluzione del duale	300
12.7 Inizializzazione del metodo del simplex su reti .	305
13 Il problema del cammino di costo minimo	309
13.1 Algoritmo di Dijkstra	311
13.2 Applicazione del simplex su reti al problema del percorso di costo minimo	324
Il metodo della “grande M ”	325
14 Il problema del massimo flusso	333
14.1 Algoritmo di Ford e Fulkerson	333
Finitezza dell'algoritmo di Ford & Fulkerson	353
14.2 Dualità nei problemi di massimo flusso	362
Il teorema flusso massimo - sezione minima	366
14.3 Reti con capacità minima	370

IV Programmazione Lineare Intera	373
15 Introduzione alla Programmazione Lineare Intera	375
15.1 Introduzione	375
15.2 Connessioni tra <i>PL</i> e <i>PL</i> intera.	379
15.3 Metodi di taglio	394
15.4 Il metodo Branch & Bound	405
Metodo Branch & Bound per problemi di ottimizzazione combinatoria	417
Bibliografia	427
Indice analitico	431

1

Introduzione

1. INTRODUZIONE

Questo volume contiene una raccolta degli appunti delle lezioni di Fondamenti di Ricerca Operativa, tenute per diversi anni presso la Facoltà di Ingegneria dell’Università degli Studi di Firenze. Il materiale, presente da tempo sotto forma di dispense, è stato utilizzato principalmente per corsi introduttivi alla Ricerca Operativa e, in particolare, alla teoria ed ai metodi fondamentali dell’ottimizzazione lineare. Non vengono trattati argomenti avanzati, quali gli algoritmi a punto interno o le tecniche più recenti di ottimizzazione lineare intera, destinati a corsi di livello più avanzato. Nei volumi successivi di questa serie verranno presentati i modelli e le tecniche di modellazione, utili per capire la vastità e la rilevanza delle applicazioni dei modelli lineari e dei modelli contenenti variabili a valori discreti. Successivamente dovrebbe apparire un volume dedicato ai modelli ed agli algoritmi specializzati per alcune applicazioni gestionali avanzate, quali i modelli di scorte, di programmazione della produzione, di instradamento veicoli, ecc. Tutto il materiale didattico che ho in mente di pubblicare in futuro fa in qualche modo riferimento alle basi contenute in questo libro.

Una pagina¹ del sito <http://gol.dsi.unifi.it> sarà dedicata a questo volume e conterrà errata–corrige, integrazioni, esempi e aggiornamenti del materiale qui presentato.

Nonostante molto del materiale qui contenuto sia stato usato da un gran numero di miei allievi, è inevitabile che molti errori già presenti non siano stati notati e che altri, nuovi, siano comparsi in questa edizione. Sarò molto grato a chiunque vorrà segnalarmi errori, imprecisioni, punti poco chiari, . . . Questo volume non sarebbe mai apparso se non avvessi avuto i contributi e l’aiuto dei miei tanti allievi: a loro ed agli allievi futuri dedico queste mie pagine.

Firenze, 15 Febbraio 2012.

¹gol.dsi.unifi.it/lab/it/content/fondamenti-di-ricerca-operativa

Parte I

Fondamenti di Ottimizzazione Lineare

Introduzione

2.1 Esempi elementari

Modelli di miscelazione

Uno dei più semplici modelli di ottimizzazione, descritto in qualunque testo introduttivo di Ricerca Operativa, è il cosiddetto modello di miscelazione, presentato a volte sotto forma di modello per la formulazione di una dieta; un modello per la pianificazione di dieci di costo minimo con vincoli di tipo nutrizionale è stato uno dei primi modelli di ottimizzazione risolti con il metodo del simplesso. Tale modello venne risolto a mano da un gruppo di ricercatori, e richiese parecchie ore di calcolo “parallelo”: è forse inutile osservare che quel modello, basato su una proposta del premio Nobel per l'economia del 1982 George Stigler (Stigler, 1945; Dantzig, 1990; Garille & Gass, 2001), può essere risolto in una frazione di secondo su un modesto Personal Computer.

Nella sua forma base, il modello di miscelazione corrisponde al problema di decidere come comporre una miscela di costo minimo e qualità garantita, ottenuta da diverse materie prime. Uno dei campi industriali nei quali il modello di miscelazione trova vasta applicazione è quello della produzione di alluminio da rottami. Si immagini di poter acquistare, in quantitativi limitati, alcuni rottami

2. INTRODUZIONE

costituiti in massima parte da alluminio, ma disponibili in lega con altri elementi chimici, e di volerli miscelare in modo tale da ottenerne, tramite fusione, un materiale che contenga quantitativi prefissati dei vari elementi chimici. Naturalmente, per correggere la qualità dei rottami disponibili, è sempre possibile acquistare, in quantità teoricamente illimitate, ma ad un prezzo sensibilmente più alto, elementi chimici puri. Il problema diventa quindi quello di stabilire le quantità di rottami e di materiali puri da impiegare in una miscela che rispetti i vincoli imposti sulla qualità e che risulti di costo minimo.

Nella tabella seguente viene riportato il file dati, preparato secondo la sintassi del linguaggio di modellizzazione algebrica AMPL (Fourer, Gay, & Kernighan, 2002), per un ipotetico problema di miscelazione di alluminio da rottami. L'obiettivo è quello di ottenere una miscela secondo specifiche qualitative prefissate e note nell'ambiente della produzione di alluminio con il codice "6063".

```
# problema di miscelazione ottimale

set MATERIALI := Scarti_ALMC Scarti_KAC Rottami Al Si Mg;

set ELEMENTI := Si Mg Fe Cu Mn Zn Cr Ti Al Altri;

param qta_rich := 4500; # kg

param:    rich_min   rich_max   :=
Si          0.2        0.6
Mg          0.45       0.9
Fe          .           0.35
Cu          .           0.1
Mn          .           0.1
Zn          .           0.1
Cr          .           0.1
Ti          .           0.1
Al          96.9       100.0
Altri        .           0.15
;
```

2.1. Esempi elementari

```
param compos:
  Scarti_ALMC Scarti_KAC Rottami Al Si Mg := 
Si      0.50      0.50     0.30   . 100   .
Mg     0.75      0.70     0.50   .   . 100
Fe     0.20      0.20     0.35   .   .   .
Cu     0.01      0.01     0.05   .   .   .
Mn     0.02      0.02     0.05   .   .   .
Zn     0.02      0.02     0.05   .   .   .
Cr     0.02      0.02     0.05   .   .   .
Ti     0.02      0.02     0.05   .   .   .
Al    97.00     97.00    90.00  100   .   .
Altri  0.06      0.06     0.77   .   .   .
;

param:      dispon      costo :=
# costo in Euro per kg
Scarti_ALMC 500      1.230
Scarti_KAC 1200      1.230
Rottami     2200      1.230
Al          .          2.140
Si          .          1.300
Mg          .          2.442;
```

La prima parte del file dati contiene la definizione dei nomi dei materiali disponibili e degli elementi chimici di interesse. Segue la specifica della quantità totale di materiale da produrre (in kg) e, per ciascun elemento chimico, il valore della percentuale minima e massima richiesta, in base alle specifiche di definizione della miscela “6063” (il valore di default per la richiesta minima è pari a 0). Come si vede dal listato, viene richiesta una miscela di alluminio contenente almeno una piccola frazione di Silicio e Manganese. Successivamente, nel file dati vengono riportate le composizioni chimiche dei vari materiali ed infine, per ciascun materiale, il quantitativo massimo disponibile ed il costo in Euro per kg.

Per la formalizzazione del problema occorre individuare le variabili di decisione, i vincoli, l’obiettivo. Per le variabili di decisione è naturale associare una incognita x_j a ciascun materiale j ; il vincolo relativo alla quantità totale da produrre si può esprimere mediante

2. INTRODUZIONE

l'equazione

$$\sum_{j \in \text{MATERIALI}} x_j = \text{qta_rich.}$$

Per il bilancio chimico, se si indica con $\text{compos}_{i,j}$ la percentuale di elemento chimico i presente nel materiale j , la quantità percentuale dell'elemento chimico i presente nella miscela sarà data da

$$\sum_{j \in \text{MATERIALI}} \text{compos}_{i,j} \cdot \frac{x_j}{\text{qta_rich}}$$

e, pertanto, i vincoli che esprimono le richieste minime e massime di elementi chimici presenti nella miscela assumono la forma

$$\text{rich_min}_i \leq \sum_{j \in \text{MATERIALI}} \text{compos}_{i,j} \cdot \frac{x_j}{\text{qta_rich}} \leq \text{rich_max}_i.$$

Infine, per ciascun materiale j , dovrà essere imposta la non negatività e, per i materiali disponibili in quantità limitata, un limite superiore:

$$0 \leq x_j \leq \text{dispon}_j.$$

L'obiettivo, dato un costo unitario costo_j associato a ciascun materiale, sarà

$$\min \sum_{j \in \text{MATERIALI}} \text{costo}_j x_j.$$

Riassumendo, il modello completo della miscelazione di costo mi-

2.1. Esempi elementari

nimo ha la forma seguente:

$$\begin{aligned} \min_x \quad & \sum_{j \in \text{MATERIALI}} \text{costo}_j x_j \\ \sum_{j \in \text{MATERIALI}} \text{compos}_{ij} \cdot \frac{x_j}{\text{qta_rich}} & \leq \text{rich_max}_i \quad \forall i \\ \sum_{j \in \text{MATERIALI}} \text{compos}_{ij} \cdot \frac{x_j}{\text{qta_rich}} & \geq \text{rich_min}_i \quad \forall i \\ \sum_{j \in \text{MATERIALI}} x_j & = \text{qta_rich} \\ x_j & \leq \text{dispon}_j \quad \forall j \\ x_j & \geq 0 \quad \forall j. \end{aligned}$$

Il modello, in linguaggio AMPL, può essere implementato come riportato nel file seguente, dove si può notare la forte somiglianza tra il modello matematico e la sua implementazione in AMPL.

```
#-----#
# problema di miscelazione ottimale #
#-----#
#-----#
# dichiarazione insiemi
#-----#
set MATERIALI;
# materiali disponibili

set ELEMENTI;
# elementi chimici

#-----#
# dichiarazione parametri
#-----#
```

2. INTRODUZIONE

```
param qta_rich >=0 ;
# qta' richiesta

param dispon{MATERIALI}, default Infinity;
# disponibilita' dei materiali

param compos {ELEMENTI, MATERIALI} default 0, >=0, <=100;
# composizione chimica dei materiali

param rich_min {ELEMENTI} >= 0 default 0;
# richieste minime in percentuale

param rich_max {el in ELEMENTI} <= 100, >= rich_min[el];
# richieste massime

param costo {MATERIALI};

#-----#
# controllo
#-----#

check {mat in MATERIALI}: 0 < sum {el in ELEMENTI}
compos[el,mat] <= 100;

# -----#
# dichiarazione variabili e limiti inferiori e superiori
# -----#

var qta {mat in MATERIALI} >= 0, <= dispon[mat];

# -----#
# definizione obiettivo
# -----#

minimize costo_tot:
    sum {mat in MATERIALI} costo[mat] * qta[mat];

# -----#
# definizione vincoli
# -----#

subject to totale:
    sum {mat in MATERIALI} qta[mat] = qta_rich;

subject to def_qualita {el in ELEMENTI}:
rich_min[el] <=
sum {mat in MATERIALI} (qta[mat]/qta_rich) * compos[el,mat]
<=rich_max[el];

# la qta e' assoluta
# divisa per qta_rich fornisce qta percentuale;
```

Con i dati del file sopra riportati, utilizzando un opportuno

2.1. Esempi elementari

risolutore, si ottiene la seguente soluzione di costo minimo:

Al	2041.76
Mg	4.38
Si	9.70
Rottami	744.16
Scarti_ALMC	500.0
Scarti_KAC	1200.0

ed il costo, in Euro, risulta pari a 7398.988. Per quanto riguarda la qualità della miscela ottenuta, si può far riferimento alla tabella seguente:

	rich_min	miscola	rich_max
Al	96.9	96.9	100
Cr	0	0.0158	0.1
Cu	0	0.0120	0.1
Fe	0	0.1334	0.35
Mg	0.45	0.45	0.9
Mn	0	0.0158	0.1
Si	0.2	0.4542	0.6
Ti	0	0.0158	0.1
Zn	0	0.0158	0.1
Altri	0	0.15	0.15

Il modello della dieta

Il modello della dieta può essere visto come una variante del modello di miscelazione; il modello classico prevede la determinazione della dieta, o della miscela di cibi, di costo minimo soggetta a vincoli sul contenuto nutritivo. Tali vincoli, come nel modello di miscelazione, possono prevedere soglie minime (per gli elementi nutritivi vantaggiosi) e/o soglie massime (per quelli nocivi). L'unica differenza sostanziale tra il modello elementare di miscelazione e quello

2. INTRODUZIONE

di dieta consiste nel fatto che in quest'ultimo non ha in genere senso imporre vincoli sulla quantità totale; un'altra differenza è il fatto che nei problemi di dieta i contenuti nutritivi non sono espressi generalmente in percentuale, ma in unità di misura differenti, quali grammi, ad esempio per le proteine, i grassi ed i carboidrati, o calorie, nel caso del valore energetico. Spesso le etichette sulle confezioni di cibo riportano il valore riferito a 100g di prodotto, oppure il valore riferito ad una porzione tipica: in questi casi nel formulare il modello occorrerà prestare attenzione al significato delle variabili, che deve essere scelto coerentemente ai dati nutrizionali utilizzati. Spesso, almeno per alcuni componenti nutritivi quali le vitamine, si utilizza un'unità di misura detta percentuale del fabbisogno medio giornaliero: con questo termine si usa indicare la quantità che mediamente un individuo adulto dovrebbe assumere giornalmente in una dieta equilibrata. Quest'ultima unità di misura, pur avendo il nome di percentuale, deve essere trattata come una normale unità di misura: quando si indica ad esempio che una porzione di un dato cibo fornisce il 30% del fabbisogno giornaliero di ferro, significa che 2 porzioni forniranno il 60% e 4 porzioni il 120%. Il modello elementare della dieta può quindi essere presentato nel modo seguente, utilizzando una terminologia che sia la più vicina possibile a quella utilizzata per il modello di miscelazione:

$$\begin{aligned} \min_x \quad & \sum_{j \in \text{CIBI}} \text{costo}_j x_j \\ \sum_{j \in \text{CIBI}} \text{compos}_{i,j} x_j & \leq \text{rich_max}_i \quad \forall i \\ \sum_{j \in \text{CIBI}} \text{compos}_{i,j} x_j & \geq \text{rich_min}_i \quad \forall i \\ x_j & \geq 0 \quad \forall j. \end{aligned}$$

In questo modello sono stati omessi anche i limiti superiori alla quantità acquistabile per ogni cibo, in quanto si può immaginare

2.1. Esempi elementari

che la disponibilità di cibo sia sostanzialmente illimitata per una dieta singola, cioè che le quantità di cibo acquistabili in base ai vincoli nutrizionali siano sempre reperibili sul mercato. In alcuni casi è utile però inserire tali vincoli non tanto per riprodurre la situazione di cibi difficilmente reperibili sul mercato, quanto piuttosto per limitare la quantità di cibo di un dato tipo presente nella dieta. Ad esempio, in una dieta pianificata per un'intera giornata, si potrebbe richiedere che per molti, o forse tutti, i cibi il numero massimo di porzioni sia pari a 1 oppure a 2. A titolo di esempio si riporta un file dati AMPL con alcune specifiche nutrizionali (naturalmente l'esempio ha il solo scopo di illustrare la metodologia di modellizzazione e non deve essere considerato come un esempio realistico di modello di dieta, nè si deve considerare la dieta risultante dal modello come consigliabile).

```
set NUTR := Proteine Vit-A Vit-C VitB1  
                  VitB2 Niacina Calcio Ferro Sodio ;  
set CIBO := Hamburger McLean BigMac PatateFritte PolloFritto  
                  Miele ChefSalad InsalataVerde UovoMcMuffin Wheatus  
                  Vaniglia Latte SuccoArancia SuccoPompelmo SuccoMela;  
  
param      costo  :=  
Hamburger    0.59  
McLean        1.79  
BigMac        1.65  
PatateFritte   0.68  
PolloFritto    1.56  
Miele          0.00  
ChefSalad     2.69  
InsalataVerde 1.96  
UovoMcMuffin   1.36  
Wheatus        1.09  
Vaniglia       0.63  
Latte          0.56  
SuccoArancia   0.88  
SuccoPompelmo  0.68  
SuccoMela      0.68  
;
```

2. INTRODUZIONE

```
param:  
    rich_min rich_max :=  
Proteine      55      .  
Sodio         0       3000  
Vit-A        100     .  
Vit-C        100     .  
VitB1        100     .  
VitB2        100     .  
Niacina      100     .  
Calcio        100     .  
Ferro         100     . ;  
  
param compos:  
    Calorie Proteine Grassi Sodio Vit-A Vit-C VitB1 :=  
Hamburger    255     12      9   490     4     4   20  
McLean        320     22     10   670     10    10   25  
BigMac        500     25     26   890     6     2   30  
PatateFritte  220     3      12   110     0     15   10  
PolloFritto   270     20     15   580     0     0     8  
Miele          45      0      0     0     0     0     0  
ChefSalad    170     17      9   400     100    35   20  
InsalataVerde 50      4      2     70     90    35   6  
UovoMcMuffin  280     18     11   710     10     0   30  
Wheaties      90      2      1   220     20    20   20  
Vaniglia      105     4      1     80     2     0     2  
Latte          110     9      2   130     10     4     8  
SuccoArancia  80      1      0     0     0    120   10  
SuccoPompeLmo 80      1      0     0     0    100    4  
SuccoMela     90      0      0     5     0     2     2  
  
:           # continuazione tabella  
  
    VitB2 Niacina Calcio Ferro :=  
  
Hamburger    10      20     10     15  
McLean        20      35     15     20  
BigMac        25      35     25     20  
PatateFritte  0       10     0      2  
PolloFritto   8       40     0      6  
Miele          0       0      0      0  
ChefSalad    15      20     15     8  
InsalataVerde 6       2      4      8  
UovoMcMuffin  20      20     25     15  
Wheaties      20      20     2      20  
Vaniglia      10      2      10     0  
Latte          30      0      30     0  
SuccoArancia  0       0      0      0  
SuccoPompeLmo 2       2      0      0  
SuccoMela     0       0      0      4 ;
```

2.1. Esempi elementari

Un modello che rappresenta il problema della dieta sopra descritto è il seguente:

```
set NUTR;    # elementi nutritivi
set CIBO;

param costo {CIBO} >= 0;
param qta_min {CIBO} >= 0, default 0;
param qta_max {j in CIBO} >= qta_min[j], default +Infinity;
# quantita' minime e massime di ciascun cibo nella dieta - default 0 e
# infinito rispettivamente

param rich_min {NUTR} >= 0, default 0;
param rich_max {NUTR} >= 0, default +Infinity;
# richieste minime e massime per i singoli elementi nutritivi

param compos {CIBO,NUTR} >= 0;

var Qta {j in CIBO} >= qta_min[j], <= qta_max[j];

minimize costo_totale: sum {j in CIBO} costo[j] * Qta[j];

subject to dieta {i in NUTR}:
   rich_min[i] <= sum {j in CIBO} compos[j,i] * Qta[j] <= rich_max[i];
```

L'esecuzione di un algoritmo risolutore per questo modello produce il seguente risultato: la dieta minima risulta avere un costo totale di 6.12605 (i costi sono espressi in Euro) e risulterebbe composta come segue:

cibo	qta
Hamburger	5.299
Insalata verde	0.535
Latte	1.441
Succo Pompelmo	0.381
'Wheaties'	0.812

2. INTRODUZIONE

ed i contenuti nutrizionali risultano i seguenti:

Elemento nutritivo	qtà
Calcio	100
Ferro	100
Niacina	124.047
Proteine	80.708
Sodio	3000
Vitamina A	100
Vitamina C	100
Vitamina B1	138.48
Vitamina B2	116.44

Come si vede immediatamente, la dieta ottenuta dal risolutore, sebbene corretta, non è utilizzabile in pratica, per vari motivi. Ad esempio, appare improponibile una dieta che presenti un numero di porzioni molto elevate – più di 5 – per un singolo cibo (ricordando che i dati inseriti corrispondono alla pianificazione di una dieta per un giorno). Inoltre la varietà di cibi è abbastanza limitata: sono stati scelti dal risolutore solo relativamente pochi cibi, rispetto al totale disponibile. Infine, dovendo programmare una dieta per un'unica persona, occorrerebbe evitare la presenza di valori decimali nella soluzione. Per quanto riguarda la limitazione del numero totale di porzioni assunte nel corso di un'unica giornata, nel modello è sufficiente specificare un dato diverso da quello di default per il parametro `upp_bnd`; imponendo ad esempio un limite superiore pari a 2 porzioni per ogni cibo si ottiene la soluzione seguente:

2.1. Esempi elementari

cibo	qta
Hamburger	2
InsalataVerde	0.265823
Latte	0.949367
SuccoPompelmo	0.388987
UovoMcMuffin	1.85823
Wheaties	2

con un costo pari a 7.204€ - come si può facilmente intuire, l'aumento di vincoli comporta un aumento di prezzo. Si nota anche come l'inserimento del vincolo relativo al numero massimo di porzioni ha portato ad una variazione non solo nelle quantità dei cibi, ma anche nella composizione della dieta ottimale. Aggiungendo il vincolo di interezza sulle variabili, si ottiene la soluzione

cibo	qta
BigMac	1
Hamburger	2
Insalata Verde	1
Latte	2
Patate Fritte	1
Wheaties	2

di costo 8.77€. Si noti come la soluzione intera non sia un arrotondamento della soluzione decimale – compaiono cibi nuovi ed alcuni (Latte) assumono valori molto diversi da quelli ottenuti precedentemente.

Il modello dell'abbinamento

Un modello piuttosto lontano dai due precedentemente esposti è il seguente: si consideri la situazione in cui un certo numero n di progetti (ad esempio, lo sviluppo di moduli di un progetto software)

2. INTRODUZIONE

devono essere affidati ad un ugual numero n di progettisti o di sviluppatori. Ogni progettista deve svolgere esattamente un progetto – si tratta quindi di mettere in corrispondenza biunivoca i progetti ed i progettisti (esistono varianti del modello nelle quali non tutti i progetti devono essere svolti o altre nelle quali qualche progettista può svolgere più di un progetto). Per addestrare il progettista i a realizzare il progetto j occorre un periodo di addestramento il cui costo è stimato pari a c_{ij} . Il problema di abbinamento consiste nel determinare l'accoppiamento progetto/progettista che rende minimo il costo totale di addestramento.

Per questo modello è comodo utilizzare delle variabili decisionali di tipo binario, δ_{ij} , con la convenzione che $\delta_{ij} = 1$ sta ad indicare che il progettista i svolge il progetto j e 0 altrimenti. Con questa convenzione il costo totale di addestramento si esprime facilmente come

$$\sum_{i,j} c_{ij} \delta_{ij}.$$

Il vincolo che esprime la necessità che ogni progetto sia svolto esattamente da un progettista, ricordando che le variabili δ sono binarie, si esprime come

$$\sum_i \delta_{ij} = 1 \quad \forall j \in \text{Progetti}$$

(una somma di variabili binarie vale uno se e solo se esattamente una di loro è pari ad 1), mentre il fatto che ogni progettista svolga esattamente un progetto si esprime simmetricamente come

$$\sum_j \delta_{ij} = 1 \quad \forall i \in \text{Progettisti}$$

Si vedrà più avanti che questo modello, nonostante richieda l'interesse delle variabili, può essere risolto trascurando questo vincolo.

2.1. Esempi elementari

In tutti gli esempi presentati in questo paragrafo, si possono vedere elementi comuni. Ogni modello prevede la definizione di 3 componenti principali: un insieme di variabili di decisione, o incognite, uno o più vincoli, espressi generalmente sotto forma di equazioni e/o disequazione, una funzione obiettivo da massimizzare o minimizzare. La maggior parte dei vincoli visti nei modelli qui presentati è costituita da equazioni e disequazioni lineari; naturalmente esistono molti modelli nei quali i vincoli si esprimono come equazioni o disequazioni non lineari. Un esempio particolare si è visto già nei modelli presentati, quando, ad esempio nel modello della dieta, si è richiesta l'interezza delle variabili: questo è un vincolo complesso, non esprimibile come vincolo lineare, la cui presenza spesso porta ad un'enorme crescita della complessità di risoluzione del problema.

Naturalmente, oltre alle 3 componenti viste, ogni modello prevederà la specifica di parametri, cioè dei coefficienti che compaiono in tutti i vincoli e nell'obiettivo. Sistemi di modellizzazione algebrica quali AMPL o anche sistemi basati su fogli elettronici prevedono una separazione tra le specifiche descrittive del modello ed i dati. Questa separazione, maggiore nei linguaggi algebrici che nei fogli elettronici, riduce la possibilità di errori, separando la fase di progettazione del modello, da effettuarsi a cura di un esperto nella formulazione di problemi di decisione, dalla fase di specifica dei dati di un esempio particolare (quella che con un orribile inglese-smo viene detta "istanziazione"), operazione che può essere effettuata da qualcuno non particolarmente esperto nella formulazione di modelli. Un terzo componente, separato sia dal modello che dai dati, è il risolutore, un algoritmo in grado di determinare una soluzione ad un problema di decisione rappresentato formalmente mediante modello e dati.

Problemi di ottimizzazione

In questo capitolo si introdurrà il concetto di problema di ottimizzazione e si forniranno alcune definizioni di base. Verrà poi definito un concetto di equivalenza tra problemi, utile per poter ricondurre famiglie apparentemente differenti di problemi ad un unico schema comune.

3.1 Introduzione e definizioni principali

Un generico *problema di ottimizzazione* P è dato da una terna

$$\langle \text{Opt}, X, f \rangle$$

e si indica con

$$P : \text{Opt}_{x \in X} f(x) \quad (3.1)$$

Opt può indicare l'operatore \min oppure l'operatore \max , $X \subseteq \mathbb{R}^n$ è detto *insieme ammissibile* ed è un sottoinsieme dello spazio euclideo n -dimensionale, con $n \in \mathbb{N}$, mentre f è una funzione reale $f : X \rightarrow \mathbb{R}$ detta *funzione obiettivo*.

Definizione 1. Il *problema di ottimizzazione* P si dice *non ammissibile* o *inammissibile* se $X = \emptyset$.

3. PROBLEMI DI OTTIMIZZAZIONE

Definizione 2. Un vettore $x \in \mathbb{R}^n$ si dice *soluzione ammissibile* per il problema P se $x \in X$.

Definizione 3. Un problema di ottimizzazione P di minimo si dice *illimitato* (inferiormente) se esiste almeno una successione $\{x_k\}_{k=1}^{\infty}$ di soluzioni ammissibili tale che

$$\lim_{k \rightarrow \infty} f(x_k) = -\infty.$$

Un problema di ottimizzazione P di massimo si dice *illimitato* (superiormente) se esiste almeno una successione $\{x_k\}_{k=1}^{\infty}$ di soluzioni ammissibili tale che

$$\lim_{k \rightarrow \infty} f(x_k) = +\infty.$$

Definizione 4. Un problema di ottimizzazione P di minimo *ammette ottimo finito* se e solo se

$$\exists x^* \in X : f(x^*) \leq f(x) \quad \forall x \in X. \quad (3.2)$$

Analogamente un problema P di massimo ammette ottimo finito se e solo se

$$\exists x^* \in X : f(x^*) \geq f(x) \quad \forall x \in X. \quad (3.3)$$

Ogni punto x^* che soddisfi la (3.2) per un problema di minimo o la (3.3) per uno di massimo, si dice *ottimo globale* per il problema P .

Spesso si sottintende la parola *globale* e si parla semplicemente di “ottimo” di un problema. L’insieme degli ottimi di P si indica con

$$\arg \text{opt}P.$$

In formule, dato un problema, ad esempio di minimo, $P = \langle \min, X, f \rangle$, si ha

$$\arg \min P = \{x \in X : f(x) \leq f(y) \forall y \in X\}$$

3.2 Modelli lineari

Gran parte dei modelli che verranno studiati in questo volume rientra nella categoria dei cosiddetti “modelli lineari”. E’ opportuno ricordare la seguente definizione:

Definizione 5. Una *funzione lineare* in \mathbb{R}^n è una funzione $f(\cdot)$ additiva ed omogenea, cioè

$$\begin{aligned} f(x+y) &= f(x) + f(y) & \forall x, y \in \mathbb{R}^n \\ f(\alpha x) &= \alpha f(x) & \forall x \in \mathbb{R}^n, \alpha \in \mathbb{R} \end{aligned}$$

Si dimostra che una funzione lineare può essere sempre espressa nella forma

$$f(x_1, \dots, x_n) = a_1x_1 + \dots + a_nx_n = \sum_{i=1}^n a_i x_i = a^T x$$

dove con a si è indicato il vettore colonna

$$a = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

e con a_1, \dots, a_n si sono indicate n costanti reali.

Una *funzione affine* ha invece la forma:

$$f(x_1, \dots, x_n) = a_1x_1 + \dots + a_nx_n + b = \sum_{i=1}^n a_i x_i + b = a^T x + b,$$

dove b è una costante reale.

I problemi di *programmazione lineare* sono caratterizzati da una funzione obiettivo lineare e da un insieme ammissibile rappresentabile esplicitamente mediante un numero *finito* di equazioni e/o disequazioni lineari. Un generico problema di programmazione lineare potrebbe pertanto avere la forma seguente:

3. PROBLEMI DI OTTIMIZZAZIONE

$$\text{Opt} \sum_{j=1}^n c_j x_j \quad (3.4)$$

$$\sum_{j=1}^n a_{ij} x_j = b_i \quad i = 1, \dots, m_\ell \quad (3.5)$$

$$\sum_{j=1}^n a_{ij} x_j \leq b_i \quad i = m_\ell + 1, \dots, m_h \quad (3.6)$$

$$\sum_{j=1}^n a_{ij} x_j \geq b_i \quad i = m_h + 1, \dots, m_k \quad (3.7)$$

In questo problema il simbolo Opt rappresenta l'operatore \min o l'operatore \max ; sono presenti $m_\ell \geq 0$ equazioni lineari e $(m_k - m_\ell) \geq 0$ disequazioni lineari, alcune delle quali rappresentate in forma di \leq , altre di \geq .

Nel problema sopra rappresentato sono presenti i principali elementi dei generici problemi di programmazione lineare e, in particolare:

1. un insieme di *variabili di decisione* o *incognite* x_1, x_2, \dots, x_n ;
2. una *funzione obiettivo* o *funzione di costo* da minimizzare o massimizzare;
3. alcune (in certi casi anche nessuna) *equazioni lineari* o *vincoli di egualanza*;
4. alcune (a volte nessuna) *disequazioni lineari* o *vincoli di disegualanza*;
5. alcuni *vincoli di segno* o di *vincoli di non negatività* sulle variabili; questi vincoli non sono altro che semplici disequazioni lineari, ma in genere si preferisce mantenerli separati dagli altri vincoli del problema.

Si noti che nel modello appena introdotto non sono presenti vincoli di diseguaglianza stretta, cioè caratterizzati dalla relazione $>$ o $<$; né sono presenti vincoli caratterizzati dalla relazione \neq .

In un modello lineare, la funzione obiettivo, cioè la funzione da minimizzare o massimizzare, è lineare, così come i vincoli sono espressi da equazioni e/o disequazioni lineari. La funzione obiettivo potrebbe anche essere affine, cioè avere la forma $a^T x + b$, con $b \in \mathbb{R}$, anche se dal punto di vista algoritmico non vi è alcuna differenza nell'ottimizzare funzioni che differiscono fra loro per una costante.

La notevole importanza data ai modelli lineari si può spiegare in vari modi:

1. molti concetti e proprietà relativi ai modelli lineari si possono ritrovare, con opportune varianti, anche in alcuni modelli più complessi;
2. i modelli lineari possono essere risolti molto efficientemente;
3. moltissimi modelli applicativi sono effettivamente o possono ricondursi a modelli lineari;
4. molti modelli non lineari possono essere approssimati efficacemente mediante opportuni modelli lineari.

Naturalmente l'ipotesi di linearità è molto forte. Formulare un problema mediante un modello lineare corrisponde ad assumere la validità delle seguenti ipotesi:

1. additività: il contributo al "costo" (o al membro di sinistra di equazioni o disequazioni) è la somma dei contributi dovuti alle singole incognite;
2. proporzionalità marginale: il contributo (ad esempio, al costo) di ciascuna variabile è di tipo proporzionale: a valori doppi

3. PROBLEMI DI OTTIMIZZAZIONE

della variabile corrisponde un contributo doppio. Spesso questa ipotesi non è ragionevole – si pensi ad esempio al caso degli sconti per acquisti di grosse quantità, oppure ai modelli di imposizione fiscale, nei quali ad imponibili via via crescenti corrispondono percentuali di prelievo crescenti.

3. divisibilità: le incognite possono assumere qualsiasi valore reale compatibile con i vincoli. Cioè, se una variabile può assumere sia il valore 3 che il valore 7, allora potrà anche assumere ciascuno dei valori dell'intervallo $[3, 7]$. Questa è forse l'ipotesi che più spesso non è verificata nelle applicazioni reali. Molto spesso alcune incognite possono assumere solo valori interi, a volte soltanto binari. Nel capitolo 15 verranno presentati i modelli di programmazione lineare *intera*.

Anche se quello appena visto è il più generale problema di programmazione lineare che incontreremo in questo libro, è importante definire una forma “standard”, alla quale sia facile ricondurre altre forme di problemi di programmazione lineare. Avendo a disposizione tale forma standard sarà possibile sviluppare un algoritmo risolutivo “standard”, senza dover cioè presentare forme algoritmiche differenti per le diverse tipologie di problemi.

Forma standard di un problema di programmazione lineare

Useremo la convenzione di definire *standard* un problema di ottimizzazione che si presenta sotto il seguente aspetto:

$$\begin{aligned} \min \quad & c_1x_1 + c_2x_2 + \cdots + c_nx_n \\ & a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ & a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ & \vdots \qquad \qquad \vdots \qquad \vdots \\ & a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \\ & x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0 \end{aligned} \tag{3.8}$$

Il problema standard consiste quindi nella minimizzazione di una funzione obiettivo lineare nelle n incognite reali x_1, \dots, x_n ; i vincoli sono rappresentati da un sistema di m equazioni lineari e dall'imposizione del vincolo di non negatività su *tutte* le incognite.

Naturalmente lo stesso problema può anche essere espresso in forma più concisa come

$$\begin{aligned} & \min \sum_{j=1}^n c_j x_j \\ & \sum_{j=1}^n a_{ij} x_j = b_i \quad i = 1, \dots, m \\ & x_j \geq 0 \quad j = 1, \dots, n \end{aligned}$$

In molti casi sarà conveniente utilizzare una rappresentazione ancora più concisa. Riunendo i coefficienti del sistema di equazioni in una matrice $m \times n$:

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

i costi in un vettore n -dimensionale:

$$c = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$

i termini noti in un vettore m -dimensionale

$$b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

3. PROBLEMI DI OTTIMIZZAZIONE

ed, analogamente, le incognite in un vettore n -dimensionale:

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

si può scrivere

$$\min_{x \in \mathbb{R}^n} c^T x \quad (3.9)$$

$$Ax = b \quad (3.10)$$

$$x \geq 0 \quad (3.11)$$

Con la notazione c^T si indica il trasposto del vettore c (quindi, in questo caso, un vettore riga $1 \times n$) e con $c^T x$ si rappresenta il prodotto interno (o prodotto scalare) $\sum_{j=1}^n c_j x_j$ tra i vettori c ed x .

Trasformazione di programmi lineari in forma standard

Spesso un problema di programmazione lineare si presenta già nella forma standard precedentemente introdotta. Ad esempio, un semplice problema di abbinamento 3×3

$$\begin{aligned} \min & \sum_{i=1}^3 \sum_{j=1}^3 c_{ij} x_{ij} \\ & \sum_{j=1}^3 x_{ij} = 1 \quad i = 1, 2, 3 \\ & \sum_{i=1}^3 x_{ij} = 1 \quad j = 1, 2, 3 \\ & x_{ij} \geq 0, \quad i = 1, 2, 3; j = 1, 2, 3 \end{aligned}$$

è già scritto in forma standard e può essere rappresentato in forma matriciale ponendo ad esempio

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

$$b = [1 \ 1 \ 1 \ 1 \ 1 \ 1]^T$$

$$c = [c_{11} \ c_{12} \ c_{13} \ c_{21} \ c_{22} \ c_{23} \ c_{31} \ c_{32} \ c_{33}]^T$$

$$x = [x_{11} \ x_{12} \ x_{13} \ x_{21} \ x_{22} \ x_{23} \ x_{31} \ x_{32} \ x_{33}]^T$$

In altri casi però il problema originale non si presenta in forma standard – si pensi, ad esempio, al problema della dieta, nel quale i vincoli sono disequazioni lineari che rappresentano richieste minime o massime di elementi nutritivi. Fortunatamente è facile ricondurre un generico problema di programmazione lineare in forma standard.

Massimizzazione

Se il problema dato richiede la *massimizzazione* di un obiettivo lineare, ci si può ricondurre ad un problema standard semplicemente cambiando segno al vettore dei costi. Questo perché, data una qualunque funzione f ed un generico insieme $X \subseteq \mathbb{R}^n$,

$$\max\{f(x) : x \in X\} = -\min\{-f(x) : x \in X\}$$

e se $x^* \in X$ è tale che $f(x^*) = \max\{f(x) : x \in X\}$, allora è anche vero che lo stesso x^* è tale che $-f(x^*) = \min\{-f(x) : x \in X\}$. In genere, vale

$$\arg \max_{x \in X} f(x) = \arg \min_{x \in X} -f(x)$$

3. PROBLEMI DI OTTIMIZZAZIONE

La dimostrazione di questa importante proprietà è molto elementare e viene lasciata al lettore.

Dovendo ad esempio risolvere un problema di massimizzazione lineare con vincoli lineari di egualanza e variabili vincolate in segno come il seguente:

$$\begin{aligned} \max c^T x \\ Ax = b \\ x \geq 0 \end{aligned}$$

avendo a disposizione un codice di calcolo predisposto per la minimizzazione, basterà risolvere il problema

$$\begin{aligned} -\min -c^T x \\ Ax = b \\ x \geq 0 \end{aligned}$$

dove il segno $-$ prima dell'operatore \min indica che, al termine, occorre cambiare segno al valore dell'ottimo trovato (solo al *valore* dell'ottimo, non all'ottimo stesso).

Vincoli di diseguaglianza

Se in un problema di programmazione lineare compare un vincolo del tipo

$$\sum_{j=1}^n a_{ij}x_j \leq b_i,$$

ci si può ricondurre alla forma standard introducendo una variabile in più. L'espressione del vincolo può essere scritta in modo equivalente come

$$b_i - \sum_{j=1}^n a_{ij}x_j \geq 0.$$

3.2. Modelli lineari

Ora, definendo una nuova variabile s_i come

$$s_i = b_i - \sum_{j=1}^n a_{ij}x_j,$$

il vincolo originale si riduce al vincolo di non negatività $s_i \geq 0$. In altre parole, al posto del vincolo originale scriveremo

$$\sum_{j=1}^n a_{ij}x_j + s_i = b_i, \quad s_i \geq 0,$$

coerentemente con quanto richiesto dalla forma standard.

La variabile s_i viene detta variabile di *slack* o di *scarto*, in quanto essa rappresenta lo scarto esistente fra i due membri della disequazione originale. Se nel problema comparissero diversi vincoli di diseguaglianza, si potrebbe operare una simile trasformazione per ciascuno di essi, ricordandosi di utilizzare, per ciascun vincolo, una diversa variabile di slack.

Se in un problema comparissero disequazioni di tipo “ \geq ”, si potrebbe utilizzare un procedimento del tutto analogo; in particolare una disequazione del tipo

$$\sum_{j=1}^n a_{ij}x_j \geq b_i$$

può essere portata in forma standard con l'introduzione di una variabile aggiuntiva (detta di *surplus* o di *ecedenza*):

$$\begin{aligned} \sum_{j=1}^n a_{ij}x_j - s_i &= b_i \\ s_i &\geq 0 \end{aligned}$$

3. PROBLEMI DI OTTIMIZZAZIONE

In forma matriciale, si può quindi concludere che, se un problema ha la forma

$$\begin{aligned} \min c^T x \\ Ax \leq b \\ x \geq 0 \end{aligned}$$

allora un problema equivalente in forma standard è il seguente

$$\begin{aligned} \min c^T x \\ Ax + s = b \\ x, s \geq 0 \end{aligned}$$

cioè

$$\begin{aligned} \min [c \quad 0]^T \begin{bmatrix} x \\ s \end{bmatrix} \\ [A \quad I] \begin{bmatrix} x \\ s \end{bmatrix} = b \\ \begin{bmatrix} x \\ s \end{bmatrix} \geq 0 \end{aligned}$$

Nel caso di vincoli di \geq la situazione è analoga:

$$\begin{aligned} \min c^T x \\ Ax \geq b \\ x \geq 0 \end{aligned}$$

è equivalente a

$$\begin{aligned} \min [c \quad 0]^T \begin{bmatrix} x \\ s \end{bmatrix} \\ [A \quad -I] \begin{bmatrix} x \\ s \end{bmatrix} = b \\ \begin{bmatrix} x \\ s \end{bmatrix} \geq 0 \end{aligned}$$

Vincoli sul segno delle variabili

A volte, anche se nelle applicazioni ciò è relativamente raro, un'incognita è vincolata ad assumere valori non positivi: $x_j \leq 0$; ovviamente, in questo caso, sostituendo ogni occorrenza della variabile x_j con il suo opposto $-x_j$ ci si riporta alla forma standard, con vincoli di non negatività sulle incognite.

In altri casi invece sull'incognita x_j non è richiesto, o non è a priori ipotizzabile, alcun vincolo di segno – si pensi ad esempio al caso di un'incognita che rappresenti il livello di un conto corrente bancario: in base ai flussi di cassa, l'incognita potrà assumere sia valori negativi che positivi.

Uno dei sistemi più comunemente utilizzati per trattare questo caso consiste nel sostituire tale incognita *libera* con la differenza di due incognite non negative:

$$x_j = x_j^+ - x_j^- \quad \text{dove } x_j^+ \geq 0, x_j^- \geq 0$$

Per qualsiasi valore di x_j esisteranno infinite coppie di valori x_j^+ ed x_j^- tali che x_j è uguale a $x_j^+ - x_j^-$. In generale x_j^+ e x_j^- non coincidono con la parte positiva e la parte negativa di x_j ; questo avviene se e solo se almeno una delle due variabili x_j^+ o x_j^- è nulla – in questo caso l'altra variabile corrisponde alla parte positiva (o negativa) di x_j .

In forma matriciale, si ha che il problema

$$\begin{aligned} & \min c^T x + d^T y \\ & Ax + By = b \\ & x \geq 0 \end{aligned}$$

3. PROBLEMI DI OTTIMIZZAZIONE

si trasforma nell'equivalente

$$\begin{aligned} \min & [c \quad d \quad -d]^T \begin{bmatrix} x \\ y^+ \\ y^- \end{bmatrix} \\ & [A \quad B \quad -B] \begin{bmatrix} x \\ y^+ \\ y^- \end{bmatrix} = b \\ & \begin{bmatrix} x \\ y^+ \\ y^- \end{bmatrix} \geq 0 \end{aligned}$$

Grazie alle trasformazioni appena viste, un problema generico di programmazione lineare può sempre essere ricondotto ad un problema equivalente in forma standard; in genere tale problema avrà un numero di incognite maggiore del problema originale.

Esempio 1. Si consideri il seguente esempio, costruito in modo artificiale per poter illustrare tutte le tecniche di trasformazione sopra presentate.

$$\begin{aligned} & \max 3x - 2y + z + 4w \\ & 2x - y + 4z - 2w \geq 2 \\ & -3x + 2y - z \geq 1 \\ & x + 2y + 3w \leq 1 \\ & x + 2y - z - w = 4 \\ & x, y, z \geq 0 \\ & w \text{ libera in segno} \end{aligned}$$

Operando in modo graduale, si può iniziare dalla funzione obiettivo, che diviene

$$-\min -3x + 2y - z - 4w$$

3.2. Modelli lineari

Passando poi all'assenza di vincolo sul segno per l'incognita w , scritta come $u - v$, si ottiene

$$\begin{aligned} & -\min 3x - 2y + z - 4u + 4v \\ & 2x - y + 4z - 2u + 2v \geq 2 \\ & -3x + 2y - z \geq 1 \\ & x + 2y + 3u - 3v \leq 1 \\ & x + 2y - z - u + v = 4 \\ & x, y, z, u, v \geq 0 \end{aligned}$$

Portando poi in forma standard i vincoli di diseguaglianza, si arriva a:

$$\begin{aligned} & -\min 3x - 2y + z - 4u + 4v \\ & 2x - y + 4z - 2u + 2v - s_1 = 2 \\ & -3x + 2y - z - u_2 = 1 \\ & x + 2y + 3u - 3v + s_3 = 1 \\ & x + 2y - z - u + v = 4 \\ & x, y, z, u, v, s_1, s_2, s_3 \geq 0 \end{aligned}$$

o, in forma matriciale,

$$\begin{aligned} & -\min [3 \quad -2 \quad 1 \quad -4 \quad 4 \quad 0 \quad 0 \quad 0] w \\ & \begin{bmatrix} 2 & -1 & 4 & -2 & 2 & -1 & 0 & 0 \\ -3 & 2 & -1 & 0 & 0 & 0 & -1 & 0 \\ 1 & 2 & 3 & -3 & 0 & 0 & 0 & 1 \\ 1 & 2 & -1 & -1 & 1 & 0 & 0 & 0 \end{bmatrix} w = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 4 \end{bmatrix} \\ & w \geq 0 \end{aligned}$$

dove

$$w^T = [x \quad y \quad z \quad u \quad v \quad s_1 \quad s_2 \quad s_3]$$

3.3 Equivalenza tra problemi di ottimizzazione

L'equivalenza tra diversi problemi di programmazione lineare è stata postulata nel paragrafo precedente appoggiandosi su una base intuitiva. Una definizione rigorosa di equivalenza tra problemi di ottimizzazione raramente viene fornita nei libri di testo; poiché però alcune trasformazioni portano a problemi abbastanza differenti fra loro (si pensi al caso dell'eliminazione delle variabili libere in segno appena presentato) pare utile a questo punto definire il concetto di equivalenza in modo più rigoroso e non basare il concetto di equivalenza esclusivamente sull'intuizione. Per giungere ad una definizione formale che contenga come casi particolari i problemi "evidentemente" equivalenti, ma che non estenda l'equivalenza a problemi qualsiasi, pare conveniente introdurre una definizione formale di equivalenza che, purtroppo, non risulta particolarmente intuitiva. Gli unici testi nei quali si fa cenno al problema di equivalenza, forniscono una definizione del tipo *il problema P è equivalente al problema Q se e solo se o sono entrambi inammissibili, o sono entrambi illimitati oppure esistono opportune trasformazioni che permettano, data una soluzione ottimale di uno dei due problemi, di trovare una soluzione ottimale dell'altro.* Purtroppo questa non è una definizione utile, poiché se la trasformazione sopra indicata fosse quella che mappa un ottimo globale di un problema su un ottimo globale di un altro, allora da questa "definizione" risulterebbe che, dati due problemi di ottimizzazione, essi sono sempre equivalenti fra loro.

Ci pare quindi meglio, sacrificando un po' di intuizione, fornire come definizione di equivalenza la seguente che, in vari testi, viene invece introdotta come proprietà:

Definizione 6. Due problemi di minimizzazione

$$P : \min_x f(x)$$

$$x \in X \subseteq \mathbb{R}^n$$

3.3. Equivalenza tra problemi di ottimizzazione

e

$$Q : \min_y g(y)$$
$$y \in Y \subseteq \mathbb{R}^m$$

si dicono *equivalenti* se e solo se vale $X = Y = \emptyset$, oppure esistono due funzioni

$$\phi : X \rightarrow Y$$
$$\psi : Y \rightarrow X$$

tali che:

$$g(\phi(x)) \leq f(x) \quad \forall x \in X$$

e

$$f(\psi(y)) \leq g(y) \quad \forall y \in Y$$

Anche se relativamente poco intuitiva, questa definizione permette di ottenere lo scopo di individuare, anche in modo operativo, l'equivalenza fra problemi di minimizzazione. Infatti vale il seguente

Teorema 1. *Dati due problemi di minimizzazione equivalenti*

$$P : \min_x f(x)$$
$$x \in X \subseteq \mathbb{R}^n$$

e

$$Q : \min_y g(y)$$
$$y \in Y \subseteq \mathbb{R}^m$$

allora

3. PROBLEMI DI OTTIMIZZAZIONE

1. Se uno dei problemi è illimitato, anche l'altro è illimitato
2. Se uno dei due problemi ammette ottimo, allora anche l'altro problema ammette ottimo ed i valori ottimali di entrambi i problemi coincidono
3. Se $x^* \in \arg \min_{x \in X} f(x)$ allora $\phi(x^*) \in \arg \min_{y \in Y} g(y)$ e se $y^* \in \arg \min_{y \in Y} g(y)$ allora $\psi(y^*) \in \arg \min_{x \in X} f(x)$

Dimostrazione.

1. Si ipotizzi, senza perdere generalità, che P sia illimitato inferiormente. Allora

$$\begin{aligned}\exists \{x_k\}_{k=0}^{\infty} : x_k &\in X \\ \lim_{k \rightarrow \infty} f(x_k) &= -\infty\end{aligned}$$

In questo caso, poiché la successione $\{\phi(x_k)\}$ è ammissibile per Q , risulta

$$g(\phi(x_k)) \leq f(x_k) \rightarrow -\infty$$

e si è pertanto individuata una successione di soluzioni ammissibili per Q , $\{y_k = \phi(x_k)\}$, di costo divergente a $-\infty$.

2. Anche qui si ipotizzi che x^* sia una soluzione ottimale del problema P . Allora, ponendo $y^* = \phi(x^*)$, si ha

$$\begin{aligned}g(y^*) &= g(\phi(x^*)) \\ &\leq f(x^*) \\ &\leq f(\psi(y)) && \forall y \in Y \\ &\leq g(y) && \forall y \in Y\end{aligned}$$

e, pertanto, y^* è una soluzione ottimale di Q . Inoltre, per quanto riguarda il valore degli ottimi, si ha

$$g(y^*) \leq f(x^*)$$

3.3. Equivalenza tra problemi di ottimizzazione

ma, poichè $g(y^*) \geq f(\psi(y^*))$, segue che,

$$f(\psi(y^*)) \geq f(x^*)$$

da cui

$$g(y^*) = f(x^*)$$

3. la dimostrazione di questa parte è già contenuta nella precedente.

□

Si può anche estendere la definizione sopra riportata al caso di problemi di massimo:

Definizione 7. Due problemi di massimizzazione

$$\begin{aligned} P : & \max_x f(x) \\ & x \in X \subseteq \mathbb{R}^n \end{aligned}$$

e

$$\begin{aligned} Q : & \max_y g(y) \\ & y \in Y \subseteq \mathbb{R}^m \end{aligned}$$

si dicono *equivalenti* se e solo se vale $X = Y = \emptyset$, oppure esistono due funzioni

$$\begin{aligned} \phi : & X \rightarrow Y \\ \psi : & Y \rightarrow X \end{aligned}$$

tali che:

$$g(\phi(x)) \geq f(x) \quad \forall x \in X$$

e

$$f(\psi(y)) \geq g(y) \quad \forall y \in Y$$

3. PROBLEMI DI OTTIMIZZAZIONE

E' possibile, infine, definire un concetto di equivalenza anche per problemi con direzione dell'ottimizzazione differente:

Definizione 8. Due problemi di ottimizzazione

$$P : \min_x f(x)$$
$$x \in X \subseteq \mathbb{R}^n$$

e

$$Q : -\max_y g(y)$$
$$y \in Y \subseteq \mathbb{R}^m$$

si dicono *equivalenti* se e solo se vale $X = Y = \emptyset$, oppure esistono due funzioni

$$\phi : X \rightarrow Y$$
$$\psi : Y \rightarrow X$$

tali che:

$$-g(\phi(x)) \leq f(x) \quad \forall x \in X$$

e

$$-g(y) \leq f(\psi(y)) \quad \forall y \in Y$$

Si può verificare che in entrambi i casi queste definizioni permettono di dimostrare un analogo teorema relativo alla corrispondenza fra ottimi dei due problemi. Per l'ultima definizione, si ha che le soluzioni ottimali di uno dei due problemi possono essere trasformate in soluzioni ottimali dell'altro problema, mentre per quanto riguarda i *valori* di tali soluzioni, essi sono uno l'opposto dell'altro.

Possiamo ora verificare formalmente l'equivalenza tra alcune famiglie particolari di problemi di ottimizzazione.

3.3. Equivalenza tra problemi di ottimizzazione

Teorema 2. *Il problema di ottimizzazione*

$$\min_{x \in X} f(x)$$

è equivalente al problema di ottimizzazione

$$-\max_{x \in X} (-f(x))$$

Dimostrazione. In questo caso, vista la definizione di equivalenza tra un problema di minimo ed uno di massimo, la dimostrazione è immediata, ponendo

$$\begin{aligned}\phi(x) &= x & \forall x \in X \\ \psi(y) &= y & \forall y \in X\end{aligned}$$

□

Come ulteriore esempio di applicazione formale della nozione di equivalenza si consideri il caso dei problemi lineari con vincoli di diseguaglianza.

Teorema 3. *I problemi:*

$$\begin{aligned}P : \min c^T x \\ Ax \leq b \\ x \geq 0\end{aligned}$$

e

$$\begin{aligned}Q : \min c^T x \\ Ax + y = b \\ x, y \geq 0\end{aligned}$$

sono equivalenti.

3. PROBLEMI DI OTTIMIZZAZIONE

Dimostrazione. Si considerino gli insiemi ammissibili

$$X = \{x \in \mathbb{R}^n : Ax \leq b, x \geq 0\}$$

$$Y = \{(x, y) \in \mathbb{R}^{n+m} : Ax + y = b, x, y \geq 0\}$$

e le trasformazioni

$$\phi(x) = \begin{bmatrix} x \\ b - Ax \end{bmatrix}$$

e

$$\psi(x, y) = x$$

1. occorre dimostrare che

$$X = \{x : Ax \leq b, x \geq 0\} = \emptyset$$

se e solo se

$$Y = \{(x, y) : Ax + y = b, x, y \geq 0\} = \emptyset$$

Infatti se esiste $(x, y) \in Y$ allora, essendo $Ax + y = b$ e $y \geq 0$ si avrebbe $Ax \leq b, x \geq 0$, ovvero $X \neq \emptyset$. Analogamente se $x \in X$, scelto $y = b - Ax$ si ha $(x, y) \in Y$, e perciò $Y \neq \emptyset$.

2. per la definizione di equivalenza, ricordando che per i problemi in esame

$$f(x) = c^T x$$

$$g(x, y) = [c \quad 0]^T \begin{bmatrix} x \\ y \end{bmatrix}$$

occorre verificare che

$$g(\phi(x)) \leq f(x)$$

$$f(\psi(x, y)) \leq g(x, y)$$

3.3. Equivalenza tra problemi di ottimizzazione

Infatti

$$\begin{aligned} g(\phi(x)) &= g(x, b - Ax) \\ &= c^T x = f(x) \end{aligned}$$

e

$$\begin{aligned} f(\psi(x, y)) &= f(x) \\ &= c^T x = [c \quad 0]^T \begin{bmatrix} x \\ y \end{bmatrix} \end{aligned}$$

□

Teorema 4. *I problemi:*

$$P : \min_{x \in X} f(x)$$

e

$$\begin{aligned} Q : \min & f(u - v) \\ u - v &\in X \\ u, v &\geq 0 \end{aligned}$$

sono equivalenti.

Dimostrazione. 1. i due problemi sono o entrambi ammissibili o entrambi non ammissibili. Infatti, se $\exists \bar{x} \in X$, scegliendo

$$\bar{u}_i = \max\{0, \bar{x}_i\} \quad i = 1, \dots, n \quad (3.12)$$

$$\bar{v}_i = -\min\{0, \bar{x}_i\} \quad i = 1, \dots, n \quad (3.13)$$

si ottiene una soluzione $\begin{bmatrix} \bar{u} \\ \bar{v} \end{bmatrix}$ ammissibile per Q ; viceversa se $\begin{bmatrix} \bar{u} \\ \bar{v} \end{bmatrix}$ è ammissibile per Q , si vede immediatamente che $\bar{x} = \bar{u} - \bar{v}$ è ammissibile per P .

3. PROBLEMI DI OTTIMIZZAZIONE

2. definendo

$$\phi(x) = \begin{bmatrix} u \\ v \end{bmatrix}$$

con u e v definiti tramite le (3.12) e (3.13) e

$$\psi(u, v) = u - v$$

si ottengono due mappe tra gli insiemi ammissibili dei due problemi che soddisfano i requisiti della definizione di equivalenza. Infatti

$$\begin{aligned} g(\phi(x)) &= g\left(\begin{bmatrix} \max(0, x) \\ -\min(0, x) \end{bmatrix}\right) \\ &= f(\max(0, x) + \min(0, x)) \\ &= f(x) \end{aligned}$$

e

$$f(\psi(u, v)) = f(u - v) = g(u, v)$$

□

Vale la pena osservare a questo punto che spesso uno stesso problema può essere trasformato in diversi problemi equivalenti fra loro. Ad esempio si può vedere come è possibile trasformare un problema con variabili libere in segno in uno equivalente che ha solo una variabile in più rispetto all'originale. Qui, come più sopra, per semplicità ci si limiterà a presentare il caso in cui tutte le variabili siano libere in segno:

Teorema 5. *I problemi*

$$P : \min_{x \in X} f(x)$$

3.3. Equivalenza tra problemi di ottimizzazione

e

$$\begin{aligned} Q : \min f(y - v\mathbf{1}) \\ y - v\mathbf{1} \in X \\ y \geq 0 & \quad y \in \mathbb{R}^n \\ v \geq 0 & \quad v \in \mathbb{R} \end{aligned}$$

sono equivalenti.

(con $\mathbf{1}$ si è indicato un vettore di dimensione opportuna i cui elementi sono tutti pari ad 1)

Dimostrazione. La trasformazione che permette di dimostrare l'equivalenza è la seguente:

$$\phi : x \in X \mapsto \begin{bmatrix} y \\ v \end{bmatrix}$$

con

$$v = -\min\{0, \min_{j=1,\dots,n} \{x_j\}\}$$

e

$$y_j = x_j + v,$$

mentre

$$\psi(y, v) = x$$

con

$$x_j = y_j - v \quad j = 1, \dots, n.$$

La dimostrazione formale dell'equivalenza viene lasciata come esercizio. \square

3. PROBLEMI DI OTTIMIZZAZIONE

Dai due teoremi precedenti si ottiene quindi immediatamente che il problema di PL non standard

$$\begin{aligned} & \min c^T x + d^T y \\ & Ax + By = b \\ & x \geq 0 \end{aligned}$$

è, come si è già presentato, equivalente al problema

$$\begin{aligned} & \min c^T x + d^T y^+ - d^T y^- \\ & Ax + By^+ - By^- = b \\ & x, y^+, y^- \geq 0 \end{aligned}$$

ma anche al problema

$$\begin{aligned} & \min c^T x + d^T z - (d^T \mathbf{1}) v \\ & Ax + Bz - B\mathbf{1}v = b \\ & x, z, v \geq 0 \end{aligned}$$

Teorema 6. *Dato un qualunque insieme $P \subseteq \mathbb{R}^n$, e due funzioni reali f_1, f_2 da P in \mathbb{R} , definite e finite per ciascun punto $x \in P$, i problemi*

$$\begin{aligned} & \min_x \max \{f_1(x), f_2(x)\} \tag{3.14} \\ & x \in P \end{aligned}$$

e

$$\begin{aligned} & \min_{x,z} z \tag{3.15} \\ & z \geq f_1(x) \\ & z \geq f_2(x) \\ & x \in P \end{aligned}$$

sono equivalenti.

3.3. Equivalenza tra problemi di ottimizzazione

Dimostrazione. La funzione obiettivo e la direzione dell'ottimizzazione sono le medesime nelle due famiglie di problemi. Una coppia di trasformazioni utilizzabile per la dimostrazione di equivalenza è data da

$$\begin{aligned}\phi(x) &= (x, \max\{f_1(x), f_2(x)\}) \\ \psi(x, z) &= x\end{aligned}$$

1. Se l'insieme $X = P$ è vuoto, allora deve essere vuoto anche $Y = \{x, z : z \geq f_1(x), z \geq f_2(x), x \in P\}$ dal momento in cui $Y \subseteq P$; viceversa, se fosse vuoto Y allora sarebbe sicuramente vuoto P : se infatti esistesse una soluzione ammissibile $\bar{x} \in P$, allora la coppia $(\bar{x}, \max\{f_1(\bar{x}), f_2(\bar{x})\})$ apparterrebbe ad Y .
2. Per quanto riguarda il caso in cui i due problemi siano ammissibili, si vede immediatamente che le funzioni sopra indicate permettono di dichiarare i due problemi equivalenti. Infatti, sia $x \in P$; allora:

$$\begin{aligned}g(\phi(x)) &= g(x, \max\{f_1(x), f_2(x)\}) \\ &= \max\{f_1(x), f_2(x)\} \\ &= f(x)\end{aligned}$$

Sia invece $(x, z) \in Y$; segue che

$$\begin{aligned}f(\psi(x, z)) &= f(x) \\ &= \max\{f_1(x), f_2(x)\} \\ &\leq z = g(x, z)\end{aligned}$$

grazie ai vincoli del problema (3.15).

□

Grazie a questo teorema è quindi possibile trasformare problemi di minimax in problemi equivalenti. Si dimostra facilmente che

3. PROBLEMI DI OTTIMIZZAZIONE

L'equivalenza sopra dimostrata si può estendere al massimo tra un qualsiasi numero finito di funzioni. Ad esempio, il problema non lineare

$$\begin{aligned} \min_x \max & \left\{ c_1^T x, c_2^T x, \dots, c_k^T x \right\} \\ & Ax = b \\ & x \geq 0 \end{aligned}$$

risulta equivalente al problema di *PL*

$$\begin{aligned} \min_{z,x} & z \\ & z \geq c_1^T x \\ & z \geq c_2^T x \\ & \vdots \\ & z \geq c_k^T x \\ & Ax = b \\ & x \geq 0. \end{aligned}$$

Da questo esempio si può anche osservare che, affinché due problemi siano equivalenti, non è necessario che esista una trasformazione *biunivoca* dell'insieme ammissibile di un problema nell'insieme ammissibile dell'altro: ad esempio, attraverso la trasformazione ϕ non è possibile ottenere soluzioni ammissibili (x, z) del problema (3.15) con $z > f_1(x), z > f_2(x)$. E' importante, come si vede in questo caso, che le mappe ϕ e ψ abbiano come dominio un insieme che contiene i possibili punti di ottimo del problema trasformato, ma non necessariamente tutti i punti ammissibili. Come si vede nell'esempio, un problema "equivalente" ad un altro, non è "identico" all'altro: in questo caso il problema trasformato in forma lineare ammette soluzioni ammissibili che non corrispondono a soluzioni ammissibili del problema originale. Tuttavia le soluzioni *ottimali* sono in corrispondenza tra loro.

Teoria della Programmazione Lineare

In questo capitolo si presenteranno alcune proprietà fondamentali dei problemi di programmazione lineare, proprietà dalle quali traggono origine alcuni dei più utilizzati algoritmi di ottimizzazione lineare.

4.1 Basi e soluzioni di base nella Programmazione Lineare

Si consideri un problema di programmazione lineare in forma standard:

$$\min_{x \in \mathbb{R}^n} c^T x \quad (4.1)$$

$$Ax = b$$

$$x \geq 0$$

con $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$.

Per compatibilità con la maggior parte dei testi di Ricerca Operativa, si definisce *soluzione* del problema (4.1) un vettore $x \in \mathbb{R}^n$ tale che

$$Ax = b$$

4. TEORIA DELLA PROGRAMMAZIONE LINEARE

Quindi una *soluzione* è in realtà solamente soluzione del sistema di equazioni del problema. Si dice invece, coerentemente con quanto già definito in precedenza, *soluzione ammissibile* di un problema di *PL* un vettore $x \in \mathbb{R}^n$ tale che

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

Si dice infine *soluzione ottimale* un vettore $x^* \in \mathbb{R}^n$ tale che

$$\begin{aligned} Ax^* &= b \\ x^* &\geq 0 \\ c^T x^* &\leq c^T x \quad \forall x : Ax = b, x \geq 0 \end{aligned}$$

E' possibile, senza perdita di generalità, assumere la validità delle seguenti ipotesi:

Ipotesi 1. Il numero m di righe della matrice $A \in \mathbb{R}^{m \times n}$ è minore o uguale al numero n di colonne;

Ipotesi 2. La matrice A ha rango pari ad m .

E' chiaro che in molti casi queste ipotesi non sono verificate. Tuttavia è possibile ridurre l'analisi dei problemi di *PL* alle sole situazioni nelle quali le due ipotesi sono valide. Si hanno infatti i seguenti casi:

1. $m > n$: in questo caso nel sistema di equazioni $Ax = b$ compaiono più equazioni che incognite. Segue che o il sistema è privo di soluzioni, nel qual caso il problema di *PL* sarà, a maggior ragione, inammissibile; altrimenti, se esiste almeno una soluzione del sistema $Ax = b$, necessariamente almeno un'equazione è ridondante (cioè si può ottenere come combinazione lineare delle altre) e può essere eliminata, riducendo

4.1. Basi e soluzioni di base nella Programmazione Lineare

così il numero di equazioni da m ad $m - 1$. Si può ripetere l'operazione, eliminando via via equazioni ridondanti, fino ad ottenere un sistema nel quale $m = n$;

2. $m \leq n$: in questo caso esistono tre possibilità:

- a) il sistema $Ax = b$ ammette una ed una sola soluzione.
Questo è possibile se e solo se la matrice A è quadrata ed invertibile. In questo caso, evidentemente, il rango di A coincide con il numero di righe m ;
- b) il sistema $Ax = b$ non ammette soluzioni; come visto prima, in questo caso il problema di PL è inammissibile;
- c) il sistema $Ax = b$ ammette infinite soluzioni; in questo caso, se $\text{rango}(A) < m$, sicuramente almeno un'equazione è ridondante e, quindi, può essere eliminata senza cambiare l'insieme ammissibile. Ripetendo questa eliminazione un numero finito di volte si arriva ad un sistema equivalente all'originale in cui il rango della matrice dei coefficienti è pari al numero di righe.

Si osservi che, nel caso $m = n$, se il sistema ammette un'unica soluzione, allora il problema di PL risulta banale. Infatti o tale unica soluzione soddisfa i vincoli di non negatività $x \geq 0$ e, in tal caso, essendo l'unica soluzione ammissibile sarà anche la soluzione ottimale del problema; oppure, se almeno un vincolo di segno fosse violato, allora il problema di PL sarebbe inammissibile.

Per questo motivo, in pratica, si ipotizza spesso che il rango m della matrice A sia *strettamente* minore di n .

Le considerazioni fatte qui sopra ci autorizzano a considerare solo problemi di PL che soddisfano alle due condizioni date. Tuttavia occorre ricordare che volendo giungere ad un algoritmo risolutivo, occorrerà identificare un metodo che permetta di trasformare un problema in uno equivalente che soddisfi le ipotesi. Fortunatamente lo stesso metodo del simplex che verrà utilizzato per la

4. TEORIA DELLA PROGRAMMAZIONE LINEARE

risoluzione dei problemi di *PL*, potrà essere utilizzato anche per la verifica delle ipotesi, come si vedrà più avanti.

Si ipotizzi quindi che il

$$\text{rango}(A) = m \leq n$$

In questo caso, per la definizione di rango, è sempre possibile scegliere tra le n colonne della matrice A un sottoinsieme costituito da m colonne fra loro linearmente indipendenti. Ammettendo che tali colonne abbiano indici corrispondenti ad un dato insieme B e indicando con $N = \{1, \dots, n\} \setminus B$ l'insieme degli indici delle colonne rimanenti, si potrà scrivere (a parte un eventuale scambio di colonne)

$$A = [A_B \quad A_N]$$

dove A_B è una matrice (m, m) con determinante non nullo che viene detta *base*.

Il nome di "base" deriva dal fatto che le m colonne di A_B costituiscono una base per lo spazio euclideo m -dimensionale, cioè ogni elemento di \mathbb{R}^m può essere espresso in modo univoco tramite una combinazione lineare delle colonne di A_B . Ricordando che anche il termine noto b è un vettore m -dimensionale, si ha che il sistema di equazioni

$$A_B x = b$$

ammetterà sempre una (ed una sola) soluzione

$$\bar{x}_B = A_B^{-1}b \in \mathbb{R}^m.$$

Se si considera il vettore n -dimensionale $\bar{x} = \begin{bmatrix} \bar{x}_B \\ 0 \end{bmatrix}$ si vede immediatamente che tale vettore è una soluzione del sistema $Ax = b$. Infatti

$$A\bar{x} = [A_B \quad A_N] \begin{bmatrix} \bar{x}_B \\ \bar{x}_N \end{bmatrix} = A_B \bar{x}_B + A_N 0 = A_B A_B^{-1}b + 0 = b.$$

4.2. Il teorema fondamentale della programmazione lineare

Una soluzione \bar{x} così costruita viene chiamata *soluzione di base* associata alla base A_B . Le variabili x_B si dicono *variabili di base*.

Si noti che una soluzione di base possiede sempre *almeno* $n - m$ componenti nulle. Potrebbe tuttavia avere anche più di $n - m$ componenti pari a zero; in tal caso la corrispondente soluzione di base si dice *soluzione di base degenere*. Per estensione, in questi casi, anche la base si dirà degenere.

4.2 Il teorema fondamentale della programmazione lineare

Le soluzioni di base sono soluzioni particolari di un problema di *PL*. Il teorema seguente mostra che, sotto debolissime condizioni, l'esistenza di soluzioni ammissibili e di soluzioni ottimali per un problema di *PL* standard equivale all'esistenza di soluzioni (rispettivamente ammissibile e ottimali) che siano anche di base.

Teorema 7. *Si consideri un problema di programmazione lineare in forma standard, con m vincoli, n incognite, $m \leq n$ e con $\text{rango}(A) = m$.*

1. *esiste una soluzione ammissibile se e solo se esiste almeno una soluzione ammissibile che è anche di base;*
2. *esiste una soluzione ottimale se e solo se esiste almeno una soluzione ottimale che è anche di base.*

Siamo quindi autorizzati a limitarci all'esplorazione delle soluzioni di base, poiché tra di esse, se il problema ammette ottimo, si troverà la soluzione ottimale; ciò non vuol dire che non possano esistere soluzioni ottimali non di base, ma almeno una delle soluzioni ottimali deve essere di base. Le eventuali soluzioni ottimali alternative avranno, naturalmente, tutte lo stesso costo. Si osservi che questo teorema permette, in un certo senso, di passare da un problema in cui lo spazio delle soluzioni da analizzare per trovare

4. TEORIA DELLA PROGRAMMAZIONE LINEARE

l'ottimo è un “continuo”, ad un insieme di candidati ottimi discreto, anzi, finito: infatti le soluzioni di base nascono dall’identificazione di una sottomatrice A_B invertibile costituita da m colonne della matrice originale A , e pertanto il numero complessivo di possibili basi non potrà superare il numero (finito) di modi in cui è possibile scegliere m colonne da un insieme di n , e cioè

$$\binom{n}{m} = \frac{n!}{m!(n-m)!}.$$

Naturalmente la finitezza del numero massimo di basi, se da un lato permette di pensare ad una procedura finita per la risoluzione di un generico problema di programmazione lineare per il quale sia nota l’esistenza di una soluzione ottima, dall’altro non rappresenta un risultato molto potente. Il numero di basi di un problema di programmazione lineare cresce, al crescere di n ed m ad una velocità tale da rendere totalmente impraticabile la strada dell’enumerazione esplicita di tutte le soluzioni di base. I migliori metodi basati sull’esplorazione di soluzioni di base, quali ad esempio il metodo del simplex che verrà presentato più avanti, si sono spesso dimostrati, in pratica, estremamente efficienti nel determinare una base ottimale evitando di generare esplicitamente tutte le soluzioni di base, compiendo cioè un’enumerazione implicita delle basi.

D’altra parte, però, la proprietà enunciata nel Teorema Fondamentale è stata spesso considerata come una caratteristica irrinunciabile degli algoritmi per la programmazione lineare: solo negli anni ’90 sono stati sviluppati algoritmi radicalmente diversi dai metodi costruiti sull’esplorazione di basi che procedono generando via via soluzioni intermedie non di base.

Dimostrazione. 1. La condizione sufficiente si dimostra banalmente; per quanto riguarda la condizione necessaria, si consideri

4.2. Il teorema fondamentale della programmazione lineare

una soluzione ammissibile \bar{x} ¹ del sistema $Ax = b$. La matrice A può essere vista come accostamento di n vettori colonna:

$$A = [A_1 \ A_2 \ \cdots \ A_n]$$

e pertanto il sistema di equazioni $Ax = b$ può essere scritto nella forma

$$\sum_{j=1}^n A_j x_j = b;$$

in altri termini si può esprimere il vettore b come combinazione lineare delle colonne di A mediante i coefficienti x_j . Sia p , ($0 \leq p \leq n$), il numero di componenti non nulle nel vettore \bar{x} . Si può sempre supporre che le eventuali componenti non nulle occupino i primi p posti nel vettore \bar{x} : sia cioè

$$\bar{x} = [\bar{x}_1 \ \bar{x}_2 \ \dots \ \bar{x}_p \ 0 \ \cdots \ 0]^T.$$

Avendo espresso \bar{x} in questo modo, sarà vero che

$$\sum_{j=1}^p A_j \bar{x}_j = b.$$

Si considerino ora le colonne della matrice A associate alle componenti non nulle di \bar{x} , cioè A_1, \dots, A_p . Si possono distinguere due situazioni possibili:

- a) le colonne A_1, \dots, A_p sono *linearmente indipendenti*.

In questo caso p è certamente non maggiore di m , poiché, grazie all'ipotesi sul rango di A , non possono esistere più di m colonne linearmente indipendenti tra le n

¹in genere si useranno i simboli barrati, come \bar{x} per indicare un vettore particolare, mentre si indicheranno senza barra (x) per indicare vettori di incognite, variabili

4. TEORIA DELLA PROGRAMMAZIONE LINEARE

della matrice A . Se fosse $p = m$, la soluzione \bar{x} sarebbe una soluzione di base (la corrispondente sottomatrice $A_B = [A_1 \dots A_m]$ di A è invertibile); se invece fosse $p < m$, per una proprietà elementare delle matrici, sicuramente tra le colonne A_{p+1}, \dots, A_n se ne potrebbero scegliere $m - p$ che, assieme alle prime p colonne, formano un sistema di m vettori linearmente indipendenti (cioè una *base*). Siano $A_{i_1}, A_{i_2}, \dots, A_{i_{m-p}}$ tali colonne. Il vettore \bar{x} è una soluzione di base corrispondente alla base

$$A_B = \begin{bmatrix} A_1 & A_2 & \cdots & A_p & A_{i_1} & A_{i_2} & \cdots & A_{i_{m-p}} \end{bmatrix}$$

Infatti, la combinazione lineare

$$A_1\bar{x}_1 + \cdots + A_p\bar{x}_p + A_{i_1}0 + \cdots + A_{i_{m-p}}0 = b$$

è ancora valida e, dato che i coefficienti formano una matrice invertibile, la soluzione $\bar{x}_1, \dots, \bar{x}_p, 0, \dots, 0$ sarà l'unica soluzione (di base) del sistema $A_Bx = b$. In questo caso, in cui sono nulle alcune delle componenti della soluzione di base (oltre naturalmente a quelle fuori base), la soluzione di base e la base stessa sono *degeneri*.

- b) le colonne A_1, \dots, A_p sono *linearmente dipendenti*. La tecnica dimostrativa consiste in questo caso nel generare soluzioni ammissibili con un numero decrescente di componenti non nulle, fino a ridursi al caso in cui i coefficienti di tali variabili formino un insieme di vettori linearmente indipendenti (caso a). Dalla definizione di vettori linearmente dipendenti si deduce che sicuramente esistono p coefficienti d_1, d_2, \dots, d_p non tutti nulli tali che

$$\sum_{j=1}^p A_j d_j = 0.$$

4.2. Il teorema fondamentale della programmazione lineare

Ricordando che vale la relazione

$$\sum_{j=1}^p A_j \bar{x}_j = b$$

si ottiene immediatamente che

$$\sum_{j=1}^p A_j (\bar{x}_j + \varepsilon d_j) = b \quad \forall \varepsilon \in \mathbb{R}.$$

Introducendo il vettore n -dimensionale

$$d = [d_1 \quad d_2 \quad \dots \quad d_p \quad 0 \quad 0 \quad \dots \quad 0]^T,$$

si ha

$$A(\bar{x} + \varepsilon d) = b$$

Il vettore d è un elemento particolare del nucleo (Ker) della matrice A e, come noto dall'algebra lineare, può essere utilizzato per generare soluzioni del sistema di equazioni.

Per qualsiasi valore di ε , il vettore $\bar{x} + \varepsilon d$ è *soluzione* del sistema $Ax = b$; non è detto che in generale tale vettore sia anche soluzione *ammissibile* – lo è certamente per $\varepsilon = 0$. L'idea è a questo punto di scegliere ε in modo tale da annullare almeno una componente di $\bar{x} + \varepsilon d$ senza perdere l'ammissibilità. Si consideri ciascuna componente del vettore $\bar{x} + \varepsilon d$, con $\varepsilon \in \mathbb{R}$ (ci si può limitare alle componenti di indice $j = 1, \dots, p$, essendo tutte le altre pari a 0):

- se $d_j = 0$ allora $\bar{x}_j + \varepsilon d_j = \bar{x}_j > 0 \quad \forall \varepsilon$;
- se $d_j < 0$ allora $\bar{x}_j + \varepsilon d_j \geq 0 \quad \forall \varepsilon \leq -\frac{\bar{x}_j}{d_j}$;

4. TEORIA DELLA PROGRAMMAZIONE LINEARE

- se $d_j > 0$ allora $\bar{x}_j + \varepsilon d_j \geq 0 \quad \forall \varepsilon \geq -\frac{\bar{x}_j}{d_j}$.

Dovendo rispettare il vincolo sul segno di ciascuna componente del vettore $\bar{x}_j + \varepsilon d_j$, occorrerà scegliere per ε la condizione più restrittiva. Siano:

$$\varepsilon_\ell = \max_{j:d_j>0} \left\{ -\frac{\bar{x}_j}{d_j} \right\}$$

$$\varepsilon_u = \min_{j:d_j<0} \left\{ -\frac{\bar{x}_j}{d_j} \right\}$$

Si noti che $\varepsilon_\ell < 0$ e che $\varepsilon_u > 0$ e che almeno uno dei due termini $\varepsilon_\ell, \varepsilon_u$ è finito, grazie al fatto che non tutti gli elementi $d_j, j = 1, \dots, p$ sono nulli.

Supponendo che ad esempio ε_u sia finito, allora il vettore

$$\hat{x} := \bar{x} + \varepsilon_u d$$

è ancora soluzione, ma con *al più* $p - 1$ componenti non nulle. Se invece $\varepsilon_u = +\infty$, sicuramente si avrà $-\infty < \varepsilon_\ell < 0$, nel qual caso sarà il vettore

$$\hat{x} := \bar{x} + \varepsilon_\ell d$$

ad avere al più $p - 1$ componenti non nulle. Sostituendo al posto di \bar{x} la nuova soluzione ammissibile \hat{x} , si può ripetere il procedimento, individuando quindi le colonne dei coefficienti di A associate alle componenti non nulle di questa nuova soluzione, fino a giungere, in non più di p iterazioni, ad una soluzione corrispondente a colonne linearmente indipendenti (caso a), cioè ad una soluzione di base.

2. Per dimostrare la seconda parte del teorema, cioè che se esiste una soluzione ottimale, esiste pure una soluzione ottimale

4.2. Il teorema fondamentale della programmazione lineare

di base (il viceversa è banale), si procede sulla falsariga della dimostrazione della prima parte. Si ipotizzi di avere a disposizione una soluzione \bar{x} ottimale. Si distinguono ancora i due casi:

- a) le colonne A_1, \dots, A_p sono linearmente indipendenti. In questo caso la soluzione considerata è già (a parte l'eventuale aggiunta di zeri) una soluzione di base, ed il teorema è dimostrato;
- b) le colonne sono linearmente dipendenti. Procedendo come nella prima parte della dimostrazione, si osserva che, perturbando il vettore \bar{x} , il costo

$$c^T \bar{x}$$

viene modificato in

$$c^T(\bar{x} + \varepsilon d).$$

Da quanto già illustrato, si vede immediatamente che $\bar{x} + \varepsilon d$ è ammissibile per ogni $\varepsilon \in [\varepsilon_\ell, \varepsilon_u]$. Da qui si deduce che $c^T d$ deve essere necessariamente nullo. Infatti, essendo \bar{x} ottimale, sicuramente dovrà valere

$$c^T \bar{x} \leq c^T(\bar{x} + \varepsilon d) \quad \forall \varepsilon \in [\varepsilon_\ell, \varepsilon_u]$$

cioè

$$\varepsilon c^T d \geq 0 \quad \forall \varepsilon \in [\varepsilon_\ell, \varepsilon_u]$$

Ma poiché $\varepsilon_\ell < 0$ e $\varepsilon_u > 0$, deve necessariamente essere $c^T d = 0$.

Pertanto, essendo $c^T d = 0$, lo spostamento ammissibile da \bar{x} a $\bar{x} + \varepsilon d$, con $\varepsilon \in [\varepsilon_\ell, \varepsilon_u]$, effettuato come indicato nella prima parte della dimostrazione mantiene, oltre

4. TEORIA DELLA PROGRAMMAZIONE LINEARE

all'ammissibilità, anche l'*ottimalità*. La nuova soluzione, ancora ottimale, avrà al più $p - 1$ componenti non nulle; iterando il procedimento ci si riconduce in un numero finito di passi, al caso a.

□

La dimostrazione precedente è importante da un punto di vista metodologico poiché evidenzia una delle tecniche principali su cui è basato il più popolare algoritmo per la risoluzione di problemi di PL: il *metodo del simplex*. L'aspetto più interessante è la tecnica mediante la quale si passa da una soluzione ammissibile ad un'altra mediante uno spostamento ammissibile. L'algoritmo del simplex, come si vedrà nei prossimi paragrafi, sfrutta esattamente lo stesso procedimento, passando da una soluzione ammissibile di base all'altra mediante spostamenti ammissibili. A questo proposito, si ricorda che, considerato un insieme generico $S \subseteq \mathbb{R}^n$ ed un punto $x \in S$, si dice che un vettore $d \in \mathbb{R}^n$ rappresenta *direzione ammissibile in* x se esiste $\bar{\varepsilon} > 0$ tale che

$$x + \varepsilon d \in S \quad \forall \varepsilon \in [0, \bar{\varepsilon}].$$

In base alla definizione la direzione degenere $d = 0$ è sempre ammissibile. Nel caso di insiemi ammissibili di problemi di programmazione lineare in forma standard, si ha

$$S = \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$$

quindi le direzioni ammissibili in \bar{x} saranno caratterizzate da

$$\begin{aligned} A(\bar{x} + \varepsilon d) &= b \\ \bar{x} + \varepsilon d &\geq 0 \end{aligned}$$

cioè da

$$\begin{aligned} Ad &= 0 \\ \bar{x} + \varepsilon d &\geq 0 \end{aligned}$$

4.2. Il teorema fondamentale della programmazione lineare

Si vede quindi che le direzioni ammissibili sono identificabili con un sottoinsieme degli elementi del nucleo (*Ker*) della matrice A ; nella dimostrazione del teorema fondamentale, in particolare nella seconda parte, gioca un ruolo fondamentale la possibilità di compiere un passo non nullo sia in direzione d che in direzione $-d$. Questo equivale, nella scelta di una direzione ammissibile d , a poter compiere spostamenti ammissibili caratterizzati da un passo ϵ sia positivo che negativo; da qui segue immediatamente che, per garantire la non negatività, per ciascuna componente $\bar{x}_j = 0$ si avrà necessariamente $d_j = 0$. Le direzioni utilizzate nella dimostrazione del teorema fondamentale sono quindi “bi-ammissibili” (cioè ammissibili in entrambe le direzioni). Si è quindi anche implicitamente dimostrato che in corrispondenza di una soluzione ammissibile non di base, esiste sempre almeno una direzione bi-ammissibile non nulla. Nell'esempio seguente viene mostrata la tecnica di perturbazione ammissibile vista nella dimostrazione del teorema.

Esempio 2. Si consideri il seguente insieme dei vincoli di un problema di PL:

$$\begin{bmatrix} 1 & -1 & 1 & 0 & 2 & 1 & 2 \\ -1 & 0 & 0 & 1 & 1 & 2 & 1 \\ 0 & 0 & 1 & 1 & -1 & 1 & 1 \end{bmatrix} x = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}$$
$$x \geq 0$$

È immediato verificare che il vettore $\bar{x} = [1 \ 1 \ 1 \ 1 \ 1 \ 0 \ 0]^T$ è una soluzione ammissibile (cioè verifica i vincoli sopra indicati). Tuttavia sicuramente non è una soluzione di base, avendo 5 elementi non nulli. Seguendo la tecnica usata nella dimostrazione del teorema si individua una perturbazione ammissibile della soluzione \bar{x} .

Le colonne corrispondenti agli elementi non nulli di A sono le prime 5; è facile vedere che una combinazione lineare di tali colonne

4. TEORIA DELLA PROGRAMMAZIONE LINEARE

con coefficienti ad esempio pari a:

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

fornisce il vettore nullo – naturalmente esistono infinite altre combinazioni lineari non banali. Lo spostamento ammissibile avviene quindi nella direzione $d = [1 \ 0 \ -1 \ 1 \ 0 \ 0 \ 0]^T$. Scegliendo un passo ε si ottiene il nuovo punto

$$x(\varepsilon) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \varepsilon \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 + \varepsilon \\ 1 \\ 1 - \varepsilon \\ 1 + \varepsilon \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

Per mantenere l'ammissibilità si dovrà scegliere il passo $\varepsilon \in [-1, 1]$. Scegliendo ad esempio il passo -1 si ottiene la soluzione ammissibile $\bar{x} = [0 \ 1 \ 2 \ 0 \ 1 \ 0 \ 0]^T$ che è di base, essendo le colonne numero 2, 3, 5 della matrice originale fra loro linearmente indipendenti; scegliendo invece il passo $\varepsilon = 1$, si ottiene la soluzione ammissibile $\bar{x} = [2 \ 1 \ 0 \ 2 \ 1 \ 0 \ 0]^T$ che non è di base; iterando il procedimento, scegliendo ad esempio come coefficienti delle colonne numero 1, 2, 4, 5 i valori $2 \ 4 \ 1 \ 1$ si ottiene lo spostamento ammissibile $d = [2 \ 4 \ 0 \ 1 \ 1 \ 0 \ 0]^T$; applicando il criterio per la determinazione del passo massimo in entrambe le direzioni si ottiene che, essendo d non negativo, $\varepsilon_u = +\infty$ – come è

4.2. Il teorema fondamentale della programmazione lineare

facile verificare, infatti, tutti i punti

$$x(\varepsilon) = \begin{bmatrix} 2 + 2\varepsilon \\ 1 + 4\varepsilon \\ 0 \\ 2 + \varepsilon \\ 1 + \varepsilon \\ 0 \\ 0 \end{bmatrix}$$

sono soluzioni ammissibili, per qualunque valore di $\varepsilon \geq 0$. Per quanto riguarda invece gli spostamenti in direzione negativa, si ha

$$\varepsilon_\ell = \max_{j:d_j > 0} \left\{ -\frac{\bar{x}_j}{d_j} \right\} = \max \left\{ -\frac{2}{2}, -\frac{1}{4}, -\frac{2}{1}, -\frac{1}{1} \right\} = -\frac{1}{4}$$

da cui si ottiene la nuova soluzione ammissibile

$$\bar{x} = \left[\frac{3}{2} \quad 0 \quad 0 \quad \frac{7}{4} \quad \frac{3}{4} \quad 0 \quad 0 \right]$$

E' facile verificare che la sottomatrice costituita dalla prima, quarta e quinta colonna di A è invertibile; la soluzione trovata è dunque una soluzione di base.

Nei prossimi capitoli si vedrà come questa tecnica di spostamento lungo direzioni ammissibili possa essere sfruttata per mettere a punto efficienti algoritmi risolutivi per problemi di *PL*.

Geometria della programmazione lineare

In questo capitolo verranno fornite alcune definizioni e proprietà elementari relative alla geometria della *PL*. Le considerazioni geometriche che verranno presentata in questo capitolo saranno di tipo elementare e saranno orientate al miglioramento della comprensione di alcuni aspetti della programmazione lineare. L'argomento, trattato in modo approfondito, richiederebbe una trattazione ben più estesa; si è ritenuto qui di fornire solo alcune nozioni elementari per facilitare la comprensione di alcuni aspetti della programmazione lineare, rimandando a testi specializzati, come ad esempio (Schrijver, 1998) per approfondimenti e per estensioni.

Si consideri lo spazio vettoriale n -dimensionale \mathbb{R}^n . Si ricorda che uno *spazio* vettoriale è un insieme chiuso rispetto alla somma ed al prodotto scalare.

Definizione 9. Un sottoinsieme $S \subseteq \mathbb{R}^n$ chiuso rispetto alla somma vettoriale ed al prodotto scalare si dice *sottospazio* di \mathbb{R}^n .

Ogni sottospazio S di \mathbb{R}^n può essere rappresentato come l'insieme dei punti dello spazio che soddisfano ad un sistema di equazioni lineari omogenee, cioè

$$S = \{x \in \mathbb{R}^n : Ax = 0\}$$

5. GEOMETRIA DELLA PROGRAMMAZIONE LINEARE

dove A è una matrice (m, n) ; senza perdere di generalità, si supporrà che $m \leq n$.

Definizione 10. La *dimensione* di un sottospazio S , è il massimo numero di vettori linearmente indipendenti contenuti in S .

Esempio 3. In \mathbb{R}^3 l'insieme

$$S = \{(x, y, z) : x + y + z = 0\}$$

rappresenta un sottospazio (con $m = 1, n = 3$). E' facile vedere che i vettori linearmente indipendenti

$$v_1 = [1 \ -1 \ 0], v_2 = [1 \ 0 \ -1]$$

appartengono ad S ; quindi la dimensione di questo sottospazio è almeno 2. Si può anche dimostrare che non è possibile trovare tre vettori indipendenti in S .

Vale una proprietà che rende più semplice il calcolo della dimensione di un sottospazio.

Teorema 8. La dimensione di un sottospazio $S = \{x \in \mathbb{R}^n : Ax = 0\}$ è pari a

$$n - \text{rango}(A)$$

Nell'esempio precedente, il rango della matrice dei coefficienti

$$A = [1 \ 1 \ 1]$$

è pari ad 1 e, pertanto, la dimensione di S è 2.

Definizione 11. Un *sottospazio affine* è un sottoinsieme di \mathbb{R}^n ottenuto traslando un sottospazio.

Equivalentemente un sottospazio affine è rappresentabile come un sottoinsieme S di \mathbb{R}^n che soddisfa un sistema di equazioni lineari non necessariamente omogenee:

$$S = \{x \in \mathbb{R}^n : Ax = b\}.$$

Definizione 12. La dimensione di un sottospazio affine è la dimensione del sottospazio corrispondente, cioè quella del sottospazio ottenuto ponendo $b = 0$.

Più in generale:

Definizione 13. La dimensione di un qualunque insieme $E \subseteq \mathbb{R}^n$ è la minima dimensione di un sottospazio affine che contenga E .

Ad esempio, un insieme costituito da due punti distinti in \mathbb{R}^n , con $n \geq 1$, ha dimensione 1, come pure ha dimensione 1 un segmento; un cerchio con raggio strettamente positivo ha dimensione 2, ma anche la sola circonferenza ha dimensione pari a 2.

L'insieme ammissibile di un problema di *PL*

$$Ax = b$$

$$x \geq 0$$

è un sottoinsieme dello spazio affine $\{x : Ax = b\}$ e pertanto avrà dimensione non superiore a $n - \text{rango}(A)$.

Sia $\alpha \in \mathbb{R}^n$ un vettore non nullo e β un numero reale.

Definizione 14. Si definisce *iperpiano* il sottospazio affine di dimensione $n - 1$:

$$H = \left\{ x \in \mathbb{R}^n : \alpha^T x = \beta \right\}$$

Il vettore α si dice *normale* ad H . Un iperpiano suddivide lo spazio in due regioni dette *semispazi affini* la cui espressione analitica è:

$$H^\leq = \{x \in \mathbb{R}^n : \alpha^T x \leq \beta\}$$

$$H^\geq = \{x \in \mathbb{R}^n : \alpha^T x \geq \beta\}.$$

Si può vedere facilmente che l'orientamento del vettore α è verso l'interno del semispazio affine H^\geq . Infatti, se $\bar{x} \in H$, effettuando

5. GEOMETRIA DELLA PROGRAMMAZIONE LINEARE

uno spostamento positivo da \bar{x} in direzione α , si ottiene il punto

$$\bar{x} + \varepsilon\alpha$$

con $\varepsilon \geq 0$. Sostituendo le coordinate di questo punto nell'espressione di H si ottiene

$$\begin{aligned}\alpha^T(\bar{x} + \varepsilon\alpha) &= \alpha^T\bar{x} + \varepsilon\alpha^T\alpha \\ &= b + \varepsilon\|\alpha\|^2 \geq b\end{aligned}$$

Un sistema di disequazioni lineari

$$Ax \leq b$$

rappresenta quindi l'intersezione di un numero finito di semispazi.

Definizione 15. L'intersezione di un numero *finito* di semispazi affini è una figura convessa detta *poliedro*.

Si ricorda che un insieme si dice *convesso* se e solo se, dati due suoi punti qualsiasi, il segmento che li unisce è tutto contenuto nell'insieme stesso. Cioè un insieme $X \subseteq \mathbb{R}^n$ è convesso se e solo se dati $x_1, x_2 \in X$ anche $\lambda x_1 + (1 - \lambda)x_2 \in X$ per ogni $\lambda \in [0, 1]$.

I semispazi affini, gli iperpiani ed i poliedri sopra definiti sono insiemni convessi. Poiché un qualunque sistema di equazioni e disequazioni lineari può essere ricondotto ad un sistema di disequazioni lineari (ad esempio, sostituendo le equazioni $Ax = b$ con la coppia di disequazioni $Ax \geq b, Ax \leq b$), segue che l'insieme ammissibile di un qualunque problema di *PL* è un poliedro ed è, pertanto, un insieme convesso.

Definizione 16. Un poliedro limitato e non vuoto è detto *politopo*.

Esempio 4. Si consideri l'insieme ammissibile del seguente problema di PL in \mathbb{R}^2 :

$$\begin{aligned} -x + y &\geq 0 \\ x + y &\geq 1 \\ x + 2y &\leq 6 \\ y &\leq 4 \\ x &\leq 2 \\ x &\geq 0 \\ y &\geq 0 \end{aligned}$$

Questo insieme è un politopo che può essere rappresentato graficamente come in figura 5.1.

Nella figura, l'area in grigio rappresenta il poliedro ottenuto intersecando i semipiani corrispondenti alle disequazioni sopra riportate. L'espressione analitica di ciascuna disequazione è riportata, nel grafico, dal lato corrispondente al semipiano individuato dalla disequazione (con l'eccezione dei vincoli di non negatività, la cui espressione non è riportata).

Come si può vedere dal grafico alcuni vincoli possono risultare ridondanti: nell'esempio la regione ammissibile non cambierebbe se si eliminassero i vincoli $y \geq 0, y \leq 4, x \leq 2$.

Si consideri ora un poliedro $P \subseteq \mathbb{R}^n$;

Definizione 17. Una disequazione $\alpha^T x \leq \beta$ si dice *valida* per P se

$$P \subseteq \left\{ x \in \mathbb{R}^n : \alpha^T x \leq \beta \right\}.$$

Definizione 18. Data una disequazione

$$\alpha^T x \leq \beta$$

valida per P , l'iperpiano

$$H = \left\{ x \in \mathbb{R}^n : \alpha^T x = \beta \right\}$$

5. GEOMETRIA DELLA PROGRAMMAZIONE LINEARE

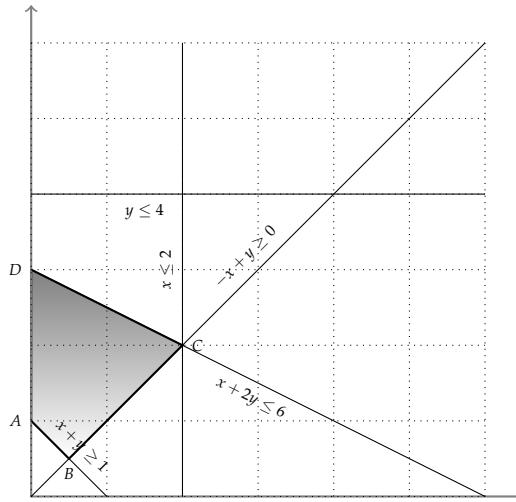


Figura 5.1: Un esempio di politopo

si dice *iperpiano di supporto* per P se e solo se

$$P \cap H \neq \emptyset.$$

Se H è un iperpiano di supporto per P , allora

$$P \cap H$$

si chiama *faccia* del poliedro.

I vincoli che definiscono un generico poliedro sono ovviamente diseguaglianze valide. Se con a_i^T e b_i si indicano rispettivamente i coefficienti dell' i -esimo vincolo ed il corrispondente termine noto, si ha che l'intersezione di ogni iperpiano della forma

$$a_i^T x = b_i$$

con il poliedro P , se non vuota, costituisce una faccia; inoltre l'intersezione di più facce di P è una faccia.

Una faccia di dimensione zero (cioè un punto) si dice *vertice* mentre una faccia di dimensione 1 si dice *spigolo*.

Se il poliedro, immerso in \mathbb{R}^n , ha dimensione $d \leq n$, allora una sua faccia di dimensione $d - 1$ è detta *facetta* oppure *faccia massimale*

Esempio 4 (cont.)

Con riferimento al precedente esempio, si osservi che i segmenti AB, BC, CD, DA sono spigoli (e anche facce massimali) del poliedro i cui vertici sono A, B, C, D . In particolare lo spigolo AB può essere generato dalla disequazione valida

$$-x - y \leq -1$$

Infatti la retta $x + y = 1$ interseca il poliedro lungo lo spigolo AB . L'intersezione fra la faccia AB e la faccia BC corrisponde al vertice B . Tale vertice è una faccia di dimensione 0; può ad esempio essere generato dalla disequazione valida

$$-2y \leq -1$$

La retta $-2y = -1$ è di supporto al poliedro, e la sua intersezione con il poliedro è il vertice B . Il vincolo $y \leq 4$ è naturalmente una disequazione valida, ma la retta $y = 4$ non interseca il poliedro: non costituisce pertanto un supporto e non definisce una faccia.

Definizione 19. Un punto x di un insieme convesso C si dice *estremo* se non è rappresentabile come combinazione convessa in senso stretto di alcuna coppia di punti distinti di C .

In altri termini, x è un estremo di C se e solo se

$$\begin{array}{ll} \nexists y, z \in C & y \neq z : \\ x = \lambda y + (1 - \lambda)z & \lambda \in (0, 1) \end{array}$$

5. GEOMETRIA DELLA PROGRAMMAZIONE LINEARE

Per i poliedri, che sono insiemi convessi particolari, è possibile ottenere una caratterizzazione più precisa dei punti estremi. Dato un poliedro $P = \{Ax \leq b\}$ ed un punto \bar{x} , si indichi con $I^=(\bar{x})$ l'insieme degli indici dei vincoli di P attivi in \bar{x} , cioè sia

$$\begin{aligned} a_i^T \bar{x} &= b_i & \forall i \in I^=(\bar{x}) \\ a_i^T \bar{x} &< b_i & \forall i \notin I^=(\bar{x}) \end{aligned}$$

Sia $A^=$ la sottomatrice di A costituita dalle righe il cui indice sia in $I^=(\bar{x})$ e analogamente si definisca $b^=$; si ha quindi $A^=\bar{x} = b^=$. Vale il seguente

Teorema 9. *Un punto \bar{x} appartenente ad un poliedro $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ di dimensione "piena" n è un estremo per P se e solo se $\text{rango}(A^=) = n$.*

Dimostrazione. Sia \bar{x} un estremo di P e si supponga per assurdo che $\text{rango}(A^=) < n$; allora il sistema di equazioni

$$A^=x = 0$$

ammette almeno una soluzione $y \neq 0$. I vettori

$$\begin{aligned} z &= \bar{x} + \epsilon y \\ w &= \bar{x} - \epsilon y \end{aligned}$$

sono tali che, per qualsiasi valore di ϵ , $A^=z = b^=$ e $A^=w = b^=$. Inoltre, poiché per i vincoli con indice $i \notin I(\bar{x})$ vale

$$a_i^T \bar{x} < b_i$$

e poiché tali vincoli sono in numero finito, esisterà una perturbazione ϵ sufficientemente piccola ma strettamente positiva tale che

$$\begin{aligned} a_i^T z &= a_i^T \bar{x} + \epsilon y \leq b_i \\ a_i^T w &= a_i^T \bar{x} + \epsilon w \leq b_i \end{aligned}$$

per $i \notin I(\bar{x})$. Ne segue che z e w appartengono al poliedro P . Però vale

$$\bar{x} = \frac{1}{2}z + \frac{1}{2}w$$

che è assurdo, essendo \bar{x} un punto estremo.

Per dimostrare il viceversa, si supponga che $\text{rango}(A^=) = n$ e che, per assurdo, \bar{x} non sia un punto estremo per P . Devono quindi esistere due punti z e w , con $z \neq w$, in P ed un numero reale $\lambda \in (0, 1)$ tali che

$$\bar{x} = \lambda z + (1 - \lambda)w.$$

Si consideri ora il sistema $A^=x = b^=$; sostituendo si ha che

$$\lambda A^=z + (1 - \lambda)A^=w = b^=.$$

Pertanto deve essere

$$\begin{aligned} A^=z &= b^= \\ A^=w &= b^= \end{aligned}$$

poiché, se per una delle due soluzioni valesse il vincolo con il segno di $<$ l'altra risulterebbe non ammissibile. Si è giunti all'assurdo, poiché in questo caso il sistema di equazioni $A^=x = b^=$ avrebbe più di una soluzione, contraddicendo l'ipotesi $\text{rango}(A^=) = n$. \square

Esistono poliedri privi di punti estremi: ogni poliedro $\{Ax \leq b\}$ con A matrice di rango minore di n non può avere punti estremi.

Esempio 5. Il poliedro

$$\{(x, y) \in \mathbb{R}^2 : x + y \leq 1\}$$

è privo di punti estremi.

Teorema 10. *In un poliedro P vertici e punti estremi coincidono.*

5. GEOMETRIA DELLA PROGRAMMAZIONE LINEARE

Dimostrazione. Sia \bar{x} un vertice di $P = \{Ax \leq b\}$; se \bar{x} non fosse un estremo, dovrebbero esistere due elementi $z, w \in P$ tali che

$$\bar{x} = \lambda z + (1 - \lambda)w$$

con $\lambda \in (0, 1)$. Essendo \bar{x} un vertice, deve esistere un iperpiano di supporto

$$H = \{x : \alpha^T x = \beta\}$$

tale che

$$P \subseteq \{\alpha^T x \leq \beta\}$$

e

$$H \cap P = \bar{x}.$$

Poiché z e w sono in P deve valere $\alpha^T z < \beta$ e $\alpha^T w < \beta$ (la disequazione deve essere stretta poiché \bar{x} è l'unico punto di P che appartiene ad H). Ma questo è assurdo perché, se fosse vero, si avrebbe

$$\alpha^T \bar{x} = \lambda \alpha^T z + (1 - \lambda) \alpha^T w < \beta.$$

Viceversa, se \bar{x} è un punto estremo, vale

$$A^= \bar{x} = b^=$$

con $A^=$ matrice di rango n . L'iperpiano

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

con $\alpha = \mathbf{1}^T A^=$ e $\beta = \mathbf{1}^T b^=$ è di supporto per P . Infatti per ogni punto $x \in P$ vale $Ax \leq b$ e, conseguentemente,

$$\mathbf{1}^T A^= x \leq \mathbf{1}^T b^=.$$

L'iperpiano H è quindi di supporto per P e vale $\bar{x} \in H \cap P$; resta da

dimostrare che \bar{x} è l'unico punto in $H \cap P$. Se $y \in H \cap P$ deve valere $A^=y \leq b^=$ e $\alpha^T y = \beta$; queste relazioni sono equivalenti a

$$\begin{aligned} b_i - a_i^T y &\geq 0 \quad \forall i \in I(\bar{x}) \\ \sum_{i \in I(\bar{x})} (b_i - a_i^T y) &= 0; \end{aligned}$$

ma la somma di quantità non negative può essere nulla solo se ciascun addendo è pari a 0. Pertanto vale

$$A^=y = b.$$

Questo è assurdo, poiché \bar{x} è l'unica soluzione del sistema $A^=x = b$. \square

Per i poliedri si possono quindi usare indifferentemente i termini di vertice e di punto estremo. I vertici hanno anche una corrispondenza con le soluzioni ammissibili di base dei problemi di PL. Vale infatti:

Teorema 11. *Ogni soluzione ammissibile di base di un problema di PL standard*

$$\begin{aligned} \min c^T x \\ Ax = b \\ x \geq 0 \end{aligned}$$

è un vertice del poliedro

$$P = \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}.$$

Dimostrazione. Una soluzione di base ammissibile è soluzione unica di un sistema

$$A_B x_B = b$$

con A matrice (m, m) invertibile. Almeno $n - m$ componenti di una soluzione di base sono nulle e, pertanto, almeno $n - m$ vincoli

5. GEOMETRIA DELLA PROGRAMMAZIONE LINEARE

$x \geq 0$ sono attivi in corrispondenza della soluzione di base. Si può quindi determinare la soluzione di base come unica soluzione di un sistema di n equazioni

$$\begin{bmatrix} A_B & 0 \\ 0 & I \end{bmatrix} x = \begin{bmatrix} b \\ 0 \end{bmatrix}.$$

Si è quindi dimostrato che la soluzione di base corrisponde ad un vertice di P , in base al teorema 9. \square

Il teorema appena dimostrato afferma che ad ogni soluzione di base corrisponde un vertice del poliedro P ; non è detto però che a soluzioni associate a basi differenti corrispondano vertici differenti: più soluzioni di base potrebbero essere rappresentate come un unico vertice. Questo avviene nel caso di soluzioni di base degeneri.

Esempio 4 (cont.)

Si consideri l'insieme ammissibile dell'esempio di pag. 68 scritto in forma standard:

$$\begin{aligned} -x + y - s_1 &= 0 \\ x + y - s_2 &= 1 \\ x + 2y + s_3 &= 6 \\ y + s_4 &= 4 \\ x + s_5 &= 2 \\ x, y, s_{1,2,3,4,5} &\geq 0 \end{aligned}$$

Se si considera il vertice C , questo corrisponde alla soluzione $x = 2, y = 2, s_2 = 3$. Si può facilmente verificare che le basi

$$\begin{aligned} &\{x, y, s_1, s_2, s_3\} \\ &\{x, y, s_1, s_2, s_5\} \\ &\{x, y, s_2, s_3, s_5\} \end{aligned}$$

5.1. Soluzione grafica di problemi di *PL*

hanno tutte come soluzione di base (degenero) $x = 2, y = 2, s_2 = 3$. Si tratta quindi di 3 diverse basi alle quali corrisponde un unico vertice.

5.1 Soluzione grafica di problemi di *PL*

I problemi di *PL* in \mathbb{R}^2 o in \mathbb{R}^3 possono essere rappresentati e risolti per via grafica. Naturalmente la risoluzione grafica di tali problemi ha esclusivamente interesse didattico e nessuna reale utilità pratica. Per poter risolvere per via grafica un problema di *PL* in \mathbb{R}^2 si può iniziare dalla rappresentazione grafica dell'insieme ammissibile; successivamente si possono aggiungere al grafico alcune curve di livello della funzione obiettivo. Per i problemi di *PL* la funzione obiettivo ha curve di livello esprimibili come

$$H_\ell = \{x \in \mathbb{R}^n : c^T x = \ell\}.$$

Tali curve sono quindi iperpiani, tutti paralleli fra loro, aventi il vettore dei costi c come normale. Si ricorda che il vettore normale c è orientato nella direzione del semispazio $\{c^T x \geq \ell\}$, cioè in direzione crescente della funzione obiettivo. Per risolvere graficamente un problema di *PL* si può quindi rappresentare il vettore dei costi c e disegnare una retta che abbia c come normale. Facendo “scivolare” tale retta parallelamente a se stessa, in direzione opposta a quella indicata da c , fino ad incontrare l'ultimo punto di P si ottiene l'ottimo del problema di *PL*.

5. GEOMETRIA DELLA PROGRAMMAZIONE LINEARE

Esempio 6. Si consideri il problema

$$\begin{aligned} \min & -x - \frac{1}{2}y \\ & -x + y \geq 0 \\ & x + y \geq 1 \\ & x + 2y \leq 6 \\ & y \leq 4 \\ & x \leq 2 \\ & x, y \geq 0 \end{aligned}$$

che ammette la rappresentazione grafica di figura 5.2. Nel disegno sono rappresentate, a titolo di esempio, le curve di livello a livello $0, -1, -2, -3$. Si vede come la curva a livello -3 sia quella di livello minimo tra quelle che intersecano P . Si deduce quindi che il punto C è la soluzione ottimale del problema di PL dato.

Esempio 7. Se nell'esempio precedente l'obiettivo fosse stato

$$\min -\frac{1}{2}x - y$$

la rappresentazione geometrica sarebbe stata quella indicata in figura 5.3. In questo caso la linea di livello minima interseca il poliedro P lungo uno spigolo, CD : esistono pertanto infinite soluzioni ottimali, due delle quali, C e D corrispondenti a vertici e, quindi, a soluzioni di base. Le soluzioni lungo il segmento aperto (C, D) sono ottimali, ma non di base. Si può anche verificare che il vertice D è in corrispondenza con un'unica soluzione di base, mentre il vertice C corrisponde a soluzioni ottimali di base associate a basi differenti.

5.1. Soluzione grafica di problemi di *PL*

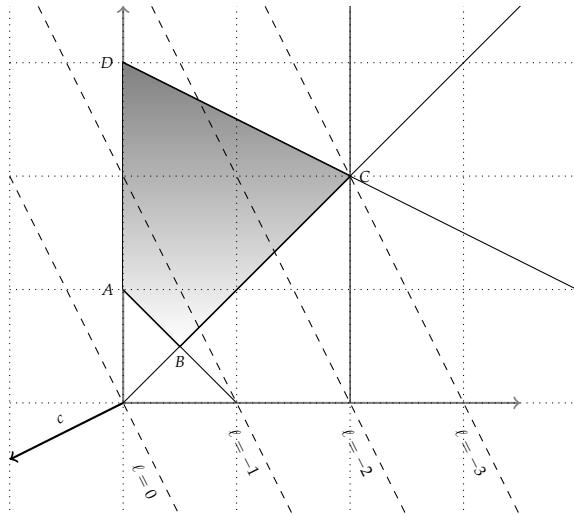


Figura 5.2: Risoluzione geometrica

Esempio 8. Si consideri il problema

$$\begin{aligned} \min \quad & -x - \frac{1}{2}y \\ \text{s.t.} \quad & -x + y \geq 0 \\ & x + y \geq 1 \\ & x, y \geq 0 \end{aligned}$$

rappresentato nella figura 5.4. Questo problema è illimitato inferiormente, come si può anche vedere dal fatto che le curve di livello normali al vettore dei costi $c = [-1 \quad \frac{1}{2}]^T$ intersecano la regione ammissibile a livelli piccoli a piacere. Se nello stesso problema

5. GEOMETRIA DELLA PROGRAMMAZIONE LINEARE

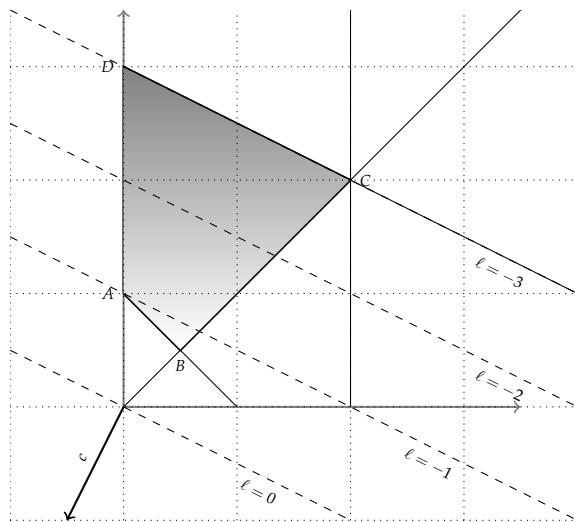


Figura 5.3: Risoluzione geometrica: infinite soluzioni ottimali

l'obiettivo fosse

$$\max -x - \frac{1}{2}y$$

la soluzione ottimale sarebbe il vertice A .

5.1. Soluzione grafica di problemi di *PL*

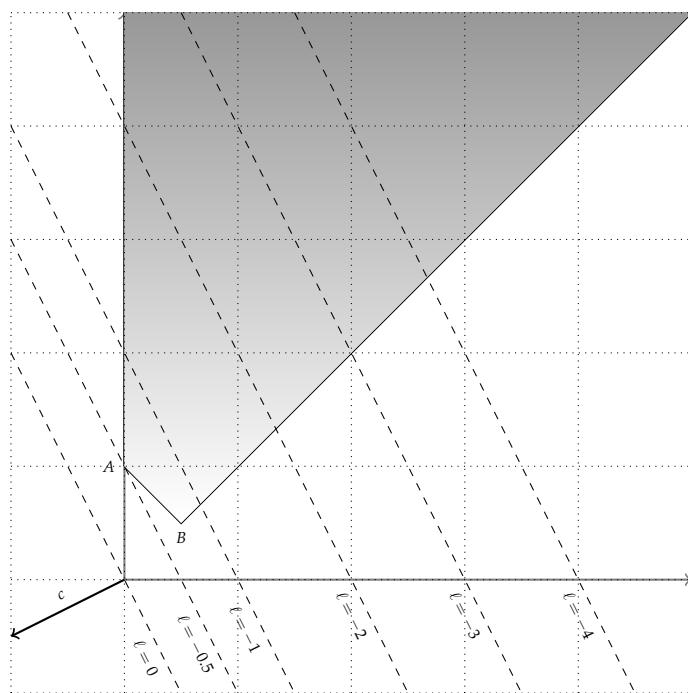


Figura 5.4: Risoluzione geometrica: problema illimitato

Il metodo del simplesso

6.1 Formulazione matriciale

Il metodo del *simplesso* costituisce uno dei più diffusi ed efficienti algoritmi per la risoluzione di problemi di *PL*, anche di notevoli dimensioni. L'algoritmo è legato ai metodi per la risoluzione di sistemi di disequazioni sviluppati, in particolare, da Fourier e Motzkin (si veda ad esempio (Schrijver, 1998) per un resoconto storico molto accurato sulle origini del metodo). Venne sviluppato principalmente da George B. Dantzig nel dopoguerra (Dantzig, 1951; Dantzig & Thapa, 1997, 2003) e valse il premio Nobel per l'Economia a George Koopmans e Leonid Kantorovich (Kantorovich, 1940), ma, incredibilmente, non al suo inventore, Dantzig. I testi dedicati all'ottimizzazione lineare ed al metodo del simplesso sono moltissimi. Senza pretesa di esaustività, se ne citano qui alcuni, oltre ai già citati, quali ad esempio (Chvátal, 1983; Bazaraa, Jarvis, & Sherali, 1990; Bertsimas & Tsitsiklis, 1997; Padberg, 1999; Vanderbei, 2001; Luenberger, 2003).

In questo paragrafo verrà analizzato l'algoritmo nella cosiddetta formulazione matriciale.

Il criterio di base dell'algoritmo consiste nell'analizzare i vertici della regione ammissibile (cioè le soluzioni ammissibili di base) procedendo da un vertice ad un vertice adiacente caratterizzato da

6. IL METODO DEL SIMPLEXSO

un valore della funzione obiettivo non superiore a quello del vertice di partenza.

La struttura dell'algoritmo può essere scomposta in modo naturale nelle seguenti fasi:

1. scelta di una soluzione ammissibile di base iniziale;
2. controllo di ottimalità della soluzione di base corrente;
3. scelta di una nuova soluzione di base.

Il primo punto, cioè la determinazione di una soluzione ammissibile di base con cui iniziare l'algoritmo, viene per il momento lasciato in sospeso e verrà trattato solo come ultimo punto, nel paragrafo 6.3: per determinare una soluzione ammissibile di base con cui iniziare il metodo del simplexso, occorrerà eseguire il metodo del simplexso stesso su un problema ausiliario, con una tecnica che, in altri contesti, viene detta di *bootstrap*. Per ora quindi si supporrà di conoscere una base A_B ammissibile e la corrispondente soluzione ammissibile di base \bar{x} per il problema di *PL* in forma standard

$$\begin{aligned} \min c^T x \\ Ax = b \\ x \geq 0. \end{aligned}$$

Si supporrà, al solito, che A sia una matrice $m \times n$ di rango m , con $m \leq n$. Si ricorda che una sottomatrice A_B di A di dimensioni $m \times m$ costituisce una base ammissibile per il problema se e solo se A_B è invertibile e $\bar{x}_B = A_B^{-1}b \geq 0$.

Ottimalità di una soluzione ammissibile di base

Per comodità di notazione, ma ovviamente senza che ciò sia in pratica necessario, si supporrà che le colonne della matrice dei coefficienti A , le componenti del vettore dei costi c , quelle della generica

6.1. Formulazione matriciale

soluzione x e quelle della soluzione particolare \bar{x} siano state permutate in modo tale che le prime m colonne corrispondano alla base A_B , cioè:

$$\begin{aligned} A &= \left[\begin{array}{cc} A_B & A_N \end{array} \right] \\ x &= \left[\begin{array}{c} x_B \\ x_N \end{array} \right] \\ \bar{x} &= \left[\begin{array}{c} \bar{x}_B \\ 0 \end{array} \right] \end{aligned}$$

dove A_B è una matrice $m \times m$ invertibile, A_N è una matrice $m \times n - m$ e $\bar{x}_B = A_B^{-1}b$.

Nel punto \bar{x} la funzione obiettivo ha valore $\bar{z} = c^T \bar{x}$, esprimibile anche come

$$\begin{aligned} \bar{z} &= \left[\begin{array}{cc} c_B^T & c_N^T \end{array} \right] \left[\begin{array}{c} \bar{x}_B \\ 0 \end{array} \right] \\ &= c_B^T \bar{x}_B \end{aligned}$$

Affinché la soluzione \bar{x} sia ottimale, non deve esistere alcuna altra soluzione ammissibile con valore funzionale strettamente minore di $c_B^T \bar{x}_B$. Ogni soluzione ammissibile deve soddisfare al sistema di equazioni

$$Ax = b$$

che, espresso in funzione della base corrente B , si può scrivere come

$$A_B x_B + A_N x_N = b$$

cioè

$$\begin{aligned} x_B &= A_B^{-1}b - A_B^{-1}A_N x_N \\ &= \bar{x}_B - A_B^{-1}A_N x_N \end{aligned}$$

6. IL METODO DEL SIMPLEXSO

In corrispondenza di una generica soluzione x la funzione obiettivo ha valore

$$c^T x = c_B^T x_B + c_N^T x_N \quad (6.1)$$

$$\begin{aligned} &= c_B^T A_B^{-1} b + (c_N^T - c_B^T A_B^{-1} A_N) x_N \\ &= c^T \bar{x} + \hat{c}_N^T x_N. \end{aligned} \quad (6.2)$$

Gli elementi del vettore

$$\hat{c}_N^T = c_N^T - c_B^T A_B^{-1} A_N$$

sono detti *coefficienti di costo ridotto*. Si ha quindi:

$$z = c^T \bar{x} + \hat{c}_N^T x_N \quad (6.3)$$

$$x_B = \bar{x}_B - A_B^{-1} A_N x_N \quad (6.4)$$

La rappresentazione matriciale (6.3-6.4) viene anche detta *dizionario* del metodo del simplexso.

Poiché ogni soluzione ammissibile deve soddisfare la condizione $x_N \geq 0$, è evidente che la soluzione di base corrente \bar{x} sarà ottimale se nessuno dei coefficienti di x_N è negativo. Si è quindi dimostrato la seguente condizione di ottimalità per la PL.

Teorema 12. *Data una base ammissibile B , se i coefficienti di costo ridotto ad essa associati*

$$\hat{c} = c_N - c_B^T A_B^{-1} A_N$$

sono tutti non negativi, allora la base B è ottimale per il problema di PL.

Esempio 9. Si consideri il problema di PL seguente:

$$\begin{aligned} &\min [2 \quad -1 \quad 3 \quad 1 \quad 0 \quad 0] x \\ &\begin{bmatrix} 1 & 1 & -2 & 1 & 1 & 0 \\ 2 & -1 & 1 & 1 & 0 & 1 \end{bmatrix} x = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \\ &x \geq 0 \end{aligned}$$

6.1. Formulazione matriciale

La soluzione corrispondente alla base $A_{5,6}$ è data da

$$\bar{x}_{5,6} = A_{5,6}^{-1}b = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

ed è caratterizzata da un costo pari a 0 e da coefficienti di costo ridotto

$$\begin{aligned} \hat{c}_{1,2,3,4}^T &= c_{1,2,3,4}^T - c_{5,6}^T A_{5,6}^{-1} A_{1,2,3,4} \\ &= [2 \quad -1 \quad 3 \quad 1] - [0 \quad 0] \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & -2 & 1 \\ 2 & -1 & 1 & 1 \end{bmatrix} \\ &= [2 \quad -1 \quad 3 \quad 1] \end{aligned}$$

Tale soluzione ammissibile di base non soddisfa la condizione sufficiente di ottimalità, essendo il coefficiente di costo ridotto della variabile x_2 negativo. Si può cercare di costruire qualche soluzione di base nella quale la variabile x_2 sia di base. Ad esempio, considerando la base $A_{1,2}$ si ottiene

$$\bar{x}_{1,2} = \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

alla quale sono associati i coefficienti di costo ridotto

$$\begin{aligned} \hat{c}_{3,4,5,6}^T &= [3 \quad 1 \quad 0 \quad 0] - [2 \quad -1] \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}^{-1} \begin{bmatrix} -2 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix} \\ &= [2 \quad 0 \quad 0 \quad -1] \end{aligned}$$

che ancora non soddisfa la condizione sufficiente di ottimalità. Questa soluzione di base ha costo pari a 2, ed è degenere. Provando a costruire una base comprendente sia la variabile x_2 che la variabile x_6 , si ottiene la soluzione di base

$$\bar{x}_{2,6} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

6. IL METODO DEL SIMPLEXSO

alla quale sono associati i coefficienti di costo ridotto

$$\begin{aligned}\hat{c}_{1,3,4,5}^T &= [\ 2 \ 3 \ 1 \ 0 \] - [\ -1 \ 0 \] \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & -2 & 1 & 1 \\ 2 & 1 & 1 & 0 \end{bmatrix} \\ &= [\ 3 \ 1 \ 2 \ 1 \].\end{aligned}$$

Questa soluzione ammissibile di base soddisfa il criterio sufficiente di ottimalità; si può pertanto affermare che, nel problema in esame, una soluzione ottimale è costituita dal vettore

$$\bar{x} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 3 \end{bmatrix}$$

ed il costo ottimale è pari a -1 .

E' importante ricordare che la condizione di ottimalità è *sufficiente, ma non necessaria*.

Esempio 10. Si consideri un problema ottenuto dal precedente tramite una variazione del termine noto:

$$\begin{aligned}\min & [\ 2 \ -1 \ 3 \ 1 \ 0 \ 0 \] x \\ & \begin{bmatrix} 1 & 1 & -2 & 1 & 1 & 0 \\ 2 & -1 & 1 & 1 & 0 & 1 \end{bmatrix} x = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ & x \geq 0;\end{aligned}$$

se si considera di nuovo la soluzione di base associata ad $A_{2,6}$, si ottiene

$$\bar{x}_{2,6} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

6.1. Formulazione matriciale

e i coefficienti di costo ridotto risultano, come nell'esempio precedente, pari a

$$\hat{c}_{1,3,4,5}^T = [\ 3 \ 1 \ 2 \ 1 \].$$

Il costo della soluzione ottimale è quindi pari a 0. Se ora si riprende in considerazione la prima soluzione ammissibile di base individuata nel precedente esempio, e cioè quella associata alla base $A_{5,6}$, si vede immediatamente che tale soluzione di base è ammissibile:

$$\bar{x}_{5,6} = A_{5,6}^{-1}b = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

e, poiché la formula dei coefficienti di costo ridotto non dipende dal termine noto b , ma solo dal vettore dei costi e dalla ripartizione della matrice dei coefficienti in matrice di base e non, i coefficienti di costo ridotto saranno identici a quelli già calcolati:

$$\hat{c}_{1,2,3,4}^T = [\ 2 \ -1 \ 3 \ 1 \].$$

Questa soluzione di base quindi non soddisfa la condizione sufficiente di ottimalità; tuttavia, valutando il costo associato a questa soluzione di base, che è pari a 0, ci si accorge che essa ha lo stesso costo della soluzione ottimale ed è, pertanto, ottimale. Tra l'altro, si può vedere immediatamente che in realtà non si tratta di una diversa soluzione ottimale, ma sempre della stessa, sebbene ottenuta tramite una diversa rappresentazione della base. Si noti che in questo caso la soluzione ottimale di base è *degenera*.

La presenza in un problema di *PL*, di soluzioni ottimali con alcuni coefficienti di costo ridotto negativi non è assolutamente rara – anzi, nei problemi incontrati nella pratica, si può dire che costituisca la norma, non l'eccezione. Tuttavia, più avanti si dimostreranno le seguenti fondamentali proprietà:

Teorema 13. *Se in un problema di PL tutte le soluzioni di base ottimali sono non degeneri, allora ogni soluzione ottimale di base avrà coefficienti di*

6. IL METODO DEL SIMPLEXSO

costo ridotto non negativi. La condizione di non negatività dei coefficienti di costo ridotto è, in questo caso, necessaria e sufficiente per l'ottimalità.

Teorema 14. *In ogni problema di PL che ammette almeno una soluzione ottimale di base, esiste almeno una base ottimale alla quale sono associati coefficienti di costo ridotto tutti non negativi.*

In base a queste proprietà e, in particolare, all'ultima, si può pensare di costruire un algoritmo che utilizzi la condizione di non negatività dei coefficienti di costo ridotto come regola d'arresto. Naturalmente se si potesse riconoscere l'ottimalità di una soluzione di base anche in presenza di coefficienti di costo ridotto negativi, la velocità di esecuzione dell'algoritmo sarebbe maggiore; in certi casi questo è possibile: a volte, ad esempio, è noto a priori il *valore* dell'ottimo, ma non l'ottimo stesso. In questi casi si può arrestare l'algoritmo appena si incontra una soluzione ammissibile con costo pari all'ottimo. Più spesso è nota una limitazione inferiore al valore dell'ottimo: anche in questi casi ci si può arrestare se si incontra una soluzione ammissibile il cui costo è pari a tale limitazione inferiore. In generale tuttavia gli algoritmi del tipo del metodo del simplexso attendono di aver determinato una soluzione di base con tutti i coefficienti di costo ridotto non negativi (o di aver dimostrato che tale soluzione non può esistere).

Scelta di una nuova soluzione ammissibile di base - operazione di pivot

Se, come supporremo qui, si dispone di una soluzione ammissibile di base con almeno un coefficiente di costo ridotto negativo, occorre un metodo per modificare, se possibile, la base corrente in una nuova cui corrisponda una soluzione "migliore" di quella attuale. Questo criterio si compone di due fasi: la scelta di una *direzione* e la successiva scelta di uno *spostamento* lungo tale direzione. In genere si cercherà di determinare una direzione *ammissibile*, che come si

6.1. Formulazione matriciale

è già visto in precedenza, rappresenta una direzione lungo la quale sia possibile effettuare uno spostamento non nullo senza violare alcun vincolo. Sia

$$\bar{x} = \begin{bmatrix} \bar{x}_B \\ 0 \end{bmatrix}$$

la soluzione di base corrente. L'obiettivo di questo paragrafo è la determinazione di una nuova soluzione ammissibile

$$\tilde{x} = \begin{bmatrix} \tilde{x}_B \\ \tilde{x}_N \end{bmatrix}$$

che sia ancora di base e non peggiore di \bar{x} , nel senso che $c^T \tilde{x} \leq c^T \bar{x}$. Si noti che la suddivisione di \tilde{x} nelle componenti \tilde{x}_B ed \tilde{x}_N è fatta, in questo momento, rispetto agli indici della base B corrispondente ad \bar{x} .

Come si è visto, per mantenere l'ammissibilità, le componenti \tilde{x}_B e \tilde{x}_N della nuova soluzione \tilde{x} devono soddisfare le relazioni:

$$\begin{aligned} \tilde{x}_B &= A_B^{-1}b - A_B^{-1}A_N\tilde{x}_N \\ &= \bar{x}_B - A_B^{-1}A_N\tilde{x}_N \\ \tilde{x} &\geq 0 \end{aligned}$$

cioè

$$\begin{bmatrix} \bar{x}_B - A_B^{-1}A_N\tilde{x}_N \\ \tilde{x}_N \end{bmatrix} \geq 0,$$

ovvero

$$\begin{bmatrix} \bar{x}_B \\ 0 \end{bmatrix} + \begin{bmatrix} -A_B^{-1}A_N \\ I \end{bmatrix} \tilde{x}_N \geq 0.$$

La scelta $\tilde{x}_N = 0$ corrisponde alla soluzione di base corrente, cioè a $\tilde{x} = \bar{x}$. Ricordando che, dalla definizione dei coefficienti di costo ridotto, si ha

$$c^T \tilde{x} = c^T \bar{x} + \hat{c}_N^T \tilde{x}_N,$$

6. IL METODO DEL SIMPLEXSO

se la soluzione \bar{x} non è ottimale esisterà certamente almeno una componente, ad esempio quella di indice $j \in N$, del vettore dei coefficienti di costo ridotto strettamente negativa, $\hat{c}_j < 0$. La scelta del metodo del simplexso corrisponde ad effettuare lo spostamento ammissibile più semplice, consistente nel modificare una sola componente fuori base. Incrementando (se possibile) la componente del vettore \tilde{x}_N corrispondente al coefficiente di costo ridotto negativo, il valore della funzione obiettivo diminuirà. Indicando con δ l'incremento della variabile \tilde{x}_j , si ha

$$\tilde{x} = \tilde{x}(\delta) = \begin{bmatrix} \bar{x}_B \\ 0 \end{bmatrix} + \begin{bmatrix} -A_B^{-1} A_N \\ I_{n-m} \end{bmatrix} e_j \delta$$

dove I_{n-m} rappresenta una matrice identica di dimensione $(n-m) \times (n-m)$ ed e_j rappresenta il versore di \mathbb{R}^{n-m} le cui componenti sono tutte nulle ad eccezione di quella corrispondente a x_j che vale 1. Ponendo

$$d = -A_B^{-1} A_N e_j \in \mathbb{R}^m$$

si ottiene

$$\tilde{x}(\delta) = \bar{x} + \begin{bmatrix} d \\ e_j \end{bmatrix} \delta.$$

Per costruzione, qualunque sia il valore di δ , il punto $\tilde{x}(\delta)$ è soluzione del problema di PL , cioè soddisfa i vincoli di eguaglianza. Per essere ammissibile deve inoltre soddisfare la condizione di non negatività $\tilde{x} \geq 0$. Scegliendo $\delta \geq 0$ le componenti di \tilde{x}_N sono non negative (anzi, ne esiste soltanto una, quella di indice j , che potrebbe divenire strettamente positiva). Per quanto riguarda le componenti di \tilde{x}_B , si ha invece

$$\tilde{x}_i = \bar{x}_i + d_i \delta \quad i \in B \tag{6.5}$$

Ricordando che si suppone che la soluzione di base corrente sia ammissibile, e che pertanto \bar{x}_i sia non negativo, si ha, analizzando i casi possibili:

6.1. Formulazione matriciale

- se $d_i \geq 0$, per qualunque valore di δ la corrispondente variabile \tilde{x}_i sarà non negativa;
- se $d_i < 0$, incrementando δ la corrispondente componente di x_i diminuisce; per imporre l'ammissibilità si deve pertanto richiedere che $\bar{x}_i + d_i\delta \geq 0$ e cioè che

$$\delta \leq -\frac{\bar{x}_i}{d_i} \quad \forall i : d_i < 0.$$

L'ammissibilità della nuova soluzione è quindi garantita per ogni scelta di δ tale da soddisfare

$$0 \leq \delta \leq \delta^* = \min_{i:d_i < 0} \left\{ -\frac{\bar{x}_i}{d_i} \right\}. \quad (6.6)$$

Poiché maggiore è il valore di δ , maggiore sarà il decremento nel valore della funzione obiettivo (si veda la (6.2)), ponendo

$$\delta = \delta^*$$

si ottiene il massimo decremento possibile per la funzione obiettivo lungo la direzione

$$\begin{bmatrix} d \\ e_j \end{bmatrix} \quad (6.7)$$

Determinazione della variabile uscente di base

Se il valore δ^* così individuato è strettamente positivo, la (6.6) individua uno spostamento non nullo lungo la direzione (6.7). Indicando con $i^* \in B$ un indice in corrispondenza del quale viene raggiunto il minimo nella (6.6), cioè quell'indice (o uno di quegli indici) tale che

$$\delta^* = -\frac{\bar{x}_{i^*}}{d_{i^*}} \leq -\frac{\bar{x}_i}{d_i} \quad \forall i \in B : d_i < 0$$

6. IL METODO DEL SIMPLEXSO

cioè

$$i^* \in \arg \min_{i \in B: d_i < 0} \left\{ -\frac{\bar{x}_i}{d_i} \right\}$$

si vede immediatamente che la variabile di base x_{i^*} assume valore nullo e che, se $\delta^* > 0$, la variabile di indice $j \in N$ acquista un valore strettamente positivo. Si dimostrerà ora che la matrice ottenuta sostituendo la colonna dei coefficienti A_{j^*} con quella dei coefficienti di A_j è ancora una base ammissibile. Si suole dire in questo caso che la variabile di base x_{i^*} "esce" dalla base, mentre la variabile x_j "entra" nella nuova base. L'operazione consistente nello scambio tra queste due variabili, di cui una in base e l'altra fuori base, si chiama operazione di *pivot* o di "*cardine*"; tale operazione corrisponde, nell'interpretazione geometrica della *PL*, ad uno spostamento da un vertice ad un vertice adiacente lungo uno spigolo del poliedro. L'operazione di pivot può anche essere effettuata qualora si abbia $\delta^* = 0$ – cosa che può accadere solamente nel caso di basi degeneri – nel qual caso il passo effettuato è nullo. La soluzione di base resta invariata, ma cambia la composizione della matrice di base.

Nello scambio tra il vettore dei coefficienti di x_{i^*} (variabile uscente di base) e quelli di x_j (variabile entrante in base), non si perde l'invertibilità della matrice A_B . Vale infatti un'importante proprietà generale sulla sostituzione di vettori in matrici invertibili (cioè in basi):

Teorema 15. *Sia $A_B = \{A_1, A_2, \dots, A_m\} \in \mathbb{R}^{m \times m}$ una base di \mathbb{R}^m e sia $v \in \mathbb{R}^m$ un vettore generico. Siano $\lambda_1, \dots, \lambda_m$ gli unici coefficienti di una combinazione lineare delle colonne di A_B che genera v :*

$$v = \sum_{i=1}^m A_i \lambda_i$$

Allora la matrice $A_{B'} = A_B \setminus A_j \cup v$ (ottenuta scambiando v con una colonna A_j di A_B) è ancora una base se e solo se $\lambda_j \neq 0$.

6.1. Formulazione matriciale

Dimostrazione. La condizione $\lambda_j \neq 0$ è sufficiente. Infatti, si considerino le colonne

$$A_{B'} = \{A_1, A_2, \dots, A_{j-1}, v, A_{j+1}, \dots, A_m\}$$

e se ne effettui una combinazione lineare con coefficienti μ_1, \dots, μ_m :

$$\sum_{i \neq j} A_i \mu_i + v \mu_j \quad (6.8)$$

Sfruttando la rappresentazione di v mediante le colonne di A_B questa è equivalente a

$$\sum_{i \neq j} A_i \mu_i + \mu_j \sum_i A_i \lambda_i$$

cioè a

$$\sum_{i \neq j} (\mu_i + \mu_j \lambda_i) A_i + \mu_j \lambda_j A_j$$

Poiché A_B è una base, l'unica combinazione lineare delle sue colonne che produce il vettore nullo è quella banale. Per annullare la combinazione (6.8) occorre avere

$$\begin{aligned} \mu_j \lambda_j &= 0 \\ \mu_i + \mu_j \lambda_i &= 0 \quad i = 1, \dots, m, i \neq j \end{aligned}$$

ed, essendo per ipotesi $\lambda_j \neq 0$, dovrà essere $\mu_j = 0$. Ma, essendo $\mu_j = 0$, dovranno essere nulli anche tutti i coefficienti $\mu_i, i \in \{1, \dots, m\}, i \neq j$. Pertanto l'unica combinazione lineare che annulla le colonne di $A_{B'}$ è quella banale tutta nulla: $A_{B'}$ è quindi una base.

La condizione è anche necessaria. Infatti, se si avesse $\lambda_j = 0$, allora nella rappresentazione di v mediante A_B si avrebbe

$$v = \sum_{i \neq j} A_i \lambda_i$$

Quindi v sarebbe dipendente linearmente dai vettori $A_i, i \neq j$. Non potrebbe pertanto formare, con loro, una base. \square

6. IL METODO DEL SIMPLEXSO

Questa proprietà ci permette di caratterizzare tutti e soli gli scambi possibili tra un vettore in base ed uno esterno alla base stessa. Per poter applicare quanto visto all'operazione di scambio nel simplexso, osserviamo che la scelta del vettore uscente di base avviene tramite il criterio

$$i^* \in \arg \min_{i \in B: d_i < 0} \left\{ -\frac{\bar{x}_i}{d_i} \right\} \quad (6.9)$$

con $d = -A_B^{-1}A_j$, dove j è l'indice della variabile entrante in base.

Il vettore d non è altro che (a meno di un cambio di segno) il vettore dei coefficienti nella rappresentazione di A_j tramite le colonne di A_B . Infatti si osservi che

$$\sum_{i \in B} A_i d_i = A_B d = A_B (-A_B^{-1}A_j) = -A_j.$$

Il vettore al posto del quale $v = A_j$ viene inserito in base compare esplicitamente in questa rappresentazione (in base al criterio (6.9)) il coefficiente d_{i^*} è sicuramente non nullo, anzi, strettamente negativo). Quindi l'operazione di pivot è valida, nel senso che trasforma basi in basi. In particolare, data la scelta del segno di d_i , in realtà tale operazione trasforma basi *ammissibili* in basi *ammissibili*.

L'operazione di pivot appena descritta viene eseguita in modo tale che la funzione obiettivo non aumenti; in particolare, se $\delta^* > 0$ è finito, la funzione obiettivo nel nuovo vertice avrà valore *strettamente minore* di $c_B^T \bar{x}_B$.

Esistono due casi in cui non è possibile ottenere un valore $\delta^* > 0$ finito:

1. $d_i \geq 0 \quad \forall i \in B$: in questo caso la componente di indice j di x può essere incrementata a piacere senza che alcun vincolo sia violato; ciò implica che il costo $c^T x$ possa essere diminuito a piacere incrementando x_j (si veda la (6.2)); pertanto in questo caso il problema di PL è *illimitato*.

6.1. Formulazione matriciale

In questo caso cioè ciascuna componente del vettore

$$\tilde{x}_i = \bar{x}_i + d_i \delta \quad i \in B$$

resta non negativa per qualsiasi scelta di $\delta \geq 0$. Poichè ad uno spostamento $\delta > 0$ corrisponde una variazione della funzione obiettivo pari a $\hat{c}_j \delta < 0$, si vede, in questo caso, che il problema risulta illimitato.

2. $\exists i \in B : \tilde{x}_i = 0, d_i < 0$; in questo caso (ovviamente possibile solo in presenza di una soluzione di base degenere), lo spostamento δ^* sarà necessariamente pari a zero; né il vettore soluzione \tilde{x} né il valore della funzione obiettivo quindi varieranno. Tuttavia sarà ancora possibile effettuare l'operazione di "pivot": in questo caso l'operazione avrà come effetto semplicemente quello di cambiare la composizione delle colonne che formano la base A_B . Si dice che in questo caso la variabile x_j entra nella nuova base "a livello zero".

Il trattamento del caso degenere è particolarmente importante sia dal punto di vista dell'analisi teorica del metodo del simplesso che dal punto di vista della realizzazione di algoritmi efficienti e delle implementazioni del metodo stesso: la presenza di soluzioni degeneri, come si è accennato, potrebbe far ciclare l'algoritmo. Vale un'importante, elementare, proprietà:

Teorema 16. *Si consideri una versione deterministica del metodo del simplesso, implementato in modo tale che le decisioni relative alle variabili entranti ed uscenti di base ad ogni iterazione dipendano esclusivamente dalla composizione della base corrente. Se tale algoritmo non termina, allora deve ciclare all'infinito.*

Dimostrazione. La dimostrazione segue banalmente dal fatto che le basi di un problema di PL sono in numero finito; pertanto un metodo del simplesso che non terminasse dovrebbe visitarne almeno

6. IL METODO DEL SIMPLEXSO

una più volte. Se l'evoluzione del metodo è completamente determinata dalla base corrente, allora l'algoritmo ripeterà ciclicamente all'infinito una stessa sequenza di basi. \square

La richiesta che il metodo sia deterministicio nasce dal fatto che se in qualche fase del metodo, come ad esempio la scelta della variabile entrante o uscente di base, si inserisse una decisione di tipo randomizzato, allora sarebbe possibile il ripetersi non ciclico di un insieme di basi. In ogni caso, deterministicio o non, dato l'andamento monotono dei valori delle funzioni obiettivo nelle basi successivamente generate dal metodo, la non terminazione è possibile esclusivamente nel caso in cui nel problema vi siano delle basi degeneri. Solo in tale caso infatti è possibile una ripetizione di basi. Vale cioè la seguente proprietà elementare:

Teorema 17. *Se nessuna base incontrata nel corso dell'esecuzione del metodo del simplexso è degenera, allora il metodo termina in un numero finito di iterazioni.*

La dimostrazione è banale.

Esempio 11. Si consideri il problema

$$\begin{aligned} \min & [1 \ 0 \ 1 \ 2 \ -1 \ -1] x \\ & \left[\begin{array}{cccccc} 1 & 0 & 2 & -1 & 1 & -1 \\ 0 & 1 & -1 & 1 & -1 & -1 \end{array} \right] x = \left[\begin{array}{c} 1 \\ 2 \end{array} \right] \\ & x \geq 0. \end{aligned}$$

Utilizzando come base iniziale la matrice identica $A_{1,2}$, corrispondente alla soluzione di base $x_{1,2} = \left[\begin{array}{c} 1 \\ 2 \end{array} \right]$, si ha

$$\begin{aligned} \hat{c}_{3,4,5,6} &= [1 \ 2 \ -1 \ -1] - [1 \ 0] \left[\begin{array}{cccc} 2 & -1 & 1 & -1 \\ -1 & 1 & -1 & -1 \end{array} \right] \\ &= [-1 \ 3 \ -2 \ 0]. \end{aligned}$$

6.1. Formulazione matriciale

Si può scegliere, come variabile entrante, sia x_3 che x_5 . Si vedrà più avanti che entrambe queste scelte sono corrette. Si scelga, ad esempio, x_5 . Per la scelta della variabile uscente, si ha

$$d = -A_{1,2}^{-1}A_5 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

L'unica possibilità per la variabile uscente è x_1 . Lo scambio avviene dunque tra x_1 e x_5 ; la nuova base diventa

$$A_{2,5} = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix}$$

e la soluzione di base ad essa associata è

$$x_{2,5} = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}.$$

I coefficienti di costo ridotto associati a questa base sono

$$\begin{aligned} \hat{c}_{1,3,4,6} &= [1 \ 1 \ 2 \ -1] - [0 \ -1] \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 2 & -1 & -1 \\ 0 & -1 & 1 & -1 \end{bmatrix} \\ &= [2 \ 3 \ 1 \ -2] \end{aligned}$$

La variabile entrante è quindi x_6 . Per determinare la variabile uscente, si cerca, al solito, il vettore direzione

$$d = -A_{2,5}^{-1}A_6 = -\begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix}^{-1} \begin{bmatrix} -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Nessuna delle componenti di d è negativa. Pertanto il problema è illimitato. Per confermare questo fatto, si osservi che la soluzione

6. IL METODO DEL SIMPLEX

di base corrente corrisponde al vettore

$$\bar{x} = \begin{bmatrix} 0 \\ 3 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

e che, attraverso lo spostamento individuato dal vettore d appena calcolato, sono ammissibili tutte le soluzioni

$$x(\delta) = \begin{bmatrix} 0 \\ 3 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \delta = \begin{bmatrix} 0 \\ 3+2\delta \\ 0 \\ 0 \\ 1+\delta \\ \delta \end{bmatrix},$$

il cui costo è $z(\delta) = -(1+\delta) - \delta = -1 - 2\delta \rightarrow -\infty$ per $\delta \rightarrow +\infty$.

La (6.5) fornisce una prova costruttiva dell'illimitatezza del problema e, in alcuni casi, può aiutare a indicarne il motivo; in un problema applicativo, in genere l'illimitatezza è causata da vincoli troppo deboli o assenti. L'individuazione di una direzione di spostamento illimitata in discesa può aiutare nella scelta di un vincolo da aggiungere al problema in modo da ottenere un ottimo finito.

Scelta della variabile entrante in base

Nel caso in cui più coefficienti di costo ridotto siano negativi, si pone il problema di determinare quale sarà l'indice del vettore entrante in base. Esistono varî metodi per compiere questa scelta. Sia \hat{j} l'indice della variabile entrante in base;

1. uno dei metodi più semplici, ma relativamente poco utilizzati, è la scelta casuale dell'indice \hat{j} nell'insieme $N \cap \{j : \hat{c}_j < 0\}$.

6.1. Formulazione matriciale

2. $\hat{j} = \min\{j \in N : \hat{c}_j < 0\}$, che corrisponde alla scelta del *primo* coefficiente di costo ridotto negativo osservato scandendo gli elementi dell'insieme N .
3. $\hat{j} \in \arg \min_{j \in N} \{\hat{c}_j\}$. Questo criterio, consistente nella scelta della variabile cui è associato il coefficiente di costo ridotto minimo, è utilizzato in molte implementazioni. Il criterio è anche noto con il nome di “metodo del gradiente nello spazio delle variabili non di base”, in quanto, osservando la (6.2), si vede come il gradiente della funzione obiettivo rispetto alle variabili non di base sia proprio il vettore \hat{c} ; un incremento della componente della soluzione cui è associato il coefficiente di costo ridotto minimo equivale quindi alla scelta di una direzione lungo l'asse cartesiano associato alla più ripida discesa della funzione obiettivo rappresentata in funzione delle variabili con indice in N . Si noti tuttavia che il solo fatto di “scendere” lungo la direzione più ripida non è garanzia di efficienza, in quanto il decremento effettivo della funzione obiettivo è funzione non solo della direzione, ma anche dello spostamento possibile lungo tale direzione. Addirittura, nel caso di soluzione degenere, lo spostamento potrebbe essere nullo. Inoltre, la scelta è solo locale e non permette di prevedere la qualità dei passi successivi. Inoltre non è detto che un passo di discesa ripida porti necessariamente ad una terminazione più rapida dell'algoritmo.
4. Uno dei metodi più utilizzati nelle implementazioni attuali è un compromesso tra i due metodi precedenti – si sceglie cioè la variabile entrante tra quelle che possiedono il coefficiente di costo minimo in un sottoinsieme costituito da un certo numero, k , degli indici di N associati a coefficienti di costo ridotto negativo.

6. IL METODO DEL SIMPLEX

È da sottolineare il fatto che nessuno dei metodi sopra illustrati si configura come “ottimale”; è possibile costruire esempi in cui ciascuno dei criteri sopra esposti porta all’esplorazione di un numero eccessivo di soluzioni di base. Il problema dell’esistenza o meno di strategie di pivoting efficienti è un problema tuttora oggetto di studio (si veda in proposito la discussione sull’efficienza del metodo del simplex nel paragrafo 6.5).

In tutti questi metodi, nel caso di ulteriore incertezza nella scelta, si adotta un qualunque criterio di “tie breaking”, ad esempio scegliendo il coefficiente di indice minimo oppure estraendo a caso.

E opportuno precisare che, nonostante i metodi suesposti richiedano costi computazionali differenti, il vantaggio in termini di tempi globali di esecuzione non sempre presenta significative differenze. Se però il problema di *PL* è dotato di particolare struttura, alcune scelte diventano vantaggiose rispetto ad altre.

Scelta del vettore uscente dalla base

La scelta della variabile uscente di base non è molto critica nel caso di soluzioni di base non degeneri; il caso in cui può esistere incertezza nella scelta della variabile uscente di base corrisponde alla situazione in cui esistono almeno due variabili di base, \bar{x}_ℓ, \bar{x}_k tali che

$$\delta^* = -\frac{\bar{x}_\ell}{d_\ell} = -\frac{\bar{x}_k}{d_k}.$$

In questo caso in genere si può scegliere indifferentemente l’una o l’altra come variabile uscente di base. Si può però vedere che, in ogni caso, la soluzione di base che si ottiene al termine dell’operazione di pivot, è degenere.

Se la soluzione di base corrente è degenere, è possibile che l’operazione di pivot avvenga a livello 0, cioè con $\delta^* = 0$. In questo caso, e solo in questo caso, è necessario utilizzare accorgimenti op-

6.1. Formulazione matriciale

portuni per evitare i cicli infiniti di operazioni di pivot e, quindi, la non terminazione del metodo.

Si può dimostrare che, nel caso di soluzioni di base degeneri, è possibile un ciclo infinito di iterazioni solo se almeno 2 variabili sono in base a livello zero e corrispondono a elementi negativi del vettore d . In questo caso occorre un criterio per decidere quale delle variabili a livello zero deve uscire di base. Il metodo più semplice, anche se non molto utilizzato nella pratica, è quello di scegliere a caso (uniformemente): il metodo garantisce l'uscita dal ciclo con probabilità uno in un numero finito di iterazioni.

Una tecnica più usata è la cosiddetta *regola di Bland*, che prescrive sia la scelta del vettore entrante che quella del vettore uscente di base. Per il vettore entrante si sceglie il primo vettore il cui coefficiente di costo ridotto sia negativo (è il criterio 2 visto nel precedente paragrafo); analogamente, per il vettore uscente, la scelta corrisponde al primo tra gli indici che realizzano il minimo nella (6.6). Cioè si sceglie come variabile uscente la variabile di indice i^* caratterizzata da:

$$i^* = \min \{i \in B : d_i < 0, \bar{x}_i = 0\}.$$

Naturalmente niente vieta di usare la regola di Bland anche nel caso di basi non degeneri: basta in questo caso scegliere la variabile uscente con indice minimo tra quelle che determinano il valore δ^* ; tuttavia nella pratica si usa generalmente una regola differente (soprattutto per la scelta della variabile entrante) e si passa alla regola di Bland solamente quando si incontra un pivot degenero.

Questo algoritmo garantisce l'impossibilità di cicli infiniti e, come conseguenza, la finitezza dell'algoritmo del simplex.

6. IL METODO DEL SIMPLEXSO

Esempio 9 (cont.)

Si consideri l'esempio di pag. 86:

$$\begin{aligned} & \min [2 \ -1 \ 3 \ 1 \ 0 \ 0] x \\ \left[\begin{array}{cccccc} 1 & 1 & -2 & 1 & 1 & 0 \\ 2 & -1 & 1 & 1 & 0 & 1 \end{array} \right] x = \left[\begin{array}{c} 1 \\ 2 \end{array} \right] \\ & x \geq 0 \end{aligned}$$

Si è visto che la soluzione $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ corrispondente alla base $A_{5,6}$ è caratterizzata da un costo pari a 0 e dai coefficienti di costo ridotto

$$\hat{c}_{1,2,3,4}^T = [2 \ -1 \ 3 \ 1]$$

Poiché il coefficiente di costo ridotto corrispondente alla seconda variabile non di base, cioè alla variabile x_2 , è negativo, è possibile effettuare un'operazione di pivot per cercare, se possibile, di diminuire il valore della funzione obiettivo. Per determinare qual è la variabile uscente dalla base si opera come segue:

$$\begin{aligned} d &= -A_B^{-1} A_N e_2 \\ &= -\left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right]^{-1} \left[\begin{array}{cccc} 1 & 1 & -2 & 1 \\ 2 & -1 & 1 & 1 \end{array} \right] \left[\begin{array}{c} 0 \\ 1 \\ 0 \\ 0 \end{array} \right] \\ &= \left[\begin{array}{c} -1 \\ +1 \end{array} \right] \\ \delta^* &= \min_{i \in B: d_i < 0} \left\{ -\frac{x_i}{d_i} \right\} = -\frac{1}{-1} = 1 \end{aligned}$$

La variabile uscente di base è dunque la prima di base, cioè x_5 . La nuova base sarà quindi $B = \{2, 6\}$. La variabile entrante in base (x_2) entrerà a livello $\delta^* = 1$.

6.1. Formulazione matriciale

Si vede facilmente (riferendosi all'esempio 9) che questa soluzione di base è ottimale.

Esempio 10 (cont.)

$$\begin{aligned} \min & \begin{bmatrix} 2 & -1 & 3 & 1 & 0 & 0 \end{bmatrix} x \\ & \begin{bmatrix} 1 & 1 & -2 & 1 & 1 & 0 \end{bmatrix} x = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ & x \geq 0 \end{aligned}$$

si considerino la soluzione di base associata a $A_{5,6}$, con $\bar{x}_{5,6} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ ed i coefficienti di costo ridotto $\hat{c}_{1,2,3,4}^T = [2 \ -1 \ 3 \ 1]$.

E' possibile effettuare un'operazione di pivot facendo entrare in base la variabile x_2 . Per la determinazione della variabile uscente, si procede in modo simile a quanto visto nell'esempio precedente:

$$\begin{aligned} d &= -A_{5,6}^{-1} A_2 \\ &= \begin{bmatrix} -1 \\ +1 \end{bmatrix} \\ \delta^* &= \min_{i \in B: d_i < 0} \left\{ -\frac{x_i}{d_i} \right\} = -\frac{0}{-1} = 0 \end{aligned}$$

In questo caso il passo di pivot è degenere. Tuttavia lo scambio tra la variabile entrante in base e quella uscente, x_5 , viene comunque effettuato. La nuova base $B = \{2, 6\}$, pur corrispondendo alla stessa soluzione ($x_1 = x_2 = x_3 = x_4 = x_5 = 0, x_6 = 1$), permette

6. IL METODO DEL SIMPLEXSO

di certificarne l'ottimalità, avendo tutti i coefficienti di costo ridotto non negativi.

Soluzioni di base e vertici degeneri

Si consideri il sistema di equazioni che permette di determinare la soluzione di base, $A_B x_B = b$, equivalente a $\sum_{i \in B} A_i x_i = b$. Nel caso di una base A_B degenera, se $B' \subset B$ è l'insieme di indici delle componenti non nulle della soluzione di base, si vede che il sistema si riduce a $A_{B'} x_{B'} = b$. Esistono in genere diversi insiemi di colonne della matrice A che, accostate ad $A_{B'}$ formano una base. Si vede quindi che, nel caso degenero, una stessa soluzione di base può essere ottenuta attraverso basi differenti.

Dal punto di vista geometrico, nel caso non degenero una soluzione di base si trova all'intersezione degli m iperpiani con $n - m$ vincoli del tipo $x_j = 0$ e, pertanto, è l'unica soluzione di un sistema quadrato di equazioni lineari:

$$\begin{bmatrix} A_B & A_N \\ 0_{n-m,m} & I_{n-m} \end{bmatrix} x = \begin{bmatrix} b \\ 0 \end{bmatrix},$$

dove $0_{h,k}$ ed I_h rappresentano rispettivamente una matrice nulla $h \times k$ ed una matrice identica $h \times h$.

Nel caso degenero invece la soluzione di base corrente si può ottenere come soluzione di un sistema non più quadrato; indicando con N' l'insieme delle colonne di A con l'eccezione di B' , si ha

$$\begin{bmatrix} A_B & A_N \\ 0_{|N'|,|B'|} & I_{|N'|} \end{bmatrix} x = \begin{bmatrix} b \\ 0 \end{bmatrix};$$

il sistema sopra scritto contiene più equazioni che incognite ($|m + N'| > n$). Poiché si sa che il sistema ammette soluzione (la soluzione di base corrente), se ne deduce che almeno una di tali equazioni è ridondante. Ciò significa, geometricamente, che il vertice corrispondente alla base si trova all'intersezione di un numero superiore ad

6.1. Formulazione matriciale

n di iperpiani. Risulta quindi che, *per la determinazione del vertice in questione*, almeno un vincolo è ridondante. Si osservi che questo non significa che nel problema originale esistono vincoli eliminabili. Immaginando ad esempio una piramide a base pentagonale in \mathbb{R}^3 , si vede che un vertice di tale piramide si trova all'intersezione di ben 5 piani, mentre tutti gli altri si trovano all'intersezione di 3 piani. Il primo vertice è degenere – le sue coordinate potrebbero essere individuate in modo univoco come intersezione di 3 qualunque (purché indipendenti) dei piani che definiscono le facce della piramide. Tuttavia nessuna delle 5 facce può essere eliminata senza modificare la forma dell'insieme. Nel caso di una base degenere è possibile che un'operazione di pivot avvenga “a livello 0”, cioè con spostamento nullo e senza variazione della funzione obiettivo. In questo caso l'operazione di pivot non porta ad un nuovo vertice, ma semplicemente all'individuazione del vertice corrente mediante l'intersezione di un diverso insieme di iperpiani di supporto, cioè di una diversa base. Nello schema 1 viene riportato il metodo del

```

Data:  $B$ : una base ammissibile
while  $\hat{c} := c_N^T - c_B^T A_B^{-1} A_N \not\geq 0$  do
    | Sia  $j \in N : \hat{c}_j < 0$ ;
    | Sia  $d := -A_B^{-1} A_j$ ;
    | if  $d \geq 0$  then
    |   | return “problema illimitato”;
    | end
    | Sia  $i^* \in B \in \arg \min_{i \in B: d_i < 0} \left\{ -\frac{x_{B_i}}{d_i} \right\}$ ;
    |  $B \leftarrow B \cup \{j\} \setminus \{i^*\}$ ;
end
return “ $B$ : base ottimale”;
```

Algoritmo 1: Algoritmo del simplex (fase II)

simplex, nell'ipotesi di conoscere una base ammissibile, da cui il

6. IL METODO DEL SIMPLEXSO

nome “fase II”, in quanto, in generale, sarà necessario eseguire una prima fase consistente nella ricerca di una base ammissibile.

Finitezza del metodo del simplexso

Teorema 18. *Il metodo del simplexso, equipaggiato con la regola di Bland, termina sempre in un numero finito di iterazioni.*

Dimostrazione. Per la dimostrazione è sufficiente mostrare che, grazie alla regola di Bland, il metodo del simplexso non entra in ciclo; si ipotizzi, per assurdo, che l’algoritmo non termini. Dovrà quindi esistere una sequenza ciclica finita di basi generata dal metodo; siano

$$B_0, B_1, \dots, B_{k-1}, B_0$$

gli insiemi di indici corrispondenti a tali basi. Sia

$$t = \max \left\{ i \in \bigcup_{r=0}^{k-1} B_r : i \notin \bigcap_{r=0}^{k-1} B_r \right\}. \quad (6.10)$$

In altre parole, t è il massimo tra tutti gli indici che compaiono in almeno una, ma non in tutte le basi della sequenza. Nel corso delle operazioni di pivot che portano in successione alle basi B_0, B_1, \dots, B_{k-1} , ad un certo punto la variabile t uscirà di base. Si può sempre immaginare, data la ciclicità della sequenza, che t esca di base nel passaggio da B_0 a B_1 . Si indichi con s la variabile che, nel corso di questa operazione di pivot, entra in base.

Il dizionario corrispondente alla base B_0 è il seguente:

$$\begin{aligned} x_{B_0} &= b^0 - A_{B_0}^{-1} A_{N_0} x_{N_0} \\ z &= \bar{z} + \hat{c}_{N_0}^T x_{N_0} \end{aligned}$$

dove b^0 e \bar{z} indicano rispettivamente la soluzione di base corrente ed il suo valore. In base alle ipotesi fatte $t \in B_0, s \in N_0$. Per come

6.1. Formulazione matriciale

è stato definito t , esisterà un altro dizionario, corrispondente ad esempio alla base B_k , in cui la variabile x_t , dopo essere uscita di base, rientrerà. Si consideri il dizionario associato a tale base:

$$\begin{aligned}x_{B_k} &= b^k - A_{B_k}^{-1} A_{N_k} x_{N_k} \\z &= \bar{z} + \hat{c}_{N_k}^T x_{N_k}.\end{aligned}$$

Poichè le operazioni di pivot considerate sono tutte degeneri, entrambi i dizionari hanno lo stesso valore \bar{z} . Per quanto riguarda la soluzione di base, il vettore b^k non sarà necessariamente identico a b^0 , ma conterrà una permutazione degli elementi di b^0 .

Associando un coefficiente di costo ridotto nullo alle variabili di base, ed indicando con \hat{c}^0 e \hat{c}^k rispettivamente i vettori dei coefficienti di costo ridotto associati alle basi B_0 e B_k , si può scrivere

$$z = \bar{z} + \sum_{j=0}^n \hat{c}_j^0 x_j = \bar{z} + \sum_{j=0}^n \hat{c}_j^k x_j.$$

Si immagini ora di incrementare la variabile x_s nel dizionario associato a B_0 , mantenendo fissate a zero tutte le altre variabili fuori base:

$$\begin{aligned}x_{N_0} &= e_s \delta \\x_{B_0} &= b^0 + d\delta \\z &= \bar{z} + \hat{c}_s^0 \delta\end{aligned}$$

dove con d si è indicato il vettore $-A_{B_0}^{-1} A_s$. L'obiettivo può anche essere scritto in funzione dei coefficienti di costo ridotto associati alla base B_k :

$$\begin{aligned}z &= \bar{z} + \sum_{j=0}^n \hat{c}_j^k x_j \\&= \bar{z} + \hat{c}_s^k \delta + \hat{c}_{B_0}^{k^T} (b^0 + d\delta).\end{aligned}$$

6. IL METODO DEL SIMPLEXSO

Uguagliando le due espressioni del costo si ottiene

$$\bar{z} + \hat{c}_s^0 \delta = \bar{z} + \hat{c}_s^k \delta + \hat{c}_{B_0}^{k^T} (b^0 + d\delta)$$

da cui

$$\hat{c}_s^0 \delta = \hat{c}_s^k \delta + \hat{c}_{B_0}^{k^T} (b^0 + d\delta)$$

ovvero, raccogliendo δ a fattore comune,

$$(\hat{c}_s^0 - \hat{c}_s^k - \hat{c}_{B_0}^{k^T} d) \delta = \hat{c}_{B_0}^{k^T} b^0.$$

Questa relazione deve essere valida per qualunque valore di δ , e, pertanto, deve risultare

$$\begin{aligned}\hat{c}_s^0 - \hat{c}_s^k - \hat{c}_{B_0}^{k^T} d &= 0 \\ \hat{c}_{B_0}^{k^T} b^0 &= 0.\end{aligned}$$

A questo punto viene effettivamente utilizzata la regola di Bland: poichè la variabile di indice s entra in base, certamente $\hat{c}_s^0 < 0$. La variabile di indice s , al pari di quella di indice t , compare in qualcuno dei dizionari del ciclo, ma non in tutti. Si è ipotizzato che t sia l'indice massimo tra quelli che corrispondono a variabili presenti in alcuni, ma non tutti, i dizionari (si veda la 6.10). Quindi $s < t$ e sicuramente $\hat{c}_s^k \geq 0$, perchè in caso contrario, per la regola di Bland, nel dizionario k sarebbe stata scelta la variabile x_s come variabile entrante. Da queste considerazioni scende che

$$\hat{c}_{B_0}^{k^T} d < 0,$$

cioè

$$\sum_{i \in B_0} \hat{c}_i^k d_i < 0,$$

da cui segue che deve esistere almeno un indice $r \in B_0$ tale che $\hat{c}_r^k d_r < 0$.

6.1. Formulazione matriciale

Questo può succedere solo se $\hat{c}_r^k \neq 0$; ricordando che $r \in B_0$, si osserva che anche la variabile x_r appartiene a qualcuna, ma non a tutte, le basi del ciclo ed, essendo t il massimo indice delle variabili di questo tipo, si avrà $r < t$ oppure $r = t$. Si vede facilmente che in realtà non può essere $r = t$. Infatti, poichè la variabile di indice t è la variabile entrante nel dizionario associato a B_k , $\hat{c}_t^k < 0$ e $d_t < 0$ poichè la variabile di indice t esce di base nel dizionario associato a B_0 . Si ha quindi $\hat{c}_t^k d_t > 0$ e, pertanto, non può essere $r = t$, essendo $\hat{c}_r^k d_r < 0$.

Quindi $r < t$; deve essere allora $\hat{c}_r^k \geq 0$, altrimenti, per la regola di Bland, al passo k entrerebbe in base la variabile di indice r e non quella di indice t , come si è ipotizzato. Di conseguenza si avrà $d_r < 0$.

Ad ogni dizionario del ciclo, essendo ogni pivot degenere, nessuna variabile cambia valore e, pertanto, tutte le variabili che a volte sono in base e a volte no, avranno valore 0; ne segue che in particolare $b_r^0 = 0$. Questo è assurdo; infatti la scelta della variabile uscente di base, quando il dizionario è quello associato a B_0 , avviene mediante la regola usuale (6.9):

$$i^* \in \arg \min_{i \in B_0: d_i < 0} \left\{ -\frac{b_i^0}{d_i} \right\}$$

e, nel caso in esame, si avrebbe, per $i = r$, $d_r < 0$ e $b_r^0 = 0$; il pivot sarebbe dunque degenere (e questo è possibile), ma poichè $r < t$, dalla regola di Bland si sarebbe scelta la variabile r piuttosto che, come ipotizzato, la variabile t , da cui l'assurdo. \square

La dimostrazione della correttezza della regola di Bland è piuttosto laboriosa, ma non priva di eleganza. La regola stessa è estremamente semplice, efficace, anche se, a volte, non molto efficiente; esistono altri metodi per evitare i cicli nel metodo del simplex, ma generalmente non hanno la semplicità della regola di Bland. Una di queste regole, usata fino ad alcuni anni fa, nasce dall'idea che la

6. IL METODO DEL SIMPLESSO

degenerazione può essere eliminata spostando leggermente i vincoli in modo da impedire che più di n iperpiani si intersechino nello stesso punto.

Grazie alla correttezza delle regole anti-ciclaggio, come la regola di Bland, si possono dimostrare immediatamente due importanti proprietà dei problemi di PL:

Teorema 19. *Per ogni problema di PL in forma standard che ammetta una soluzione di base ottimale, esiste sempre almeno una soluzione di base ottimale alla quale sono associati coefficienti di costo ridotto non negativi.*

Dimostrazione. Se così non fosse, eseguendo il metodo del simplesso a partire da una base ottimale, ad ogni passo di pivot si troverebbe sempre almeno un coefficiente di costo ridotto negativo e l'algoritmo prima o poi entrerebbe in un ciclo. \square

Teorema 20. *Per ogni problema di PL vale una ed una sola delle seguenti:*

1. *il problema non è ammissibile;*
2. *il problema ammette ottimo finito;*
3. *il problema è illimitato.*

Dimostrazione. Se il problema è ammissibile, è possibile eseguire l'algoritmo del simplesso equipaggiato con la regola di Bland. Data la finitezza delle basi e l'impossibilità di cicli, in un numero finito di passi di pivot si arriva ad una soluzione di base ottimale con tutti i coefficienti di costo ridotto non negativi, oppure si arriva all'individuazione di una direzione di decrescita illimitata dell'obiettivo. \square

6.2. Presentazione dell'algoritmo del simplesso mediante dizionario

6.2 Presentazione dell'algoritmo del simplesso mediante dizionario

In questa parte verrà visto come il metodo del simplesso già presentato in forma matriciale, possa essere facilmente eseguito a partire dalla rappresentazione esplicita del “dizionario”, eseguendo operazioni molto simili a quelle comunemente utilizzate per la risoluzione di sistemi di equazioni lineari. Lo scopo principale di questa parte è quello di fornire uno strumento essenzialmente didattico per l'esecuzione passo passo del metodo: solo risolvendo manualmente diversi problemi si riesce ad acquisire la consapevolezza delle caratteristiche del metodo.

Ricordiamo innanzitutto lo sviluppo del metodo così come esemplificato nel precedente paragrafo. Si ipotizza di avere un dizionario iniziale nella forma:

$$\begin{aligned} z &= \bar{z} + \sum_{j=m+1}^n \hat{c}_j x_j \\ x_1 &= \bar{x}_1 + \sum_{j=m+1}^n a_{1j} x_j \\ x_2 &= \bar{x}_2 + \sum_{j=m+1}^n a_{2j} x_j \\ &\vdots \\ x_m &= \bar{x}_m + \sum_{j=m+1}^n a_{mj} x_j \end{aligned}$$

dove $\bar{x}_1, \dots, \bar{x}_m$ sono costanti *non negative* – l'ipotesi di non negatività è legata alla necessità di avere una base iniziale ammissibile.

Si tratta quindi di un problema con n incognite, m vincoli di egualanza, il vincolo di non negatività (sottointeso per tutte le incognite), un obiettivo z da minimizzare. I coefficienti \hat{c}_j nell'e-

6. IL METODO DEL SIMPLEXSO

quazione dell'obiettivo sono i coefficienti di costo ridotto; la base è formata dalle incognite x_1, \dots, x_m , mentre fuori base abbiamo le incognite x_{m+1}, \dots, x_n .

Se tutti i coefficienti di costo ridotto sono non negativi, allora questa base è ottimale e la soluzione ottimale del problema è data da

$$x_1 = \bar{x}_1, x_2 = \bar{x}_2, \dots, x_m = \bar{x}_m, x_{m+1}, \dots, x_n = 0$$

con costo pari a \bar{z} .

Altrimenti certamente esisterà almeno un coefficiente di costo ridotto strettamente minore di 0. Sia \bar{j} l'indice di un simile coefficiente: $\hat{c}_{\bar{j}} < 0$. L'incognita (fuori base) $x_{\bar{j}}$ è "competitiva" rispetto a quelle attualmente in base: un suo eventuale ingresso in base, ad un livello strettamente positivo, porterebbe ad un miglioramento del costo; in particolare, per ogni unità di $x_{\bar{j}}$ in base si avrebbe una variazione pari a $\hat{c}_{\bar{j}}$ del costo. Per capire se è possibile fare entrare in base tale variabile e, in caso affermativo, poter costruire il nuovo dizionario, occorre considerare i vincoli ed analizzare l'effetto che l'eventuale ingresso di tale variabile in base avrebbe sul segno delle altre incognite.

Consideriamo una generica equazione del dizionario, evidenziando il termine contenente la variabile $x_{\bar{j}}$:

$$x_k = \bar{x}_k + a_{k\bar{j}}x_{\bar{j}} + \dots$$

(non sono stati scritti esplicitamente gli altri addendi in quanto, nel corso delle operazioni di pivot, le variabili non di base con indice diverso da \bar{j} restano fissate a 0).

Si vede che se $a_{k\bar{j}} \geq 0$, all'aumentare di $x_{\bar{j}}$ anche x_k non diminuirà, mantenendo quindi la positività – si ricordi che si è supposto che $\bar{x}_k \geq 0$. Se al contrario il segno del coefficiente della variabile entrante in base fosse negativo, allora per non violare il vincolo sul segno di x_k si dovrebbe scegliere

$$x_{\bar{j}} \leq -\frac{\bar{x}_k}{a_{k\bar{j}}}$$

6.2. Presentazione dell'algoritmo del simplesso mediante dizionario

(che è una quantità non negativa). Analizzando la situazione di *tutte* le equazioni del dizionario, si dovrà scegliere $x_{\bar{j}}$ in modo tale da rispettare la più restrittiva di tutte le condizioni, cioè

$$x_{\bar{j}} \leq -\frac{\bar{x}_k}{a_{k\bar{j}}} \quad \forall k : a_{k\bar{j}} < 0$$

da cui si ottiene

$$x_{\bar{j}} \leq \min_k \left\{ -\frac{\bar{x}_k}{a_{k\bar{j}}} : a_{k\bar{j}} < 0 \right\}.$$

Per quanto già visto, poiché tanto maggiore è il valore della variabile entrante in base, tanto maggiore sarà la diminuzione del costo, si sceglierà

$$x_{\bar{j}} = \min_k \left\{ -\frac{\bar{x}_k}{a_{k\bar{j}}} : a_{k\bar{j}} < 0 \right\}.$$

Questa scelta provoca l'azzeramento di almeno una tra le variabili di base. Sia \bar{k} l'indice di tale variabile: sarà questa la variabile uscente di base. L'operazione di pivot può essere ora effettuata "risolvendo" l'equazione che, nel dizionario, corrisponde alla definizione di $x_{\bar{k}}$, rispetto a $x_{\bar{j}}$ e sostituendo ogni occorrenza di $x_{\bar{j}}$ nelle altre equazioni e nell'espressione dell'obiettivo.

Esempio 12. Si consideri il seguente problema :

$$\begin{aligned} \min & x - 4y \\ \text{s.t. } & x + y \leq 1 \\ & x - 2y \leq 0 \\ & -x + y \leq \frac{1}{2} \\ & x, y \geq 0 \end{aligned}$$

6. IL METODO DEL SIMPLESSO

Inserendo le opportune variabili di slack ed introducendo una variabile ausiliaria z per rappresentare la funzione obiettivo, il problema precedente può essere scritto come

$$\begin{aligned} \min z &= x - 4y \\ s &= 1 - x - y \\ t &= -x + 2y \\ u &= \frac{1}{2} + x - y \\ x, y, s, t, u &\geq 0 \end{aligned}$$

Il problema originale è rappresentato in funzione della base $B = \{s, t, u\}$; le variabili x ed y sono fuori base. Il sistema è equivalente all'originale, quindi assegnando ad x ed y valori arbitrari, si ottengono tutte le soluzioni del sistema di *equazioni* appena scritto. Assegnando valore 0 a tutte le variabili fuori base si ottiene la soluzione di base

$$x = 0, y = 0, s = 1, t = 0, u = \frac{1}{2};$$

il costo associato a questa soluzione di base è $z = 0$.

In corrispondenza della soluzione di base corrente sia x che y valgono 0. Se si provasse ad incrementare la sola variabile x non si otterrebbe un miglioramento dell'obiettivo: infatti, essendo il costo $z = x - 4y$, mantenendo $y = 0$ e rendendo positivo, se possibile, x , si otterrebbe un *aumento* di z , l'esatto opposto di ciò che desiderremmo. Questo non significa che nessuna soluzione con x positivo sia conveniente – significa soltanto che non è conveniente incrementare x in questo momento. Al contrario, un incremento di y , lasciando x a 0, porterebbe ad una diminuzione del costo – una diminuzione di 4 unità per ogni unità di incremento di y . Di quanto è possibile incrementare y , lasciando invariato $x = 0$? Per quanto riguarda il sistema di equazioni, non esistono vincoli ai valori as-

6.2. Presentazione dell'algoritmo del simplesso mediante dizionario

sumibili da y ; tuttavia occorre rispettare i vincoli di non negatività per tutte le variabili. Esaminandoli singolarmente, si osserva che:

- x resta fissato a 0;
- y è scelto non negativo per costruzione;
- s resta non negativo se e solo se $1 - y \geq 0$, cioè se $y \leq 1$;
- la non negatività di t è assicurata per qualsiasi valore di $y \geq 0$;
- quella di u solo per $y \leq \frac{1}{2}$.

Pertanto, mentre per soddisfare il sistema di equazioni avremmo potuto scegliere y liberamente, per tener conto dei vincoli sul segno siamo costretti a limitare la scelta ai valori $0 \leq y \leq \frac{1}{2}$. Poiché maggiore è il valore di y , minore sarà il valore dell'obiettivo, la scelta più conveniente pare essere $y = \frac{1}{2}$.

Per costruzione, solo le variabili di base possono assumere valore non nullo; pertanto per ottenere una soluzione di base con $y = \frac{1}{2}$ dovremo scrivere un dizionario nel quale y compaia fra le variabili di base; ma il numero di variabili nella base è fissato e pari ad m , pertanto una variabile che attualmente fa parte della base dovrà lasciare il posto ad y . Poiché si è scelto di dare ad y il massimo valore possibile, compatibilmente con i vincoli sul segno, tale scelta avrà l'effetto di rendere nulla una (in genere, *almeno* una) delle attuali variabili di base. Nel caso in esame, sarà u ad annullarsi quando $y = \frac{1}{2}$. Questa variabile è la candidata naturale per lasciare il posto in base ad y . Riscriviamo quindi il dizionario esplicitando le equazioni rispetto alle incognite s, t, y : dalla terza equazione ricaviamo $y = \frac{1}{2} + x - u$ che, sostituita nelle prime due e nell'obiettivo, porta

6. IL METODO DEL SIMPLEXSO

al dizionario seguente:

$$z = -2 - 3x + 4u$$

$$s = \frac{1}{2} - 2x + u$$

$$t = 1 + x - 2u$$

$$y = \frac{1}{2} + x - u.$$

Si noti che in questo dizionario si sono tralasciate sia l'indicazione della direzione dell'ottimizzazione (min) che i vincoli di non negatività su tutte le variabili: adottando la convenzione di associare un dizionario solo a problemi di programmazione lineare standard, sarà sempre sottinteso che i problemi saranno di minimo e con tutte le variabili non negative.

Il dizionario appena ricavato è, naturalmente, una rappresentazione esattamente equivalente alla precedente, ottenuta con operazioni standard di trasformazione di sistemi lineari (si tratta del metodo di eliminazione di Gauss). Tuttavia questa rappresentazione permette una lettura più agevole ed indica la strada verso la soluzione ottimale. In particolare, questo dizionario è associato ad una nuova soluzione di base, facilmente leggibile ponendo uguali a zero le variabili fuori base, corrispondente ai valori

$$x = 0, y = \frac{1}{2}, s = \frac{1}{2}, t = 1, u = 0$$

e caratterizzata da un costo $z = -2$ (leggibile anche direttamente dalla prima equazione, ponendo $x = u = 0$).

Il coefficiente di costo ridotto di x è pari a -3 : ciò significa che per ogni unità di incremento assegnato ad x il costo diminuirà di 3 unità. Il coefficiente di costo ridotto di u è invece positivo, pari a 4: un eventuale innalzamento del livello di u provocherebbe un aumento del costo. Si noti che la positività del coefficiente di costo ridotto della variabile appena uscita di base non è casuale,

6.2. Presentazione dell'algoritmo del simplesso mediante dizionario

ma è conseguenza del fatto che un eventuale ingresso in base di u riporterebbe il dizionario ad assumere la forma precedente, con conseguente aumento del costo.

Ripetendo il ragionamento svolto in precedenza, analizziamo l'effetto che un innalzamento del livello di x potrebbe avere sui vincoli di segno: dalla prima equazione otteniamo che la condizione di non negatività di s è $x \leq \frac{1}{4}$; dalla seconda non abbiamo vincoli (un incremento di x non può rendere negativa t) e così pure dalla terza. Quindi la scelta sarà quella di far entrare x in base a livello $\frac{1}{4}$, facendo uscire nel frattempo la variabile s (divenuta pari a 0).

Ricavando x dalla prima equazione e sostituendola nelle rimanenti, si ottiene

$$\begin{aligned} z &= -\frac{11}{4} + \frac{3}{2}s + \frac{5}{2}u \\ x &= \frac{1}{4} - \frac{1}{2}s + \frac{1}{2}u \\ t &= \frac{5}{4} - \frac{1}{2}s - \frac{3}{2}u \\ y &= \frac{3}{4} - \frac{1}{2}s - \frac{1}{2}u. \end{aligned}$$

Questo dizionario corrisponde ad una base costituita da x, y, t e rappresenta una soluzione di base con

$$x = \frac{1}{4}, y = \frac{3}{4}, s = 0, t = \frac{5}{4}, u = 0$$

e costo pari a $-11/4$. I coefficienti di costo ridotto di entrambe le variabili non di base sono positivi (basterebbe in realtà che fossero non negativi). Poiché le soluzioni ammissibili sono un sottoinsieme delle soluzioni ottenute assegnando valore arbitrario, ma non negativo, alle due variabili fuori base (t, y), dall'espressione $z = -\frac{11}{4} + \frac{3}{2}s + \frac{5}{2}u$ si deduce che nessuna soluzione ammissibile può avere obiettivo inferiore a $-11/4$. Pertanto la soluzione di base corrente è *ottimale*.

6.3 Inizializzazione del metodo del simplesso

Il metodo due fasi

Abbiamo visto finora alcune proprietà dei problemi di *PL* ammissibili e, in particolare, dei problemi di *PL* per i quali sia nota una base iniziale ammissibile. In questo paragrafo si vedrà come verificare se un problema di *PL* è ammissibile o no ed, in caso affermativo, come trovare, se esiste, una soluzione ammissibile di base. Infatti, non è detto che un problema ammissibile ammetta soluzioni ammissibili di base – il sistema di equazioni potrebbe non essere ridotto al rango.

Tra i vari metodi esistenti per la ricerca di una soluzione ammissibile di base, viene qui presentato il più noto e diffuso, il cosiddetto metodo due fasi. Il nome deriva dal fatto che alla fase di ottimizzazione vera e propria viene fatta precedere una prima fase il cui scopo è unicamente determinare, se esiste, una base ammissibile iniziale. L'idea di questa prima fase nasce dall'osservazione che in alcuni problemi di *PL* la determinazione di una base iniziale ammissibile è un problema banale. Si consideri infatti il problema:

$$\begin{aligned} \min & c^T x \\ Ax &\leq b \\ x &\geq 0. \end{aligned}$$

Se il termine noto b è non negativo, allora aggiungendo le variabili di slack si ottiene immediatamente una base ammissibile:

$$\begin{aligned} \min & [c^T \quad 0^T] \begin{bmatrix} x \\ s \end{bmatrix} \\ [A \quad I] \begin{bmatrix} x \\ s \end{bmatrix} &= b \\ x, s &\geq 0, \end{aligned}$$

6.3. Inizializzazione del metodo del simplesso

costituita dalla matrice identica. Quindi in un problema con variabili di slack e termine noto non negativo, le variabili di slack costituiscono una base ammissibile. In un problema generico, che supponiamo essere in forma standard, possiamo sempre ricondurci al caso in cui il termine noto sia non negativo (al massimo si tratterà di moltiplicare per -1 entrambi i membri di alcune equazioni). La prima fase del metodo due fasi consiste nella risoluzione del seguente problema, detto *problema prima fase*:

$$\begin{aligned} \min \quad & \mathbf{1}^T x^a \\ \text{s.t.} \quad & Ax + x^a = b \\ & x, x^a \geq 0, \end{aligned}$$

con $b \geq 0$.

Come si vede il problema prima fase ha alcune variabili, x^a dette *variabili artificiali*, che appaiono come variabili di slack. In realtà esse sono diverse dalle variabili di slack, in quanto queste ultime sono vere e proprie variabili, introdotte per scrivere in forma standard un problema di *PL*, mentre quelle artificiali sono variabili che provocano una perturbazione del problema originale. L'obiettivo del problema prima fase è la minimizzazione della somma di tali perturbazioni.

Si vede immediatamente che il problema prima fase non è né inammisibile, né può essere illimitato. Infatti una soluzione ammissibile (a patto che ci si sia ricondotti al caso $b \geq 0$) esiste sempre, ed è data da $x = 0, x^a = b$. Anzi, questa è una soluzione ammissibile *di base*; inoltre, poiché il costo è una combinazione non negativa di variabili non negative, il problema non potrà essere illimitato inferiormente. Quindi, per quanto già visto, il problema prima fase ammette sempre soluzione ottimale. In base al valore dell'ottimo possiamo dedurre se il problema originale ammette soluzioni ammissibili. Sia z^* il valore dell'ottimo del problema prima fase e siano x^* e x^{a*} le componenti di una soluzione ottimale. Vale il seguente

6. IL METODO DEL SIMPLESSO

Teorema 21. *Esiste almeno una soluzione ammissibile per il sistema $Ax = b, x \geq 0$ se e solo se $z^* = 0$.*

Dimostrazione. Se l'ottimo della prima fase è 0 allora, essendo l'obiettivo della prima fase pari alla somma delle variabili artificiali, e, poiché queste ultime sono non negative, esse dovranno essere tutte nulle. Quindi si avrà

$$\begin{aligned} Ax^* + 0 &= b \\ x^* &\geq 0 \end{aligned}$$

e pertanto x^* sarà ammissibile per il problema originario.

Se invece $z^* > 0$, non è possibile che esista una soluzione ammissibile per il problema originale. Infatti in questo caso sicuramente ogni soluzione ammissibile del problema prima fase avrà almeno una variabile artificiale non nulla e, pertanto, il sistema

$$\begin{aligned} Ax + 0 &= b \\ x &\geq 0 \end{aligned}$$

risulta impossibile.

□

Il valore ottimo del problema prima fase ci permette dunque di distinguere tra problemi ammissibili e non e, nel caso di problemi ammissibili, fornisce una soluzione particolare ammissibile. Tuttavia non è detto che tale soluzione sia di base. Il teorema fondamentale garantisce che, nel caso di problemi nei quali il sistema di equazioni lineari sia ridotto al rango, se esiste una soluzione ammissibile ne esiste una ammissibile di base. La prima fase del metodo del simplesso ci permette anche di determinare una soluzione di base ammissibile. Infatti si indichi con B l'insieme degli indici di una soluzione ottimale di base del problema prima fase, nel caso in cui la prima fase sia terminata con costo pari a 0. Si possono dare i seguenti casi:

6.3. Inizializzazione del metodo del simplesso

1. nessuna variabile artificiale x^a appartiene alla base B ; in questo caso B è una base ammissibile per il problema originario; il metodo prosegue eliminando dal dizionario tutte le variabili artificiali (fuori base) e riscrivendo la funzione obiettivo $c^T x$ in funzione delle variabili, non artificiali, fuori base – da qui può iniziare il metodo del simplesso (seconda fase) per la risoluzione del problema originale;
2. almeno una variabile artificiale compare nella base ottimale B ; sicuramente tale variabile sarà nulla, in quanto, essendo l'obiettivo della prima fase pari alla somma delle variabili artificiali, nessuna di queste potrà essere diversa da zero se l'ottimo è zero. Se è possibile, occorre effettuare un'operazione di pivot per far uscire di base tale variabile. Sia x_i^a una variabile artificiale in base a livello 0. Si consideri una variabile non artificiale x_j fuori base. Sia

$$A' = [A \quad I]$$

la matrice dei coefficienti del problema prima fase. Tramite le operazioni di pivot effettuate per evidenziare la base corrente, B , il sistema originale

$$[A \quad I] \begin{bmatrix} x \\ x^a \end{bmatrix} = b$$

è stato trasformato nel sistema equivalente

$$A_B'^{-1} Ax + A_B'^{-1} x^a = A_B'^{-1} b$$

Se in questo dizionario l' i -esima equazione

$$e_i^T A_B'^{-1} x^a = e_i^T A_B'^{-1} b - e_i^T A_B'^{-1} Ax$$

corrisponde ad una variabile artificiale in base a livello zero, come si è visto dal teorema 15, la condizione necessaria e sufficiente per poter effettuare lo scambio tra una colonna A_j di

6. IL METODO DEL SIMPLEXSO

coefficienti di x_j e quella dei coefficienti dell' i -esima variabile in base è che l' i -esima componente del vettore $d = -A_B'^{-1} A_j$ sia non nulla. Quindi sarà possibile portare fuori base la variabile artificiale associata all' i -esima equazione se e solo se esiste una variabile fuori base x_j tale che

$$d_i = -e_i^T A_B'^{-1} A_j \neq 0$$

Poiché la variabile artificiale x_i^a è a livello 0, tale operazione cambia solo la composizione della base, non la soluzione di base (si effettua un passo di lunghezza 0). Pertanto in questo caso è possibile effettuare un'operazione di pivot sia se $d_i < 0$ che se $d_i > 0$. Si osservi che, in generale un'operazione di pivot che faccia entrare in base una variabile cui è associato un coefficiente $d_i > 0$, nel metodo del simplexso non è ammessa, poiché il risultato sarebbe una perdita di ammissibilità. Anche nel metodo del simplexso però, se la variabile entrante in base ha valore 0, l'operazione è in teoria possibile, dato che un passo di lunghezza 0 non modifica la soluzione corrente. Tuttavia nel metodo del simplexso, sebbene possibile, questa operazione non ha interesse alcuno e, pertanto, non viene mai effettuata. Nella prima fase del metodo due fasi invece tale operazione di pivot può servire a far uscire di base la variabile artificiale x_i^a .

3. l'unico caso in cui l'operazione di pivot non è possibile è quello in cui

$$e_i^T A_B'^{-1} A_j = 0 \quad \forall j \in \{1, \dots, n\}$$

Da qui segue che

$$(e_i^T A_B'^{-1}) A = 0^T$$

6.3. Inizializzazione del metodo del simplex

e, pertanto, il vettore

$$e_i^T A_B'^{-1}$$

costituisce un vettore di coefficienti di una combinazione lineare delle righe di A che genera il vettore nullo. Tale vettore non è banale. Infatti, se fosse

$$e_i^T A_B'^{-1} = 0^T$$

moltiplicando ambo i membri, a destra, per la matrice invertibile A_B' , si otterrebbe

$$\begin{aligned} e_i^T A_B'^{-1} A_B' &= 0^T A_B' && \text{da cui} \\ e_i^T &= 0^T \end{aligned}$$

che è falso. Il vettore $e_i^T A_B'^{-1}$ fornisce quindi una combinazione lineare nulla non banale delle righe di A . La matrice A quindi in questo caso non ha rango massimo. Poiché il sistema $Ax = b$ era ammissibile, significa che un'equazione, in questo caso quella di indice i nel dizionario corrente, è ridondante e può essere eliminata dal dizionario finale del problema, al termine della prima fase.

Quindi si potranno fare uscire di base, attraverso operazioni di pivot, tutte e sole le variabili artificiali presenti in base e associate ad equazioni del sistema $Ax = b$ fra loro linearmente indipendenti.

Una volta terminata la prima fase del metodo due fasi, nel caso in cui si sia determinata una base ammissibile, si può iniziare con il normale metodo del simplex e determinare, se esiste, una soluzione ottimale del problema originario.

6. IL METODO DEL SIMPLEXSO

Esempio 13. Si consideri il problema di *PL* seguente

$$\begin{aligned} \min & [1 \ -1 \ 0 \ 1 \ 2] x \\ \left[\begin{array}{ccccc} 1 & -1 & -1 & 1 & 2 \\ -1 & -1 & 0 & 2 & 1 \\ 2 & 0 & -1 & -1 & 1 \end{array} \right] x &= \left[\begin{array}{c} 2 \\ 1 \\ 2 \end{array} \right] \\ x &\geq 0. \end{aligned}$$

Poiché non è evidente una base iniziale si imposta il problema prima fase:

$$\begin{aligned} \min & [1 \ 1 \ 1] x^a \\ \left[\begin{array}{ccccc} 1 & -1 & -1 & 1 & 2 \\ -1 & -1 & 0 & 2 & 1 \\ 2 & 0 & -1 & -1 & 1 \end{array} \right] x + x^a &= \left[\begin{array}{c} 2 \\ 1 \\ 2 \end{array} \right] \\ x, x^a &\geq 0. \end{aligned}$$

Utilizzando il dizionario per questo problema, si ha:

$$\begin{aligned} z &= 5 - 2x_1 + 2x_2 + 2x_3 - 2x_4 - 4x_5 \\ x_1^a &= 2 - x_1 + x_2 + x_3 - x_4 - 2x_5 \\ x_2^a &= 1 + x_1 + x_2 - 2x_4 - x_5 \\ x_3^a &= 2 - 2x_1 + x_3 + x_4 - x_5 \end{aligned}$$

Facendo entrare in base la variabile x_5 (regola del coefficiente di costo ridotto minimo), si ottiene

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

6.3. Inizializzazione del metodo del simplesso

In questo caso si può concludere che il problema dato è privo di soluzioni ammissibili, in quanto la prima fase è terminata con valore strettamente positivo. Se il termine noto del problema iniziale fosse

stato $\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$, si sarebbe invece ottenuto il dizionario seguente:

$$\begin{aligned} z &= 0 + 2x_1^a + 2x_4 \\ x_5 &= 1 - \frac{1}{2}x_1^a - \frac{1}{2}x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_3 - \frac{1}{2}x_4 \\ x_2^a &= \frac{3}{2}x_1 + \frac{1}{2}x_2 - \frac{1}{2}x_3 - \frac{3}{2}x_4 + \frac{1}{2}x_1^a \\ x_3^a &= 0 - \frac{3}{2}x_1 - \frac{1}{2}x_2 + \frac{1}{2}x_3 + \frac{3}{2}x_4 + \frac{1}{2}x_1^a \end{aligned}$$

Da questo dizionario si deduce che il problema originario ammette soluzioni: ad esempio la scelta $x_5 = 1$, con tutte le altre variabili nulle, soddisfa il sistema dei vincoli, come è facile vedere per sostituzione diretta. Quindi è nota una soluzione ammissibile, non però una soluzione ammissibile di base. Per determinare una base occorre, se possibile, far uscire di base tutte le variabili artificiali, effettuando pivot degeneri. Ad esempio, è possibile far uscire di base la variabile artificiale x_2^a facendo un'operazione di pivot con una qualsiasi delle quattro variabili x presenti nell'equazione che definisce x_2^a , indipendentemente dal segno del coefficiente. Ad esempio, scambiando x_2^a con x_2 si ottiene il dizionario:

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

che ancora non rappresenta una soluzione di base; volendo far uscire di base la variabile x_3^a , si scopre che non è possibile, in quanto

6. IL METODO DEL SIMPLEXSO

nell'equazione che definisce tale variabile non compare nessuna incognita del problema originario. Ricordando che il problema iniziale non conteneva le variabili artificiali, si può anche considerare che la terza equazione sia stata trasformata nell'equivalente equazione ridondante $0^T x = 0$, che può essere eliminata dal problema. Il problema dato può pertanto essere risolto partendo dal dizionario seguente (seconda fase), nel quale è stata inserita l'equazione del costo, sono state eliminate le variabili artificiali della prima fase ed è stata eliminata la terza equazione, combinazione lineare delle prime due:

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

Per caso il dizionario iniziale è già ottimale. Si conclude pertanto che la soluzione $x_5 = 1$ è ottima per il problema dato.

Il metodo "big M"

Esistono altre tecniche per la determinazione di soluzioni di base iniziali ammissibili. Una delle più note è chiamata "*metodo della M grande*" ("big M") e consiste nel risolvere un problema di prima fase in cui non si trascura la funzione obiettivo originale:

$$\begin{aligned} \min c^T x + M \mathbf{1}^T x^a && (6.11) \\ Ax + x^a &= b \\ x, x^a &\geq 0 \end{aligned}$$

Si può dimostrare che, a patto di scegliere M sufficientemente grande, questo problema gode di proprietà molto simili a quelle del problema prima fase.

Il problema artificiale (6.11) è sempre ammissibile, se $b \geq 0$. Infatti la soluzione di base $x = 0, x^a = b$ è ammissibile. Quindi il pro-

6.3. Inizializzazione del metodo del simplesso

blema o è illimitato oppure ammette soluzione ottimale. Valgono i seguenti teoremi:

Teorema 22. *Se il problema*

$$\begin{aligned} \min c^T x \\ Ax = b \\ x \geq 0 \end{aligned} \tag{6.12}$$

con $b \geq 0$ è illimitato, allora, per ogni $M > 0$, è illimitato anche il problema (6.11).

Dimostrazione. Ad una successione $\{x_k\}$ ammissibile per (6.12) con $c^T x_k \rightarrow \infty$ si può associare la successione $\{x_k, 0\}$ ammissibile per il problema (6.11) e ugualmente con costo divergente a $-\infty$. \square

Teorema 23. *Se il problema (6.12) non è ammissibile, allora esiste $M_0 > 0$ tale che per ogni $M \geq M_0$, o il problema (6.11) è illimitato, oppure ogni suo ottimo (x^*, x^{a^*}) è tale che $x^{a^*} \neq 0$.*

Dimostrazione. Per assurdo, ipotizziamo che esista un $M > M_0$ tale che il problema (8.3) ammetta ottimo finito $(x^*, 0)$. Ma in questo caso x^* sarebbe ammissibile per (6.12). \square

Teorema 24. *Se il problema (6.12) ammette ottimo finito x^* , allora la soluzione $(x^*, 0)$ è ottimale per il problema (6.11)*

La dimostrazione di questa parte viene rimandata al capitolo sulla dualità 8.2 a pag. 178.

Il problema della scelta di una costante M sufficientemente elevata si risolve evitando di assegnare un valore *esplicito* ad M , ma considerandola, in modo simbolico, sempre maggiore di qualsiasi altra costante incontrata nel corso della risoluzione del problema.

6.4 Problemi con infiniti ottimi

Se, al termine del metodo del simplesso, esistono dei coefficienti di costo ridotto nulli, possono esistere infinite soluzioni ottimali; infatti incrementando, se possibile, la componente non di base associata ad un coefficiente di costo ridotto nullo, la funzione obiettivo non varierà. Se l'incremento δ^* è strettamente positivo e finito, si otterrà un nuovo vertice ottimale. Se si compie un'operazione di pivot per portare in base la variabile corrispondente al coefficiente di costo ridotto nullo, se esiste almeno una componente negativa nel corrispondente vettore direzione d , l'operazione di pivot può essere eseguita e si determina così una nuova base ottimale; se invece le componenti di d sono tutte positive o nulle allora non esisterà una corrispondente soluzione di base, ma esisteranno comunque infinite soluzioni ottimali. In altri termini, se x^* è una soluzione ottimale associata ad una base A_B ed esiste un coefficiente di costo ridotto, ad esempio \hat{c}_j , nullo, allora per ogni valore di $0 \leq \delta \leq \delta^*$ dove

$$\delta^* = \begin{cases} \min\{-\frac{x_i}{d_i} : d_i < 0\} & \text{se } \exists i \in B : d_i < 0 \\ +\infty & \text{altrimenti} \end{cases} \quad (6.13)$$

il vettore

$$x(\delta) = \begin{bmatrix} A_B^{-1}b \\ 0 \end{bmatrix} + \begin{bmatrix} d \\ e_j \end{bmatrix} \delta$$

è soluzione ottimale del problema di PL . Naturalmente se nella (6.13) risulta $\delta^* = 0$, in corrispondenza di ogni coefficiente di costo ridotto nullo, allora la soluzione ottimale sarà ancora unica, pur esistendo diverse basi ottimali. Un problema di PL (in forma standard e ridotto al rango), se ammette ottimo finito, può dunque avere una e una sola soluzione ottimale (e, in questo caso, sarà sicuramente una soluzione ottimale di base), oppure infinite soluzioni ottimali (almeno una delle quali di base). Un problema di PL con uno o infiniti ottimi può ammettere diverse basi ottimali; una stessa soluzione ottimale

6.4. Problemi con infiniti ottimi

può, nel caso degenere, corrispondere a diverse basi ottimali. Un problema di *PL* può avere infinite soluzioni ottimali, tra cui anche una sola potrebbe essere di base. Si noti che, in un caso degenere, un'operazione di pivot degenere non porta ad un cambiamento della soluzione ($\delta^* = 0$); in questo caso è possibile anche effettuare un'operazione di pivot corrispondente ad un elemento *positivo* di d senza perdere l'ammissibilità e l'ottimalità – si tratta in sostanza di un passo in una direzione non ammissibile, ma di lunghezza 0 e, pertanto, non si perde l'ammissibilità.

Per comprendere meglio la struttura dei problemi di *PL* e delle loro soluzioni ottimali, è utile tenere presente le seguenti proprietà:

Teorema 25. *L'insieme ammissibile di un problema di PL è convesso.*

Dimostrazione. Si ricordi che un insieme S è convesso se e solo se comunque scelti due suoi elementi $u \in S, v \in S$, ciascuna loro combinazione convessa appartiene ad S , cioè

$$\lambda u + (1 - \lambda)v \in S \quad \forall \lambda \in [0, 1].$$

Scegliendo due soluzioni u e v , ammissibili per un problema in cui $S = \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$ si ha che la combinazione convessa $\lambda u + (1 - \lambda)v$ è non negativa (essendo combinazione non negativa di due vettori non negativi) e inoltre

$$\begin{aligned} A(\lambda u + (1 - \lambda)v) &= \lambda Au + (1 - \lambda)Av \\ &= \lambda b + (1 - \lambda)b = b, \end{aligned}$$

pertanto $\lambda u + (1 - \lambda)v$ è soluzione ammissibile per ogni $\lambda \in [0, 1]$. \square

Teorema 26. *L'insieme delle soluzioni ottimali di un problema di PL che ammette ottimo finito è convesso.*

6. IL METODO DEL SIMPLEXSO

Dimostrazione. E' sufficiente dimostrare che il costo della combinazione convessa di due soluzioni ottimali è ancora ottimale. In effetti, se u e v sono soluzioni ottimali, di valore z , si ha

$$\begin{aligned} c^T(\lambda u + (1 - \lambda)v) &= \lambda c^T u + (1 - \lambda)c^T v \\ &= \lambda z + (1 - \lambda)z = z. \end{aligned}$$

□

Il seguente teorema fornisce una rappresentazione finita dei poliedri.

Teorema 27. (*Minkowski-Weil - (Schrijver, 1998)*) *Dato un qualsiasi poliedro*

$$S = \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$$

esiste un numero finito di elementi $v_1, v_2, \dots, v_k \in S$ ed un insieme finito di vettori $y_1, y_2, \dots, y_h \in \mathbb{R}^n$ tale che ogni elemento x di S può essere rappresentato come

$$x = \sum_{i=1}^k \lambda_i v_i + \sum_{j=1}^h \mu_j y_j$$

con

$$\begin{aligned} \mu_j &\geq 0 \quad \forall j \\ \sum_{i=1}^k \lambda_i &= 1, \quad \lambda_i \geq 0 \quad \forall i. \end{aligned}$$

Il teorema di Minkowski permette di generare ogni poliedro attraverso la somma di una combinazione convessa (cioè effettuata con coefficienti non negativi a somma unitaria) di un numero finito di suoi elementi (l'insieme minimale di tali elementi è costituito dai vertici del poliedro) e di una combinazione a coefficienti non negativi di un numero finito di vettori; in questo caso l'insieme minimale

6.4. Problemi con infiniti ottimi

è costituito dai cosiddetti raggi estremi del poliedro. Un *raggio* r in un poliedro S è un vettore di \mathbb{R}^n tale che, se $u \in S$, allora

$$u + \lambda r \in S \quad \forall \lambda \geq 0.$$

I raggi sono dunque direzioni ammissibili che permettono uno spostamento illimitato in un verso, senza mai uscire dall'insieme ammissibile, a partire da un qualunque punto ammissibile. I raggi estremi sono raggi che non è possibile rappresentare come combinazione convessa non banale (cioè a coefficienti non 0 né 1) di altri due raggi.

Esempio 14. Si consideri il problema

$$\begin{aligned} & \min [\begin{array}{cccc} 0 & 0 & 0 & 1 \end{array}] x \\ & \left[\begin{array}{cccc} 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & 2 \end{array} \right] x = \left[\begin{array}{c} 2 \\ 1 \end{array} \right] \\ & x \geq 0 \end{aligned}$$

che possiede la soluzione ottimale $x = [\begin{array}{cccc} 0 & 3 & 2 & 0 \end{array}]^T$, corrispondente alla base $A_{2,3}$. Per confermare che questa è la base ottimale basta calcolare i coefficienti di costo ridotto:

$$\begin{aligned} \hat{c}_{1,4} &= [\begin{array}{cc} 0 & 1 \end{array}] - [\begin{array}{cc} 0 & 0 \end{array}] A_{2,3}^{-1} A_{1,4} \\ &= [\begin{array}{cc} 0 & 1 \end{array}]. \end{aligned}$$

Il coefficiente di costo ridotto della variabile x_1 è nullo; è possibile pertanto che esistano infinite soluzioni ottimali. Per determinarle si cerca se possibile di fare entrare in base la variabile x_1 :

$$d = -A_{2,3}^{-1} A_1 = - \left[\begin{array}{cc} 0 & 1 \\ 1 & -1 \end{array} \right]^{-1} \left[\begin{array}{c} 1 \\ 0 \end{array} \right] = \left[\begin{array}{c} -1 \\ -1 \end{array} \right].$$

6. IL METODO DEL SIMPLEXSO

La variabile uscente di base viene determinata con la regola usuale:

$$\begin{aligned} \min \left\{ -\frac{\bar{x}_2}{d_2}, -\frac{\bar{x}_3}{d_3} \right\} &= \min \left\{ -\frac{3}{-1}, -\frac{2}{-1} \right\} \\ &= 2 = -\frac{\bar{x}_3}{d_3}. \end{aligned}$$

La variabile uscente di base è quindi la x_3 ; la nuova base diviene $B = \{1, 2\}$. La soluzione associata a questa base è

$$\bar{x}_{1,2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

ed i coefficienti di costo ridotto ad essa associati sono

$$\begin{aligned} \hat{c}_{3,4}^T &= [0 \ 1] - [0 \ 0] \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix} \\ &= [0 \ 1]. \end{aligned}$$

Questa verifica non è in realtà necessaria per dimostrare l'ottimalità della soluzione corrente, dato che avendo fatto entrare in base una variabile con coefficiente di costo ridotto nullo, sicuramente non si avrà un cambio della funzione obiettivo. Però dai coefficienti di costo ridotto di quest'ultima soluzione di base si vede che per trovare altre soluzioni di base ottimali si dovrebbe fare rientrare in base la variabile x_3 , che ne è appena uscita. Si può quindi concludere che le due soluzioni di base trovate sono le uniche soluzioni ottimali di base. Non esistono raggi nel poliedro delle soluzioni ottimali (altrimenti si sarebbero trovati in corrispondenza di un coefficiente di costo ridotto nullo). Quindi tutte e sole le soluzioni ottimali sono ottenibili da una combinazione convessa delle

6.4. Problemi con infiniti ottimi

due appena trovate:

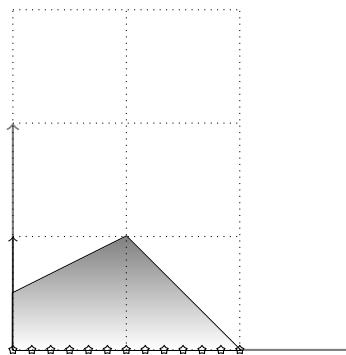
$$x^*(\lambda) = \lambda \begin{bmatrix} 0 \\ 3 \\ 2 \\ 0 \end{bmatrix} + (1 - \lambda) \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 - 2\lambda \\ 2\lambda + 1 \\ 2\lambda \\ 0 \end{bmatrix}$$

con $\lambda \in [0, 1]$.

Il problema risolto in questo esempio può essere anche rappresentato come

$$\begin{aligned} & \min_{x,y} y \\ & x + y \leq 2 \\ & -x + 2y \leq 1 \\ & x, y \geq 0 \end{aligned}$$

Questa formulazione permette una rappresentazione geometrica:



dove si è evidenziato l'insieme delle soluzioni ottimali.

6. IL METODO DEL SIMPLEXSO

Esempio 15. Se si considera il problema seguente, ottenuto modificando in parte il precedente,

$$\begin{aligned} \min & [0 \ 0 \ 0 \ 1] x \\ [1 \ 0 \ -1 \ 1] x &= [2] \\ 0 \ 1 \ -1 \ 2 & \\ x \geq 0 & \end{aligned}$$

si vede facilmente che la base $B = \{1, 2\}$ è, come nell'esempio precedente, ottimale. I coefficienti di costo ridotto associati a questa base sono ancora $\tilde{c}_{3,4}^T = [0 \ 1]$. Cercando di fare entrare in base la variabile x_3 , con coefficiente di costo ridotto nullo, si ottiene:

$$d = -A_{1,2}^{-1} A_3 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

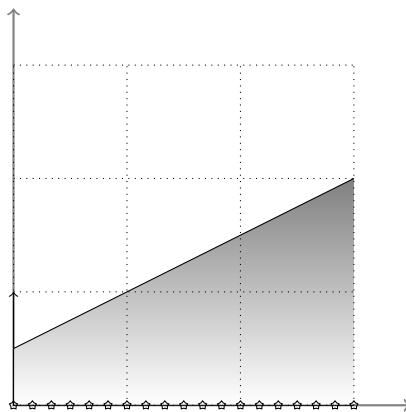
Lungo la direzione individuata dal vettore d per le variabili di base è possibile compiere uno spostamento illimitato senza perdere l'ammissibilità; lungo questa direzione inoltre il costo non cambia e, pertanto, si è determinato un raggio (estremo) ottimale:

$$\begin{aligned} x^*(\delta) &= \begin{bmatrix} \bar{x}_{1,2} \\ 0 \end{bmatrix} + \begin{bmatrix} d \\ e_1 \end{bmatrix} \delta \\ &= \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} \delta \\ &= \begin{bmatrix} 2 + \delta \\ 1 + \delta \\ \delta \\ 0 \end{bmatrix}. \end{aligned}$$

Le soluzioni $x^*(\delta)$ sono ottimali per ogni scelta di $\delta \geq 0$. Di queste infinite soluzioni, una sola, caratterizzata da $\delta = 0$, è di base.

6.4. Problemi con infiniti ottimi

Una rappresentazione geometrica di questo problema, ancora rispetto alle variabili x_3, x_4 , è la seguente:



Esempio 16. Il problema

$$\begin{aligned} \min & \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} x \\ & \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & 2 \end{bmatrix} x = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\ & x \geq 0 \end{aligned}$$

è ancora una variante dei precedenti, e possiede la soluzione ottimale $\bar{x} = [0 \ 0 \ 1 \ 0]^T$, degenera, ancora associata alla base $A_{2,3}$. La conferma è banale, in quanto la soluzione è ammissibile (come è facile controllare sostituendo il vettore \bar{x} nel sistema di equazioni) e i coefficienti di costo ridotto, essendo indipendenti dal vettore dei termini noti, sono ancora pari a

$$\hat{c}_{1,4}^T = [0 \ 1].$$

6. IL METODO DEL SIMPLEXSO

Cercando di fare entrare in base la variabile di costo ridotto nullo x_1 , si ottiene

$$d = -A_{2,3}^{-1} A_1 = \begin{bmatrix} -1 \\ -1 \end{bmatrix};$$

il passo di lunghezza massima possibile in questa direzione è dato da

$$\delta^* = \min \left\{ -\frac{0}{-1}, -\frac{1}{-1} \right\} = 0.$$

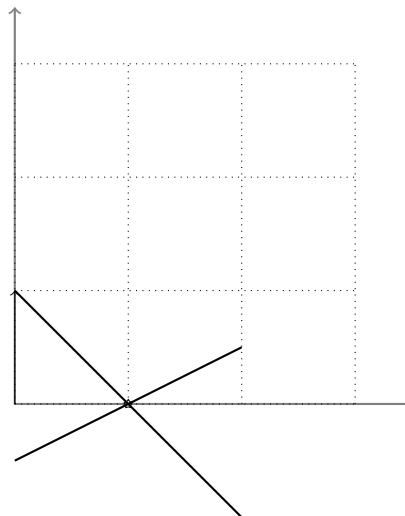
Si può dire, in questo caso, che le infinite soluzioni ottimali sono degenerate in un solo punto, la soluzione di base corrente. Si osservi che, in questo caso, la soluzione di base è degenere: se in una soluzione ottimale di base non degenere esiste almeno un coefficiente di costo ridotto nullo, allora esistono *sempre* infiniti ottimi. Naturalmente, in questo caso, è possibile fare entrare in base la variabile x_1 al posto della variabile x_2 : la nuova base $A_{1,3}$ è ancora ottimale e corrisponde alla soluzione di base

$$\bar{x}_{1,3} = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

che rappresenta sempre la stessa soluzione, $\bar{x}_3 = 1$. Esistono per tanto più basi ottimali, ma una sola soluzione ottimale (si può verificare che le basi ottimali in questo esempio sono 3).

Rappresentando il problema nello spazio x_3, x_4 , si ottiene la seguente figura:

6.5. Efficienza del metodo del simplex



dove l'unica soluzione ammissibile è il punto $(1, 0)$.

6.5 Efficienza del metodo del simplex

Esiste un'importante teoria, detta *teoria della complessità computazionale* (Garey & Johnson, 1979; Papadimitriou & Steiglitz, 1998), i cui scopi sono l'analisi e la classificazione dei problemi in classi di difficoltà differente. Non ci vogliamo addentrare qui in una discussione specialistica sull'argomento. Con molta approssimazione si può dire che una classe di problemi è considerata computazionalmente facile se esiste un algoritmo in grado di risolvere qualsiasi esempio di tale classe compiendo un numero di operazioni elementari che è asintoticamente limitato da un polinomio di grado finito nella dimensione di tale esempio. La definizione di "dimensione" di un problema non è banale: per la programmazione lineare dimensioni naturali sono m ed n , ma questo non basta a caratterizzare in modo soddisfacente la reale dimensione del problema. Esisto-

6. IL METODO DEL SIMPLEXSO

no, ad esempio, molte implementazioni del metodo del simplesso che sfruttano la sparsità della matrice dei coefficienti. E' ragionevole pensare quindi che i problemi di *PL* siano in genere tanto più difficili quanto meno sparsa è la matrice dei coefficienti. La definizione comunemente accettata di dimensione di un problema corrisponde alla lunghezza di una codifica efficiente dell'intero insieme dei coefficienti

$$\tilde{A} = \begin{bmatrix} A & b \\ c^T & 0 \end{bmatrix}$$

del problema di *PL*. Ad esempio, volendo memorizzare i dati di un problema di *PL* (con coefficienti interi o razionali) in forma binaria, ed utilizzando il minimo numero possibile di bit per ogni dato ed un separatore tra un dato e l'altro, si vede che occorreranno

$$L = O(mn + \sum_{i,j:\tilde{A}_{ij} \neq 0} \log(|\tilde{A}_{ij}|))$$

bit; si ricorda che il simbolo $O(f(k))$ indica una funzione f tale che

$$\limsup_{k \rightarrow \infty} \frac{O(f(k))}{f(k)} \leq \infty$$

Un algoritmo polinomiale per la *PL* dovrebbe poter risolvere qualsiasi problema di *PL* in un tempo (o in un numero di operazioni elementari) $O(L')$, con r costante opportuna.

La questione se la *PL* fosse o meno una classe di problemi risolvibili in modo teoricamente efficiente, al di là della notevole efficienza pratica del metodo del simplesso, è rimasta aperta fino agli anni '80, quando è stato definitivamente dimostrato da Leonid Kachiyan, col suo metodo degli ellissoidi, (Nemhauser & Wolsey, 1988) che esistono algoritmi polinomiali per la *PL*. Curiosamente, il metodo del simplesso non è un algoritmo polinomiale. Per tutte le implementazioni note del metodo si è potuta trovare una successione di problemi particolari, a lunghezza crescente, per la quale il metodo

6.5. Efficienza del metodo del simplesso

richiede un numero esponenziale di passi di pivot. Il metodo degli ellissoidi, grazie al quale si è dimostrata la polinomialità della *PL*, è un algoritmo estremamente inefficiente in pratica (mentre il simplesso, inefficiente in teoria, ha un comportamento eccellente nella pratica). Solo negli anni '90 si sono sviluppati algoritmi innovativi, detti "a punto interno", (Karmarkar, 1984; Ye, 1997), che hanno complessità polinomiale e che, almeno per problemi di grande dimensione, sono competitivi rispetto al metodo del simplesso.

Per l'implementazione basata sulla scelta della variabile entrante con coefficiente di costo ridotto minimo, il seguente esempio (Klee & Minty, 1972) fornisce un caso pessimo per il metodo del simplesso:

$$\begin{aligned} \max & \sum_{j=1}^n 10^{n-j} x_j \\ 2 \sum_{j=1}^{i-1} & 10^{i-j} x_j + x_i \leq 100^{i-1} \quad i = 1, \dots, n \\ x_j & \geq 0 \quad j = 1, \dots, n \end{aligned}$$

Esempio 17. Per $n = 4$ il problema diventa

$$\begin{aligned} \max & [1000 \ 100 \ 10 \ 1 \ 0 \ 0 \ 0 \ 0] x \\ \left[\begin{array}{ccccccc} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 20 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 200 & 20 & 1 & 0 & 0 & 0 & 1 & 0 \\ 2000 & 200 & 20 & 1 & 0 & 0 & 0 & 1 \end{array} \right] x &= \left[\begin{array}{c} 1 \\ 100 \\ 10000 \\ 1000000 \end{array} \right] \\ x &\geq 0 \end{aligned}$$

Si nota subito come questo sia un problema mal scalato, cioè con coefficienti il cui ordine di grandezza è molto diverso. A causa della diversa grandezza dei termini noti, l'insieme ammissibile non

6. IL METODO DEL SIMPLEXO

è molto diverso dall'iperrettangolo

$$\begin{aligned}0 &\leq x_1 \leq 1 \\0 &\leq x_2 \leq 100 \\0 &\leq x_3 \leq 10\,000 \\0 &\leq x_4 \leq 1\,000\,000.\end{aligned}$$

Eseguendo il metodo del simplesso a partire dalla base $A_{5,6,7,8}$, corrispondente all'origine nello spazio delle prime 4 incognite, scegliendo la variabile entrante con coefficiente di costo ridotto minimo, vengono visitati tutti i 2^4 vertici dell'iperrettangolo deformato.

L'algoritmo del simplesso ha rappresentato, e rappresenta tuttora, un campo di ricerca estremamente ricco ed aperto. Non è ancora stata fornita, ad esempio, una soddisfacente giustificazione teorica del fatto che, nonostante i risultati fortemente negativi della *teoria* della complessità, l'algoritmo del simplesso sia *in pratica*, un metodo estremamente efficiente, utilizzato per risolvere problemi con centinaia di migliaia di variabili. Esistono studi sulla complessità media del metodo (ad esempio (Borgwardt, 1982)), ma in genere i modelli stocastici su cui sono basati non sono sufficientemente vicini ai problemi reali. Studi sia empirici che teorici tendono a sostenere che il metodo del simplesso termina con un numero di passi di pivot $O(m)$, in genere compresi tra $3m$ e $4m$.

Un'altra questione tuttora aperta è la seguente: per ogni tipo di strategia di pivoting finora tentata, si è potuto determinare un controesempio per il quale il metodo richiede un numero esponenziale di pivot. Ma questo non è sufficiente per concludere che non può esistere una strategia di pivoting che renda il metodo polinomiale. A tutt'oggi non è ancora chiaro se tale strategia possa esistere o meno.

Questo problema, tuttora aperto ed oggetto di intensa ricerca, è in qualche modo legato alla cosiddetta *congettura di Hirsch* secondo la quale un poliedro in \mathbb{R}^n con d facce massimali (cioè a dimensione

6.5. Efficienza del metodo del simplesso

$n - 1$) ha “diametro” non maggiore di $d - n$, dove per *diametro di un poliedro* si intende la massima distanza fra due vertici e per *distanza fra due vertici* si intende il minimo numero di spigoli che è necessario percorrere in un cammino da un vertice ad un altro. Ad esempio, dato un pentagono in \mathbb{R}^2 , è evidente che nessuna coppia di nodi dista più di $5 - 2 = 3$ spigoli. La congettura ha a che fare con il metodo del simplesso, poichè il diametro di un poliedro è un limite inferiore al numero di passi di pivot che una qualunque implementazione dell’algoritmo dovrà effettuare nel caso peggiore. Quindi se si dimostrasse, ad esempio, che esistono poliedri per i quali il diametro è esponenziale nel numero, ad esempio, di vincoli o variabili, si sarebbe dimostrato che è impossibile costruire una regola di pivoting per il simplesso che garantisca la terminazione in un numero polinomiale di iterazioni. La congettura è stata recentemente rifiutata (Kim & Santos, 2010); tuttavia si è solo dimostrato che esistono poliedri per i quali il diametro è maggiore di $d - n$. La congettura è vera in tutti i politopi ottenuti come chiusura convessa di punti a coordinate intere limitati (ad esempio, se tutti i vertici del politopo sono binari, la congettura vale). In generale, il controesempio di Santos (Kim & Santos, 2010) dimostra che la congettura di Hirsch è falsa, ma è ancora aperta la questione se il diametro di un generico politopo sia o meno limitato da una funzione polinomiale di $d - n$ (addirittura, una funzione lineare). Fino alla chiusura, in un senso o nell’altro, di questa ipotesi, la ricerca di una strategia efficiente di pivoting sarà ancora molto attiva.

Parte II

Teoria della dualità

Introduzione alla dualità

7.1 Introduzione

Per introdurre l'idea fondamentale della dualità, si immagini di dover risolvere un problema di *PL* e di non avere un controllo diretto sull'algoritmo risolutivo, cosa che avviene in pratica molto di frequente, ma di poter semplicemente osservare una sequenza di soluzioni proposte da un meccanismo di generazione. Tale meccanismo può essere lo stesso metodo del simplesso, ma si potrebbe anche pensare che esso sia un "oracolo" in grado di fornire delle soluzioni ammissibili (non necessariamente di base) per il problema, oppure un generatore casuale di soluzioni ammissibili. Sia che si tratti di un generatore casuale di soluzioni, sia che si stia utilizzando un prodotto software commerciale, è sensato porsi il problema di determinare se le soluzioni generate sono, secondo qualche criterio, vicine a quella ottimale; nel caso di un algoritmo che termina, ci si può anche porre il problema di verificare, a posteriori, se la soluzione prodotta è effettivamente una soluzione ottimale, o per lo meno una buona approssimazione di tale soluzione. Se le soluzioni generate fossero soluzioni di base, si potrebbe pensare di analizzare i coefficienti di costo ridotto per certificare l'ottimalità: in realtà non è detto che questa sia una strategia corretta, poiché la condizione sui coefficienti di costo ridotto è solamente una condizione suffi-

7. INTRODUZIONE ALLA DUALITÀ

ciente e potrebbe capitare che un algoritmo per la programmazione lineare terminasse producendo una soluzione ottima di base con coefficienti di costo ridotto non tutti non negativi, anche se questa è una possibilità abbastanza poco realistica nella pratica, essendo il test sul segno dei coefficienti di costo ridotto il più comune test d'arresto per tutti i metodi che si ispirano al simplex. Inoltre non è detto che un algoritmo generico produca, sia nel corso delle iterazioni sia al termine, soluzioni di base. Gli algoritmi di programmazione lineare di concezione più recente, i cosiddetti metodi a punto interno, generano soluzioni strettamente ammissibili, interne al poliedro delle soluzioni ammissibili del problema di *PL*. Per questi algoritmi occorre quindi un metodo di verifica della qualità della soluzione trovata, un metodo, tra l'altro, che possa essere utilizzato come criterio d'arresto alternativo rispetto a quello basato sui coefficienti di costo ridotto.

In molti metodi di approssimazione numerica viene utilizzata una tecnica di controllo della qualità dell'approssimazione consistente nell'aggiornare, ad ogni iterazione k , due valori ℓ_k, u_k che costituiscono rispettivamente una limitazione inferiore ed una superiore per il valore della quantità cercata. Nel caso dei metodi di ottimizzazione, indicato con z^* il valore dell'ottimo, si cercano quantità tali che

$$z^* \in [\ell_k, u_k] \quad k = 0, 1, \dots$$

Inizialmente si pone in genere $\ell_0 = -\infty$ e $u_0 = +\infty$ (a meno che non si disponga di qualche informazione a priori sul problema che permetta di iniziare con stime più precise di una o entrambe le limitazioni); l'intervallo $[\ell_k, u_k]$ rappresenta un intervallo di incertezza sul valore dell'ottimo e, naturalmente, si desidera rendere minima la quantità $u_k - \ell_k$. La speranza è di giungere ad un'iterazione in cui $\ell_k = u_k$, che è garanzia di ottimalità. Una maggiorazione di z^* nei problemi di minimo è molto semplice da ottenere, ponendo

$$u_k = \min\{c^T x_k, u_{k-1}\} \quad k = 1, 2, \dots,$$

7.2. Rilassamento e duale lagrangiano

dove con x_k si è indicata la k -esima soluzione ammissibile proposta dall'algoritmo o dall'oracolo. In altri termini, u_k è il miglior valore osservato fino alla k -esima iterazione. Più complesso è determinare delle buone limitazioni inferiori.

Uno degli scopi della teoria della dualità è quello di fornire delle buone limitazioni inferiori al valore dell'ottimo del problema di PL . Anche se sono molti gli aspetti interessanti ed utili della dualità, si inizierà da questo tema per presentarne le idee fondamentali. Iniziamo con una presentazione di tipo generale, utile per generare limitazioni inferiori di buona qualità per problemi generici, non solo lineari.

7.2 Rilassamento e duale lagrangiano

Sia dato un generico problema di ottimizzazione (P):

$$\min_{x \in X} f(x)$$

dove $X \subseteq \mathbb{R}^n$ è un qualsiasi sottoinsieme.

Definizione 20. Un problema di ottimizzazione (Q) definito come

$$\min_{x \in Y} s(x)$$

con $Y \subseteq \mathbb{R}^n$ si definisce *rilassamento* del problema (P) se e solo se:

1. $X \subseteq Y$
2. se $x \in X$ allora $s(x) \leq f(x)$.

Lo scopo di un rilassamento è quello di permettere la determinazione di un limite inferiore al valore dell'ottimo di un qualsiasi problema (di minimo). Per ottenere un rilassamento, occorre “allargare” l'insieme ammissibile o, rimanendo nell'insieme ammissibile originale, minorare la funzione obiettivo. Vale una proprietà elementare:

7. INTRODUZIONE ALLA DUALITÀ

Teorema 28. Se il problema (Q) ammette ottimo finito $s^* = \min_{y \in Y} s(y)$, allora

$$s^* \leq f(x) \quad \forall x \in X$$

Dimostrazione. Scelto un qualsiasi punto $x \in X$, sia f che s sono valutabili in tale punto; inoltre vale $f(x) \geq s(x)$; ma $s(x) \geq s^*$ per definizione di minimo e, pertanto, la tesi è dimostrata. \square

Un rilassamento, pertanto, è un problema di ottimizzazione dalla cui soluzione ottimale è possibile ricavare una limitazione inferiore ai valori associati alle soluzioni ammissibili del problema originale. Se il problema originale ammette ottimo, l'ottimo del rilassamento è un limite inferiore al valore di tale ottimo.

Rilassamenti banali si ottengono, ad esempio, eliminando uno o più vincoli. L'eliminazione di vincoli è, in un certo senso, troppo drastica. Il *rilassamento lagrangiano* pur eliminando alcuni vincoli, ne tiene conto sotto forma di una modifica alla funzione obiettivo.

Per un qualunque problema di ottimizzazione

$$\min_{x \in A} f(x) \tag{7.1}$$

ed un suo rilassamento

$$\min_{x \in B} g(x) \tag{7.2}$$

vale il seguente:

Teorema 29. Data una soluzione ottimale del problema (7.2) $x^* \in B$, se vale $x^* \in A$ e $f(x^*) = g(x^*)$, allora x^* è anche soluzione ottimale del problema (7.1).

Dimostrazione. Per un problema ed un suo qualsiasi rilassamento, vale la relazione

$$z^* = \min_{x \in B \subseteq \mathbb{R}^n} g(x) \leq \min_{x \in A \subseteq \mathbb{R}^n} f(x).$$

7.2. Rilassamento e duale lagrangiano

Vale quindi $z^* = g(x^*) = f(x^*)$. Se, come ipotizzato, $x^* \in A$, deve essere ottimo anche per il problema originale. Infatti, se esistesse una soluzione ammissibile $y \in A : f(y) < f(x^*)$, si avrebbe anche $y \in B$ e $g(y) \leq f(y) < f(x^*) = g(x^*)$. \square

Si consideri un generico problema di ottimizzazione vincolata

$$\begin{aligned} \min f(x) \\ g_i(x) \geq 0 & \quad i = 1, \dots, m \\ x \in S \subseteq \mathbb{R}^n \end{aligned}$$

nel quale alcuni vincoli sono stati messi in evidenza come disequazioni, mentre altri concorrono alla definizione dell'insieme S . Si dice *rilassamento lagrangiano* rispetto ai vincoli $g_i(x) \geq 0$ il seguente problema:

$$L(\lambda) = \min_{x \in S} f(x) - \sum_{i=1}^m \lambda_i g_i(x)$$

dove $\lambda \geq 0$ è un vettore di parametri, detto vettore dei *moltiplicatori di Lagrange*. Si può verificare che

Teorema 30. *Il rilassamento lagrangiano di un qualsiasi problema di ottimizzazione è un rilassamento*

Dimostrazione. Infatti l'insieme ammissibile del rilassamento contiene quello originale, essendo ottenuto da questo tramite cancellazione dei vincoli $g_i(x) \geq 0$. Inoltre, per la funzione obiettivo, se x è una soluzione ammissibile del problema, si avrà $g_i(x) \geq 0$ per ogni i e, pertanto,

$$f(x) - \sum_i \lambda_i g_i(x) \leq f(x)$$

grazie alla non negatività dei moltiplicatori di Lagrange. \square

7. INTRODUZIONE ALLA DUALITÀ

E' anche possibile definire un rilassamento lagrangiano rispetto a vincoli di egualianza: se il problema originale ha la forma

$$\begin{aligned} & \min f(x) \\ & h_i(x) = 0 \quad i = 1, \dots, \ell \\ & x \in S \end{aligned}$$

un rilassamento lagrangiano rispetto ai vincoli $h_i(x) = 0$ è dato dal seguente problema:

$$\min_{x \in S} f(x) - \sum_{i=1}^m \lambda_i h_i(x)$$

dove i moltiplicatori di lagrange λ non sono vincolati in segno. Si può facilmente verificare che anche questo è effettivamente un rilassamento. La verifica può essere effettuata sia applicando direttamente la definizione di rilassamento, sia vedendo i vincoli di egualianza come una coppia di disequazioni ed applicando il rilassamento lagrangiano.

Come semplice applicazione, si consideri un problema di PL standard:

$$\begin{aligned} & \min c^T x \\ & Ax = b \\ & x \geq 0 \end{aligned}$$

Il rilassamento rispetto ai vincoli di egualianza ha la forma

$$L(\lambda) = \min_{x \geq 0} c^T x - \lambda^T (Ax - b)$$

o, equivalentemente,

$$L(\lambda) = \lambda^T b + \min_{x \geq 0} (c^T - \lambda^T A)x$$

7.2. Rilassamento e duale lagrangiano

Il problema di ottimizzazione risultante è banale: se λ viene scelto in modo tale che almeno una componente del vettore $(c^T - \lambda^T A)$ sia negativa, allora $L(\lambda)$ sarà illimitato: infatti in questo caso, ponendo uguali a zero tutte le componenti di x tranne una il cui coefficiente sia negativo, si vede immediatamente che diventa possibile incrementare a piacere tale componente senza perdita di ammissibilità. Al contrario, se λ è scelto in modo tale che $(c^T - \lambda^T A) \geq 0$, allora la soluzione ottimale del problema si otterrà ponendo $x = 0$. Risulta quindi:

$$L(\lambda) = \begin{cases} \lambda^T b & \text{se } c^T - \lambda^T A \geq 0 \\ -\infty & \text{altrimenti} \end{cases}$$

Con la tecnica del rilassamento lagrangiano si genera una famiglia di rilassamenti, aventi come indice il vettore dei moltiplicatori; e poiché ogni rilassamento fornisce una limitazione inferiore al valore dell'ottimo, sarebbe utile conoscere il “miglior” rilassamento lagrangiano $\max L(\lambda)$.

Definizione 21. Il problema

$$D = \max_{\lambda} L(\lambda)$$

viene detto *duale lagrangiano*.

Nella definizione qui sopra si è sottinteso l’insieme in cui λ può variare: per i problemi con vincoli di diseguaglianza, i moltiplicatori dovranno essere non negativi.

Nel caso della PL, il duale lagrangiano diviene:

$$\begin{aligned} & \max \lambda^T b \\ & \lambda^T A \leq c^T \end{aligned}$$

se esiste almeno un vettore di moltiplicatori λ tale che $\lambda^T A \leq c^T$, mentre altrimenti il duale lagrangiano avrà valore $-\infty$.

7. INTRODUZIONE ALLA DUALITÀ

La teoria della dualità è di fondamentale importanza per la comprensione delle proprietà dei problemi di programmazione lineare, per la costruzione di nuovi algoritmi, per l'analisi del metodo del simplex e di altri algoritmi e per l'interpretazione dei risultati che tale metodo consente. Questa teoria, nel caso lineare, può essere sviluppata in modo indipendente dal metodo del simplex e può, anzi, costituirne le premesse; in questo volume si è preferito porre questa trattazione a valle del metodo del simplex, ma, come si vedrà, la trattazione della teoria della dualità qui presentata si appoggerà solo in minima parte alle proprietà viste nei capitoli precedenti.

Introduzione alla dualità per la generazione di limiti inferiori

La definizione del problema duale come miglior problema per la generazione di limitazioni inferiori in un problema di *PL*, si può anche ricavare per via diretta, senza la necessità di passare attraverso la teoria lagrangiana.

Si consideri per il momento il problema di *PL* in forma standard

$$\begin{aligned} \min c^T x \\ Ax = b \\ x \geq 0. \end{aligned}$$

Si consideri ora una generica combinazione lineare dei vincoli di eguaglianza, ottenuta moltiplicando per un coefficiente λ_1 il primo vincolo

$$\lambda_1 \sum_j a_{1j} x_j = \lambda_1 b_1,$$

per un coefficiente λ_2 il secondo e così via; sommando si ottiene l'equazione:

7.2. Rilassamento e duale lagrangiano

$$\sum_{i=1}^m \lambda_i \sum_{j=1}^n a_{ij} x_j = \sum_{i=1}^m \lambda_i b_i$$

o, in forma vettoriale,

$$\lambda^T A x = \lambda^T b$$

dove λ è un vettore di m coefficienti reali. Raccogliendo a fattor comune le incognite x si ottiene

$$\sum_{j=1}^n \left(\sum_{i=1}^m \lambda_i a_{ij} \right) x_j = \sum_{i=1}^m \lambda_i b_i. \quad (7.3)$$

L'equazione ottenuta in questo modo è un'*equazione valida*, cioè soddisfatta da tutte le soluzioni ammissibili del problema di *PL*. Si immagini ora di poter scegliere i coefficienti della combinazione lineare in (7.3) in modo tale che:

$$\begin{aligned} \sum_{i=1}^m \lambda_i a_{i1} &\leq c_1 \\ \sum_{i=1}^m \lambda_i a_{i2} &\leq c_2 \\ &\vdots \\ \sum_{i=1}^m \lambda_i a_{in} &\leq c_n. \end{aligned}$$

Si avrà in questo caso

$$\begin{aligned} \sum_{i=1}^m \lambda_i b_i &= \sum_{j=1}^n \left(\sum_{i=1}^m \lambda_i a_{ij} \right) x_j \\ &\leq \sum_{j=1}^n c_j x_j \end{aligned}$$

7. INTRODUZIONE ALLA DUALITÀ

grazie alla non negatività delle incognite x . La relazione

$$\sum_{j=1}^n c_j x_j \geq \sum_{i=1}^m \lambda_i b_i$$

è quindi valida per qualsiasi soluzione ammissibile x e fornisce, pertanto, una limitazione inferiore ai valori possibili per la funzione obiettivo e, in particolare, per il valore ottimale. In forma matriciale, se il vettore $\lambda \in \mathbb{R}^m$ è scelto in modo tale che

$$\lambda^T A \leq c^T,$$

allora, essendo $x \geq 0$, si avrà anche

$$\lambda^T b = \lambda^T A x \leq c^T x.$$

Poiché si desidera la limitazione più stringente possibile, si è naturalmente condotti alla ricerca del massimo valore che la quantità $\lambda^T b$ può assumere; la ricerca di tale massimo valore corrisponde alla risoluzione di un altro problema di *PL*

$$\begin{aligned} & \max_{\lambda \in \mathbb{R}^m} \lambda^T b \\ & \lambda^T A \leq c^T \end{aligned} \tag{7.4}$$

che è il duale del problema di *PL* visto sopra. Apparentemente la limitazione inferiore “ottimale” desumibile dalla risoluzione di questo problema duale non è direttamente utilizzabile, poiché la sua determinazione richiede la risoluzione di un altro problema di *PL*; si potrebbe però utilizzare un secondo oracolo, o un algoritmo, ad esempio il simplex stesso, che proponga soluzioni ammissibili λ per il problema (7.4): si vedrà in realtà tra breve che una volta risolto correttamente un problema di *PL*, la soluzione del duale sarà desumibile in modo piuttosto semplice. Indicando con λ_k la soluzione ottenuta al passo k , è possibile aggiornare la limitazione inferiore al valore dell’ottimo tramite la regola

$$\ell_k = \max\{\lambda_k^T b, \ell_{k-1}\}.$$

7.2. Rilassamento e duale lagrangiano

Esempio 18. Si consideri il problema di *PL* seguente

$$\begin{aligned} & \min [1 \ -1 \ 2 \ 1] x \\ & \left[\begin{array}{cccc} 1 & 1 & 0 & -1 \\ 2 & -1 & 1 & 0 \end{array} \right] x = \left[\begin{array}{c} 2 \\ 2 \end{array} \right] \\ & x \geq 0 \end{aligned}$$

Combinando linearmente i vincoli si ottiene l'equazione

$$(\lambda_1 + 2\lambda_2)x_1 + (\lambda_1 - \lambda_2)x_2 + \lambda_2x_3 - \lambda_1x_4 = 2\lambda_1 + 2\lambda_2$$

e, pertanto, se i coefficienti della combinazione lineare sono scelti in modo tale che

$$\begin{aligned} (\lambda_1 + 2\lambda_2) &\leq 1 \\ (\lambda_1 - \lambda_2) &\leq -1 \\ \lambda_2 &\leq 2 \\ -\lambda_1 &\leq 1 \end{aligned}$$

ne segue che la quantità

$$2(\lambda_1 + \lambda_2)$$

costituisce una limitazione inferiore al valore dell'ottimo. Ad esempio, scegliendo $\lambda_1 = -1, \lambda_2 = 0$ i vincoli del duale sono soddisfatti e, pertanto, -2 è un limite inferiore per l'ottimo del problema e per qualunque soluzione ammissibile. Anche i moltiplicatori $\lambda_1 = -\frac{1}{2}, \lambda_2 = \frac{1}{2}$ sono ammissibili per il duale e pertanto un miglior limite inferiore è dato da $2(-\frac{1}{2} + \frac{1}{2}) = 0$. E' facile vedere che il problema di *PL* ammette, ad esempio, la soluzione ammissibile $x = [1 \ 1 \ 1 \ 0]^T$, di costo 2. L'intervallo di incertezza per il valore dell'ottimo è, a questo punto, $[0, 2]$, con un margine di errore pari a 2. Se si considera la soluzione ammissibile per il duale

7. INTRODUZIONE ALLA DUALITÀ

$\lambda_1 = -\frac{1}{3}, \lambda_2 = \frac{2}{3}$ si vede che il limite inferiore può essere aumentato al livello $\frac{2}{3}$, il che dimostra, fra l'altro, che tutte le soluzioni ammissibili del problema dato hanno costo strettamente positivo. L'intervallo di incertezza è ridotto a $[\frac{2}{3}, 2]$. Eseguendo il simplex sul problema dato, si trova che la soluzione ottimale è data da $x = [\frac{4}{3} \quad \frac{2}{3} \quad 0 \quad 0]^T$; si può verificare immediatamente che questa soluzione soddisfa tutti i vincoli del primale. Il suo costo è pari a $\frac{4}{3} - \frac{2}{3} = \frac{2}{3}$: l'intervallo di incertezza è ora ridotto a $[\frac{2}{3}, \frac{2}{3}]$, di ampiezza zero: l'ottimalità della soluzione del problema di PL è dimostrata, senza bisogno di verifiche sui coefficienti di costo ridotto. Inoltre è stata anche certificata l'ottimalità della soluzione del duale $\lambda_1 = -\frac{1}{3}, \lambda_2 = \frac{2}{3}$.

Dualità lineare

8.1 Duale di un generico problema di PL

Si consideri il problema

$$\begin{aligned} \min_x c^T x \\ Ax = b \\ x \geq 0; \end{aligned}$$

che in questo contesto verrà designato col termine di problema *prima*. Si consideri il suo *duale*:

$$\begin{aligned} \max_{\lambda} \lambda^T b \\ \lambda^T A \leq c^T. \end{aligned}$$

Per i problemi primali in forma generica, si definisce duale un problema ottenuto applicando la definizione sopra introdotta ad un problema standard equivalente al primale. Ad esempio, per un problema con vincoli di diseguaglianza

$$\begin{aligned} \min c^T x \\ Ax \geq b \\ x \geq 0, \end{aligned}$$

8. DUALITÀ LINEARE

introducendo le variabili di surplus si ottiene il problema standard equivalente:

$$\begin{aligned} \min & \left[\begin{array}{cc} c^T & 0^T \end{array} \right] \begin{bmatrix} x \\ s \end{bmatrix} \\ & \left[\begin{array}{cc} A & -I \end{array} \right] \begin{bmatrix} x \\ s \end{bmatrix} = b \\ & x, s \geq 0, \end{aligned}$$

il cui duale si può ottenere dal duale standard:

$$\begin{aligned} \max & \lambda^T b \\ \lambda^T & \left[\begin{array}{cc} A & -I \end{array} \right] \leq \left[\begin{array}{cc} c^T & 0^T \end{array} \right], \end{aligned}$$

da cui

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

Si può quindi osservare che se nel problema primale sono presenti vincoli di diseguaglianza (di \geq), nel duale le variabili, cioè i moltiplicatori, sono vincolate in segno. Si può vedere facilmente che se i vincoli del problema primale sono di \leq il vincolo di segno nel duale diventa un vincolo di non positività. Si può però anche osservare che la definizione di problema duale di un problema generico non è priva di ambiguità. Infatti non esiste un metodo unico per trasformare un problema in un equivalente problema standard e, di conseguenza, non esiste un unico problema duale di un problema dato. Ad esempio, per il problema

$$\begin{aligned} \min & c^T x \\ Ax & \leq b \\ x & \geq 0, \end{aligned}$$

8.1. Duale di un generico problema di *PL*

utilizzando la procedura prima descritta si ottiene il duale

$$\begin{aligned} \max \lambda^T b \\ \lambda^T A \leq c^T \\ \lambda \leq 0, \end{aligned}$$

mentre trasformando il problema nell'equivalente

$$\begin{aligned} \min c^T x \\ -Ax \geq -b \\ x \geq 0, \end{aligned}$$

si giunge al duale

$$\begin{aligned} \max -\lambda^T b \\ -\lambda^T A \leq c^T \\ \lambda \geq 0, \end{aligned}$$

che, anche se equivalente al duale precedentemente ottenuto, ne differisce nella forma. Si potrebbe verificare in modo formale, utilizzando la definizione di problemi equivalenti vista nel capitolo 3.3, che i duali ottenuti applicando la trasformazione standard a problemi fra loro equivalenti sono fra loro equivalenti.

Proseguendo nell'analisi dei duali di problemi non standard, se il problema primale ha alcune variabili, y , libere in segno

$$\begin{aligned} \min_{x,y} c_x^T x + c_y^T y \\ A_x x + A_y y = b \\ x \geq 0, \end{aligned}$$

8. DUALITÀ LINEARE

la trasformazione usuale in forma standard porta al problema equivalente

$$\begin{aligned} \min_{x,y} & \left[\begin{array}{ccc} c_x^T & c_y^T & -c_y^T \end{array} \right] \begin{bmatrix} x \\ y^+ \\ y^- \end{bmatrix} \\ & \left[\begin{array}{ccc} A_x & A_y & -A_y \end{array} \right] \begin{bmatrix} x \\ y^+ \\ y^- \end{bmatrix} = b \\ & x, y^+, y^- \geq 0, \end{aligned}$$

il cui duale è

$$\begin{aligned} \max \lambda^T b \\ \lambda^T \left[\begin{array}{ccc} A_x & A_y & -A_y \end{array} \right] \leq \left[\begin{array}{ccc} c_x^T & c_y^T & -c_y^T \end{array} \right], \end{aligned}$$

equivalente a

$$\begin{aligned} \max \lambda^T b \\ \lambda^T A_x \leq c_x^T \\ \lambda^T A_y = c_y^T. \end{aligned}$$

Quindi, ad ogni variabile libera in segno nel primale corrisponde un vincolo di eguaglianza nel duale.

Infine, se il problema primale è un problema non standard di massimo

$$\begin{aligned} \max c^T x \\ Ax = b \\ x \geq 0 \end{aligned}$$

8.1. Duale di un generico problema di *PL*

la trasformazione in forma standard porta al problema

$$\begin{aligned} -\min & -c^T x \\ \text{a.s.} & Ax = b \\ & x \geq 0 \end{aligned}$$

ed il duale risulta

$$\begin{aligned} -\max & \lambda^T b \\ \text{a.s.} & \lambda^T A \leq -c^T. \end{aligned}$$

Si noti il segno di fronte all'operatore max. Il duale può essere trasformato nell'equivalente

$$\begin{aligned} \min & \mu^T b \\ \text{a.s.} & \mu^T A \geq c^T. \end{aligned}$$

ottenuto ponendo $\mu = -\lambda$. Questa definizione di duale di un problema di massimo può anche essere ricavata direttamente (come le altre, del resto) dalla teoria lagrangiana.

Proseguendo nell'analisi dei duali di problemi non standard, si può arrivare alla tabella seguente.

Trasformazione primale \rightarrow duale		$\min c^T x$	$\max c^T x$
		\downarrow	\downarrow
obiettivo	\rightarrow	$\max \lambda^T b$	$\min \lambda^T b$
$x_j \geq 0$	\rightarrow	$\lambda^T A_j \leq c_j$	$\lambda^T A_j \geq c_j$
$x_j \leq 0$	\rightarrow	$\lambda^T A_j \geq c_j$	$\lambda^T A_j \leq c_j$
x_j libera in segno	\rightarrow	$\lambda^T A_j = c_j$	$\lambda^T A_j = c_j$
$a_i^T x = b_i$	\rightarrow	λ_i libera in segno	λ_i libera in segno
$a_i^T x \geq b_i$	\rightarrow	$\lambda_i \geq 0$	$\lambda_i \leq 0$
$a_i^T x \leq b_i$	\rightarrow	$\lambda_i \leq 0$	$\lambda_i \geq 0$

8. DUALITÀ LINEARE

Si sarà notato che il collegamento tra primale e duale è tale da associare ai vincoli del primale variabili del duale e viceversa. Ad ogni vincolo del primale corrisponde una ben precisa variabile del duale che, nell'interpretazione data all'inizio del capitolo, corrisponde ad un moltiplicatore del vincolo; per i vincoli di uguaglianza, il moltiplicatore può avere segno qualunque, mentre, come si è visto, per i vincoli di diseguaglianza si devono utilizzare moltiplicatori di segno costante, per poter ottenere diseguaglianze valide. Ad ogni variabile del primale è associato un unico vincolo del duale, nato dall'esigenza di costruire un minorante al coefficiente di costo della variabile. Se la variabile è non negativa, è possibile minorare il costo: da qui nasce un vincolo di diseguaglianza per il duale. Se invece una variabile è libera in segno, non è possibile costruire minoranti a partire dalla formulazione prima introdotta: l'unica possibilità è pertanto di uguagliare il costo.

Come applicazione di quanto visto finora si può cercare il duale di un problema duale standard, cioè del problema

$$\begin{aligned} \max \lambda^T b \\ \lambda^T A \leq c^T. \end{aligned}$$

Si può procedere come segue, trasformando il duale in un problema equivalente di minimo:

$$\begin{aligned} -\min -b^T \lambda \\ A^T \lambda + s = c \\ s \geq 0. \end{aligned}$$

da cui si ottiene il duale

$$\begin{aligned} -\max y^T c \\ y^T A^T = -b^T \\ y \leq 0 \end{aligned}$$

che è equivalente al problema

$$\begin{aligned} \min x^T c \\ x^T A^T = b^T \\ x \geq 0 \end{aligned}$$

che, evidentemente, coincide con il problema di partenza. Si è quindi dimostrato il seguente

Teorema 31. *Il duale del duale di un problema di PL è equivalente al problema di PL stesso.*

Parlando di problemi di programmazione lineare e dei rispettivi duali, siamo pertanto autorizzati ad usare il termine di *coppia* di problemi, l'uno duale dell'altro, senza distinzione tra un primale ed il duale; resta soltanto il fatto che uno dei due problemi consiste nella minimizzazione di un obiettivo lineare, l'altro nella massimizzazione di un altro obiettivo lineare.

8.2 Teoremi di dualità

La proprietà che ha permesso l'introduzione del problema duale è comunemente nota con il nome di *dualità debole*. Formalmente vale il seguente

Teorema 32. *(di dualità debole): Data una coppia di problemi duali che ammettono soluzioni ammissibili, il valore di ogni soluzione ammissibile del problema di minimo è limitato inferiormente dal valore di ogni soluzione ammissibile del problema di massimo.*

8. DUALITÀ LINEARE

Dimostrazione. Il problema di minimo può essere sempre trasformato nel problema standard equivalente

$$\begin{aligned} \min c^T x \\ Ax = b \\ x \geq 0 \end{aligned}$$

il cui duale è:

$$\begin{aligned} \max \lambda^T b \\ \lambda^T A \leq c^T. \end{aligned}$$

Proseguendo come nel capitolo introduttivo, si ottiene

$$\lambda^T A x = \lambda^T b$$

e, essendo $\lambda^T A \leq c^T x \geq 0$, si ottiene la limitazione

$$\lambda^T b \leq c^T x.$$

□

Date due soluzioni ammissibili, \bar{x} per un problema di *PL* di minimo (ad esempio standard) e $\bar{\lambda}$ per il suo duale, l'intervallo $[\bar{\lambda}^T b, c^T \bar{x}]$ si dice *intervallo di dualità* e la quantità $c^T \bar{x} - \bar{\lambda}^T b$ si dice *gap di dualità*. Se il gap di dualità si riduce a 0, allora è dimostrata l'ottimalità della soluzione del primale e del duale:

Corollario 1. *Data una coppia di problemi, uno duale dell'altro, se per ciascuno dei due problemi è nota una soluzione ammissibile e le due soluzioni, nei rispettivi problemi, hanno lo stesso costo, allora esse sono ottimali.*

8.2. Teoremi di dualità

Dimostrazione. Se sono note due soluzioni particolari \bar{x} del primale, di minimo, e $\bar{\lambda}$ del duale, di massimo, poiché deve valere

$$c^T x \geq \bar{\lambda}^T b = c^T \bar{x} \quad \forall x : Ax = b, x \geq 0$$

segue immediatamente che \bar{x} è una soluzione ottimale del primale. La dimostrazione per il duale è del tutto analoga. \square

Si è quindi visto che il metodo presentato per la determinazione di limitazioni inferiori al valore dell'ottimo in un problema di *PL* di minimo è corretto e permette di fornire una dimostrazione di ottimalità. Non è ancora chiaro però se tale dimostrazione possa essere sempre ottenuta con il metodo proposto. La risposta è affermativa ed è contenuta, in forma rafforzata, nel teorema seguente.

Teorema 33. (*di dualità forte*): *Se un problema di programmazione lineare ammette ottimo, anche il suo duale ammette ottimo; inoltre il costo di ogni soluzione ottimale del primale coincide con quello del duale.*

Dimostrazione. Grazie alla proprietà simmetrica della relazione di dualità si può pensare che il problema primale sia di minimo e, grazie alle proprietà di equivalenza, si può immaginare che sia in forma standard. Inoltre è anche possibile ipotizzare, senza ledere la generalità, che il problema primale sia ridotto al rango e pertanto, grazie al teorema fondamentale della *PL*, avendo ipotizzato che il problema ammetta ottimo, ammetterà almeno un ottimo *di base*. Grazie a quanto visto nella descrizione del metodo del simplex, è noto che se un problema di *PL* ammette ottimo di base, allora ammette anche una soluzione ottimale di base *con coefficienti di costo ridotto non negativi*. Sia B la base alla quale sono associati coefficienti di costo ridotto non negativi. La soluzione ottimale di base associata a B è data da

$$\bar{x} = \begin{bmatrix} A_B^{-1} b \\ 0 \end{bmatrix}$$

8. DUALITÀ LINEARE

I coefficienti di costo ridotto associati alla base A_B sono dati da

$$\hat{c}^T = c_N^T - c_B^T A_B^{-1} A_N \geq 0.$$

Il vettore

$$\lambda_B^T = c_B^T A_B^{-1}$$

è una soluzione ammissibile per il duale. Infatti,

$$\begin{aligned}\lambda_B^T A &= \lambda_B^T [A_B \quad A_N] \\ &= [\lambda_B^T A_B \quad \lambda_B^T A_N] \\ &= [c_B^T A_B^{-1} A_B \quad c_B^T A_B^{-1} A_N] \\ &= [c_B^T \quad c_B^T A_B^{-1} A_N] \\ &\leq [c_B^T \quad c_N^T]\end{aligned}$$

dove per il primo blocco la diseguaglianza è banalmente verificata, mentre per il secondo essa corrisponde alla non negatività dei coefficienti di costo ridotto. Quindi si è dimostrato che se un problema di *PL* ammette ottimo (di base con coefficienti di costo ridotto non negativi) allora il problema duale è ammissibile, e si è trovata una particolare soluzione ammissibile. Il costo, nel duale, di tale soluzione ammissibile è

$$\lambda_B^T b = c_B^T A_B^{-1} b.$$

Ma questa quantità coincide con il costo della soluzione ottimale del primale:

$$c^T \bar{x} = c_B^T \bar{x}_B = c_B^T A_B^{-1} b$$

e pertanto, grazie al corollario, si è dimostrato che il gap di dualità è ridotto a zero. \square

Il sistema individuato per la determinazione di limitazioni inferiori per i problemi di *PL* di minimo è pertanto in grado di fornire un certificato di ottimalità per qualsiasi problema di *PL* che ammetta ottimo.

Se un problema non ammette ottimo, la situazione deve ancora essere analizzata. Si ricava molto facilmente la proprietà seguente:

Teorema 34. *Se un problema di PL è illimitato, il suo duale è privo di soluzioni ammissibili.*

Dimostrazione. Si può senza perdita di generalità dimostrare la proprietà immaginando che il problema sia di minimo. Se tale problema è illimitato allora esiste una successione di soluzioni ammissibili x_1, x_2, \dots tale che

$$\lim_{k \rightarrow \infty} c^T x_k = -\infty.$$

Se il duale ammettesse una soluzione ammissibile $\bar{\lambda}$, la quantità $\bar{\lambda}^T b$ costituirebbe una limitazione inferiore ai valori delle soluzioni ammissibili del primale. Ma questo è assurdo perché in questo caso si avrebbe

$$c^T x_k \geq \bar{\lambda}^T b \quad \forall k.$$

□

Se un problema di PL è inammissibile, sicuramente il suo duale non può ammettere soluzione ottimale. Tuttavia, per il duale, sono possibili entrambe le possibilità: o il problema duale è illimitato, oppure è inammissibile (possono quindi esistere coppie primale/duale in cui entrambi i problemi non sono ammissibili).

Esempio 19. Si consideri il problema

$$\begin{aligned} & \min x \\ & x - y + z = 2 \\ & x + z + s = 1 \\ & x, y, z, s \geq 0. \end{aligned}$$

Questo problema non ammette soluzioni ammissibili, pur essendo in forma standard e con la matrice dei coefficienti ridotta al

8. DUALITÀ LINEARE

rango. La verifica di questa affermazione potrebbe essere effettuata applicando la prima fase del metodo due fasi. Tuttavia si può vedere molto semplicemente che il primo vincolo è equivalente a $x + z \geq 2$, mentre il secondo equivale a $x + z \leq 1$, due vincoli evidentemente incompatibili fra loro. Il duale di questo problema è

$$\begin{aligned} & \max 2\lambda_1 + \lambda_2 \\ & \lambda_1 + \lambda_2 \leq 1 \\ & -\lambda_1 \leq 0 \\ & \lambda_1 + \lambda_2 \leq 0 \\ & \lambda_2 \leq 0 \end{aligned}$$

e risulta illimitato superiormente. Anche in questo caso la verifica potrebbe essere fatta attraverso il metodo del simplesso, anche se, data la semplicità del problema, è possibile una verifica diretta: ad esempio scegliendo $\lambda_1 = k, \lambda_2 = -k$, con $k \geq 0$, si ottengono soluzioni ammissibili di costo pari a k , che può essere scelto grande a piacere.

Esempio 20. Si consideri la famiglia di problemi seguente, nella quale compare un parametro reale a :

$$\begin{aligned} & \min x + ay \\ & x - y + z = 2 \\ & x - y - z = 3 \\ & x, y, z \geq 0 \end{aligned}$$

Si può vedere facilmente che il primo vincolo equivale a $x - y \leq 2$ ed il secondo a $x - y \geq 3$. Quindi, per qualunque valore di a il

8.2. Teoremi di dualità

problema risulta inammissibile. Il duale ha la forma

$$\begin{aligned} \max & 2\lambda_1 + 3\lambda_2 \\ \text{s.t. } & \lambda_1 + \lambda_2 \leq 1 \\ & -\lambda_1 - \lambda_2 \leq a \\ & \lambda_1 - \lambda_2 \leq 0 \end{aligned}$$

Se $a < -1$ il problema non è ammissibile: infatti il primo vincolo $\lambda_1 + \lambda_2 \leq 1$ è in contrasto con il secondo $\lambda_1 + \lambda_2 \geq -a$. Se invece $a \geq -1$ il problema duale è ammissibile e illimitato: basta ad esempio scegliere la successione $\lambda_2 = k, \lambda_1 = 1 - k$ che è ammissibile per ogni $k \geq \frac{1}{2}$ ed ha costo pari a $2 + k$. Il problema in questo caso risulta essere illimitato. Più semplicemente, per verificare il fatto che il duale risulta illimitato, basterebbe in questo caso verificare l'esistenza di una soluzione ammissibile per il duale: infatti se il primale non ammette ottimo, nemmeno il duale può ammetterlo e, pertanto, non essendo inammissibile dovrà sicuramente essere illimitato.

Riassumendo, le situazioni possibili per una coppia primale/duale possono essere rappresentate dalla tabella seguente:

	\exists ottimo finito	illimitato	non ammissibile
\exists ottimo finito	sì	no	no
illimitato	no	no	sì
non ammissibile	no	sì	sì

8. DUALITÀ LINEARE

Dalla tabella, e da quanto mostrato finora, si può dedurre, tra l'altro, che se (e solo se) sia il primale che il duale sono ammissibili, allora entrambi ammettono soluzione ottimale.

La teoria della dualità svolta fino a questo punto sfrutta il metodo del simplesso e, in particolare, la dimostrazione costruttiva di finitezza del simplesso, laddove, nel teorema forte, si ipotizza di avere a disposizione una soluzione di base ottimale con coefficienti di costo ridotto non negativi. In realtà la teoria può essere sviluppata in modo indipendente dall'algoritmo del simplesso, appoggiandosi ai cosiddetti teoremi dell'alternativa, il più noto dei quali prende il nome di *Lemma di Farkas* (Farkas, 1902; Dax, 1997). Avendo già dimostrato i teoremi di dualità, è possibile utilizzarli per giungere in modo elementare alla dimostrazione del lemma di Farkas.

Teorema 35. (*Lemma di Farkas*) *Data una matrice $A \in \mathbb{R}^{m \times n}$ ed un vettore $b \in \mathbb{R}^m$, il sistema di vincoli*

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

è ammissibile se e solo se non lo è il sistema

$$\begin{aligned} A^T y &\leq 0 \\ b^T y &> 0. \end{aligned}$$

Dimostrazione. Il sistema $Ax = b, x \geq 0$ ammette soluzioni ammissibili se e solo se il problema di *PL*

$$\begin{aligned} \min 0^T x \\ Ax = b \\ x \geq 0 \end{aligned}$$

ammette ottimo (il cui valore sarà necessariamente 0). Utilizzando la teoria della dualità si deduce che l'ammissibilità del sistema

8.2. Teoremi di dualità

originale è quindi equivalente al fatto che anche il problema duale ammette ottimo, di valore 0. Il duale del problema di *PL* precedente è dato da

$$\begin{aligned} \max \lambda^T b \\ \lambda^T A \leq 0. \end{aligned}$$

In questo problema duale la soluzione $\lambda = 0$ è ammissibile e, pertanto, il valore 0 costituisce una limitazione inferiore al valore dell'ottimo. Se, e solo se, esistesse una soluzione y tale che $A^T y \leq 0, b^T y > 0$, il problema duale non avrebbe ottimo con valore 0 e, di conseguenza, sarebbe illimitato. \square

Un'altra importante proprietà legata alla dualità fornisce uno strumento generale per la verifica di ottimalità.

Teorema 36. (*Scarti complementari*) Si consideri la coppia di problemi

$$\begin{array}{ll} \min c^T x & \max \lambda^T b \\ Ax = b & \lambda^T A \leq c^T \\ x \geq 0 & \end{array}$$

Due soluzioni \bar{x} e $\bar{\lambda}$, ammissibili rispettivamente per il primale e per il duale sono ottimali se e solo se $\forall j \in 1, \dots, n$

$$\begin{aligned} \bar{x}_j = 0 & \quad \text{oppure} \\ \bar{\lambda}^T A_j = c_j & \end{aligned}$$

Dimostrazione. Si considerino le variabili di slack per il problema duale: $\bar{y} \geq 0 : \bar{\lambda}^T A + \bar{y}^T = c^T$. Le soluzioni date, essendo ammissibili per ipotesi, saranno ottimali se e solo se

$$c^T \bar{x} - \bar{\lambda}^T b = 0$$

8. DUALITÀ LINEARE

cioè se e solo se

$$(\bar{\lambda}^T A + \bar{y}^T) \bar{x} - \bar{\lambda}^T b = 0.$$

Essendo $\bar{\lambda}^T A \bar{x} = \bar{\lambda}^T b$, l'ottimalità è equivalente a

$$\bar{y}^T \bar{x} = 0 \quad (8.1)$$

(si noti che tale prodotto scalare coincide con il “gap” di dualità). Poiché sia le componenti di \bar{x} , sia quelle di \bar{y} sono non negative, la (8.1) è equivalente a

$$\bar{x}_j \bar{y}_j = 0 \quad \forall j$$

□

La proprietà degli scarti complementari, dimostrata nel caso standard, è valida per ciascuna coppia di problemi in relazione di dualità l'uno con l'altro. Ad esempio, nel caso di problemi in forma canonica, si ha la proprietà, analoga alla precedente:

Teorema 37. (*Scarti complementari per problemi non standard*) *Data la coppia di problemi*

$$\begin{array}{ll} \min c^T x & \max \lambda^T b \\ Ax \geq b & \lambda^T A \leq c^T \\ x \geq 0 & \lambda \geq 0 \end{array}$$

una coppia di soluzioni ammissibili $\bar{x}, \bar{\lambda}$ è ottimale se e solo se $\forall j \in 1, \dots, n$

$$\begin{array}{ll} \bar{x}_j = 0 & \text{oppure} \\ \bar{\lambda}^T A_j = c_j & \end{array}$$

e, $\forall i \in 1, \dots, m$

$$\begin{array}{ll} \bar{\lambda}_i = 0 & \text{oppure} \\ a_i^T \bar{x} = b_i & \end{array}$$

8.2. Teoremi di dualità

Dimostrazione. La dimostrazione è del tutto analoga alla precedente: siano

$$\begin{aligned}\bar{y} &= A\bar{x} - b \\ \bar{\mu}^T &= c^T - \bar{\lambda}^T A\end{aligned}$$

le variabili di slack/surplus per i vincoli dei due problemi. Le soluzioni date sono ottimali se e solo se

$$c^T \bar{x} - \bar{\lambda}^T b = 0$$

cioè se e solo se

$$\begin{aligned}(\bar{\mu}^T + \bar{\lambda}^T A)\bar{x} - \bar{\lambda}^T(A\bar{x} - \bar{y}) &= 0 \\ \bar{\mu}^T \bar{x} + \bar{\lambda}^T \bar{y} &= 0\end{aligned}$$

ma poiché $\bar{\mu} \geq 0, \bar{x} \geq 0, \bar{\lambda} \geq 0, \bar{y} \geq 0$, segue la tesi. \square

Il teorema degli scarti complementari permette di verificare l'ottimalità di una coppia di soluzioni ammissibili per un problema ed il suo duale controllando se ad ogni variabile non nulla di uno dei due problemi corrisponde, nel duale, un vincolo attivo, cioè con slack nullo. Essendo la proprietà necessaria e sufficiente, può essere anche utilizzata, data una soluzione ottimale di un problema, per determinare *tutte* le soluzioni ottimali del duale.

Esempio 21. Si consideri il problema

$$\begin{aligned}\min & [1 \quad -1 \quad -1 \quad 0] x \\ & \left[\begin{array}{cccc} 1 & 0 & 2 & -1 \\ 0 & 1 & 1 & 2 \end{array} \right] x = \left[\begin{array}{c} 2 \\ 1 \end{array} \right] \\ & x \geq 0\end{aligned}$$

8. DUALITÀ LINEARE

Una soluzione ottimale è data da

$$x = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}.$$

Per verificare l'ottimalità di questa soluzione, dopo aver controllato l'ammissibilità, basta determinare una soluzione ammissibile del duale

$$\max \begin{bmatrix} 2 & 1 \end{bmatrix} \lambda$$

$$\lambda^T \begin{bmatrix} 1 & 0 & 2 & -1 \\ 0 & 1 & 1 & 2 \end{bmatrix} \leq \begin{bmatrix} 1 & -1 & -1 & 0 \end{bmatrix}$$

che soddisfi le relazioni di scarti complementari. Poiché il problema primale è in forma standard, tutti gli slack sono nulli. Delle variabili del primale, nella soluzione ottimale, solo x_3 è non nulla. Dovrà quindi esistere una soluzione ammissibile del duale per la quale sia nullo lo slack del terzo vincolo, cioè

$$\lambda^T \begin{bmatrix} 2 \\ 1 \end{bmatrix} = -1.$$

Gli ottimi del duale saranno tutte e sole le soluzioni del sistema

$$\begin{aligned} \lambda_2 &= -1 - 2\lambda_1 \\ \lambda_1 &\leq 1 \\ -1 - 2\lambda_1 &\leq -1 \\ -\lambda_1 + 2(-1 - 2\lambda_1) &\leq 0 \end{aligned}$$

e cioè, dopo brevi calcoli,

$$\begin{aligned} \lambda_1 &\in [0, 1] \\ \lambda_2 &= -1 - 2\lambda_1 \end{aligned}$$

8.2. Teoremi di dualità

In questo modo, non solo si è certificata l'ottimalità della soluzione trovata per il primale, ma si sono determinate anche tutte le soluzioni ottimali del duale che, in questo caso, sono infinite. Volendo potremmo utilizzare la stessa tecnica per verificare se il primale ammette soluzioni ottimali alternative a quella data. Scegliendo una qualunque soluzione ottimale del duale, ad esempio $\lambda_1 = 1, \lambda_2 = -3$, si ha che i vincoli del duale non attivi sono il secondo ed il quarto. Ogni soluzione ottimale del primale dovrà dunque avere $x_2 = x_4 = 0$. Il sistema che si ottiene eliminando tali incognite dal primale è quadrato e non singolare e, pertanto, ammette come unica soluzione quella già trovata. Il primale dunque ammette un'unica soluzione ottimale.

Il teorema degli scarti complementari afferma quindi che affinché una coppia di soluzioni ammissibili, una del primale, l'altra del duale, sia ottimale è necessario e sufficiente che a ciascuna componente non nulla di una delle due soluzioni corrisponda, nell'altro problema, un vincolo attivo, cioè a slack nullo. Può succedere che siano nulli sia una variabile che lo slack del vincolo ad essa associato. Quando questo *non* succede, si suol dire che le soluzioni soddisfano alla condizione di complementarità *stretta*. Si può dimostrare che in una coppia di problemi duali che ammettono ottimo (basta in realtà che siano ammissibili, dato che l'ammissibilità della coppia primale/duale è equivalente all'esistenza di un ottimo per entrambi i problemi) esiste sempre almeno una coppia di soluzioni strettamente complementari.

Nella dimostrazione del teorema di dualità forte si è utilizzata la soluzione particolare del duale

$$\bar{\lambda}_B^T = c_B^T A_B^{-1}$$

Questa soddisfa sempre le relazioni di complementarietà con la soluzione

$$x = \begin{bmatrix} A_B^{-1} b \\ 0 \end{bmatrix}.$$

8. DUALITÀ LINEARE

Nel corso delle iterazioni del metodo del simplesso, ad ogni base generata dall'algoritmo è possibile associare un vettore $\bar{\lambda}_B$ che soddisfa le relazioni degli scarti complementari con la soluzione di base corrente; naturalmente, prima del termine dell'algoritmo, tale soluzione non sarà ammissibile per il duale. La soluzione $\bar{\lambda}_B$ viene detta *soluzione duale complementare* rispetto alla base B .

Un'applicazione della dualità alla programmazione lineare

Nel capitolo 6.3, a pag. 128, si è introdotto il metodo “big M ” per l'inizializzazione del metodo del simplesso. Grazie alla dualità diviene abbastanza semplice dimostrare un'importante proprietà di tale metodo. Si consideri il problema

$$\begin{aligned} \min c^T x \\ Ax = b \\ x \geq 0 \end{aligned} \tag{8.2}$$

ed il corrispondente problema artificiale

$$\begin{aligned} \min c^T x + M\mathbf{1}^T x^a \\ Ax + x^a = b \\ x, x^a \geq 0 \end{aligned} \tag{8.3}$$

Vale il seguente

Teorema 38. *Esiste un $M_0 > 0$ tale che se il problema (8.2) con $b \geq 0$ ammette ottimo finito x^* , allora il problema (8.3) ammette soluzione ottimale $(x^*, 0)$ per ogni $M \geq M_0$.*

Dimostrazione. Infatti, se il problema (8.2) ammette ottimo, anche il suo duale

$$\begin{aligned} \max \lambda^T b \\ \lambda^T A \leq c^T \end{aligned}$$

8.3. Informazione duale nel dizionario del simplexso

ammette ottimo λ^* con $b^T \lambda^* = c^T x^*$. Per il duale del problema (8.3)

$$\begin{aligned} & \max \lambda^T b \\ & \lambda^T A \leq c^T \\ & \lambda \leq M\mathbf{1} \end{aligned}$$

la soluzione λ^* sarà ammissibile, a patto di scegliere M sufficientemente grande. Il costo di tale soluzione nel duale è pari al costo della soluzione ammissibile $x = x^*, x^a = 0$ nel primale, da cui la tesi. \square

8.3 Informazione duale nel dizionario del simplexso

Si è appena visto che la proprietà degli scarti complementari può essere utilizzata per ricavare una o tutte le soluzioni ottimali del duale di un problema dato. Tuttavia non sempre, specialmente per i problemi di grande dimensione, tale sistema si rivela efficiente. In molti casi l'informazione relativa ad una soluzione ottimale del duale è prontamente disponibile al termine dell'esecuzione del metodo del simplexso. Si consideri il caso particolare di un problema che, inizialmente, si trova formulato nel modo seguente:

$$\begin{aligned} & \min c^T x + c_I^T y \\ & Ax + y = b \\ & x, y \geq 0, \end{aligned}$$

un problema cioè nel quale compare la matrice identica come matrice dei coefficienti di alcune variabili. Ad ogni iterazione del metodo del simplexso, data una base B , i coefficienti di costo ridotto associati alla variabile y saranno dati da

$$\begin{aligned} \hat{c}_I^T &= c_I^T - c_B^T A_B^{-1} I \\ &= c_I^T - \bar{\lambda}_B^T \end{aligned}$$

8. DUALITÀ LINEARE

da cui

$$\bar{\lambda}_B^T = c_I^T - \hat{c}_I^T,$$

cioè: data una base B la soluzione complementare del duale può essere trovata come differenza tra il costo delle variabili i cui coefficienti formano, nel problema iniziale, una matrice identica, ed il loro costo ridotto. Ovviamente, se non si è giunti alla soluzione ottimale, la soluzione duale complementare non sarà ammissibile per il duale.

Apparentemente questo metodo molto semplice per determinare la soluzione complementare associata ad una base B sembra di limitata utilità, richiedendo la presenza di una sottomatrice identica nel problema di *PL*. In realtà può essere utilizzato per *qualunque* problema di *PL*: ad esempio, se nella matrice dei coefficienti A è presente una matrice di base facilmente invertibile, ad esempio $A_{B'}$, sfruttando l'equivalenza

$$\hat{c}_{B'}^T = c_{B'}^T - c_B^T A_B^{-1} A_{B'},$$

si ottiene

$$\bar{\lambda}_B^T = (c_{B'}^T - \hat{c}_{B'}^T) A_{B'}^{-1}.$$

Inoltre, se nel problema iniziale non è presente la matrice identica, generalmente si utilizzerà il metodo due fasi, visto nel capitolo 6.3. Tale metodo prevede una matrice dei coefficienti contenente sempre una matrice identica, data dai coefficienti delle variabili artificiali. Nella presentazione del metodo si è supposto che, al termine della prima fase, le variabili artificiali vengano eliminate dal problema prima di passare alla seconda fase. Immaginando invece di mantenere tali variabili fino al termine anche della seconda fase, con la convenzione di lasciarle sempre a livello 0, si possono sfruttare le informazioni necessarie per la determinazione dell'ottimo del duale. In pratica mantenere le variabili artificiali a livello 0 nel problema

8.3. Informazione duale nel dizionario del simplesso

della seconda fase equivale a risolvere il problema

$$\begin{aligned} \min & c^T x + 0^T y \\ Ax + y &= b \\ y &= 0 \\ x &\geq 0, \end{aligned}$$

il cui duale è

$$\begin{aligned} \max & \lambda^T b + \mu^T 0 \\ \lambda^T A &\leq c^T \\ \lambda^T + \mu^T &\leq 0 \end{aligned}$$

che è equivalente al duale del problema standard originale. Data una base B , la soluzione duale complementare ad essa associata si potrà determinare come

$$\bar{\lambda}_B^T = 0^T - \hat{c}_y^T,$$

cioè come l'opposto dei coefficienti di costo ridotto delle variabili artificiali.

Esempio 22. Si consideri il problema di PL

$$\begin{aligned} \min & [1 \quad -1 \quad -1 \quad 2] x \\ \left[\begin{array}{cccc} 1 & 0 & -1 & 1 \\ 1 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 \end{array} \right] x &= \left[\begin{array}{c} 1 \\ 4 \\ 1 \end{array} \right] \\ x &\geq 0. \end{aligned}$$

8. DUALITÀ LINEARE

Eseguendo il metodo due fasi si ottiene il dizionario iniziale

$$\begin{aligned}z &= 6 - 2x_1 - x_2 + x_3 - 4x_4 \\y_1 &= 1 - x_1 + x_3 - x_4 \\y_2 &= 4 - x_1 - x_2 - 2x_4 \\y_3 &= 1 - x_4\end{aligned}$$

da cui si ottiene

$$\begin{aligned}z &= 2 - 2x_1 - x_2 + x_3 + 4y_3 \\y_1 &= -x_1 + x_3 + y_3 \\y_2 &= 2 - x_1 - x_2 + 2y_3 \\x_4 &= 1 - y_3\end{aligned}$$

e, successivamente,

$$\begin{aligned}z &= 2 + 2y_1 - x_3 + 2y_3 - x_2 \\x_1 &= -y_1 + x_3 + y_3 \\y_2 &= 2 + y_1 - x_3 + y_3 - x_2 \\x_4 &= 1 - y_3\end{aligned}$$

e

$$\begin{aligned}z &= y_1 + y_2 + y_3 \\x_1 &= 2 - x_2 - y_2 + 2y_3 \\x_3 &= 2 - x_2 - y_2 + y_1 + y_3 \\x_4 &= 1 - y_3.\end{aligned}$$

8.3. Informazione duale nel dizionario del simplesso

Passando alla seconda fase, si ottiene il dizionario iniziale

$$\begin{aligned}z &= x_1 - x_2 - x_3 + 2x_4 \\x_1 &= 2 - x_2 - y_2 + 2y_3 \\x_3 &= 2 - x_2 - y_2 + y_1 + y_3 \\x_4 &= 1 - y_3\end{aligned}$$

che, in forma canonica, diviene

$$\begin{aligned}z &= 2 - x_2 - y_1 - y_3 \\x_1 &= 2 - x_2 - y_2 + 2y_3 \\x_3 &= 2 - x_2 - y_2 + y_1 + y_3 \\x_4 &= 1 - y_3.\end{aligned}$$

Naturalmente le variabili artificiali devono essere pensate come fissate a 0 e, pertanto, anche se hanno coefficiente di costo ridotto negativo, non entreranno mai in base; inoltre il controllo di ottimalità per l'arresto del metodo del simplesso verrà effettuato basandosi esclusivamente sul segno dei coefficienti di costo ridotto delle variabili regolari. Facendo entrare in base x_2 , si ottiene:

$$\begin{aligned}z &= x_1 - y_1 + y_2 - 3y_3 \\x_2 &= 2 - x_1 - y_2 + 2y_3 \\x_3 &= x_1 + y_1 - y_3 \\x_4 &= 1 - y_3\end{aligned}$$

che è ottimale per il primale (la soluzione è $x_1 = x_3 = 0, x_2 = 2, x_4 = 1$, di costo 0). Da quanto visto risulta anche che una soluzione ottimale per il duale è

$$\lambda_1 = 1, \lambda_2 = -1, \lambda_3 = 3$$

A prima vista potrebbe sembrare uno spreco il dover mantenere le variabili artificiali anche nella seconda fase del metodo del

8. DUALITÀ LINEARE

simplesso. In realtà nelle implementazioni efficienti dell'algoritmo viene utilizzata una struttura dati differente dal dizionario, che permette di mantenere queste informazioni in modo efficiente.

8.4 Interpretazione della dualità

Se un problema di *PL* è dotato di particolare struttura e proviene da una ben precisa classe di modelli applicativi, il duale in genere si presta ad essere interpretato come problema applicativo esso stesso, in qualche modo collegato con il primale e, spesso, fornisce modelli utili direttamente.

Duale del problema della dieta

Si consideri come esempio iniziale il modello della dieta di costo minimo visto nel capitolo 2.1. Indicati con x un vettore di quantità incognite di ciascun cibo da includere nella dieta, con A una matrice di contenuti nutritivi, con b un vettore contenente i requisiti minimi richiesti per ciascun elemento nutritivo, con c un vettore di costi unitari dei cibi, la formulazione risulta essere la seguente:

$$\begin{aligned} \min & c^T x \\ \text{s.t. } & Ax \geq b \\ & x \geq 0 \end{aligned}$$

Il duale di questo problema ha la forma

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

In questo esempio si è ipotizzato il caso in cui tutti gli elementi nutritivi sono “benefici” e, pertanto, una dieta equilibrata ne fissa il

8.4. Interpretazione della dualità

quantitativo minimo (vincoli di \geq nel primale). Per interpretare il problema duale si può procedere come segue. Innanzitutto, anche se le variabili del primale e del duale appartengono a spazi euclidei differenti, i costi, in entrambi i problemi, sono reali e, nell'ottimo, grazie ai teoremi di dualità, si equivalgono. E' sensato pertanto immaginare che i costi siano espressi nella stessa unità di misura, in questo caso euro. Essendo il termine noto b un vettore che esprime quantità di elementi nutritivi, può essere una buona idea ipotizzare che gli elementi del vettore λ rappresentino prezzi per unità di elemento nutritivo e, in particolare, l'incognita λ_i rappresenti il prezzo associato all' i -esimo elemento nutritivo. L'interpretazione che si può dare del duale di questo problema di dieta può essere quella della massimizzazione del ricavo della vendita di elementi nutritivi allo stato puro (vitamine, carboidrati, ecc.): infatti, essendo b il livello minimo richiesto per una data dieta, l'acquirente razionale, potendo scegliere i singoli elementi nutritivi, e non essendo vincolato ad acquistarli "indirettamente", come elementi contenuti in proporzioni differenti all'interno dei diversi cibi, sarebbe soddisfatto dall'acquisto di quel quantitativo minimo. Quindi il produttore di elementi nutritivi allo stato puro, dovendo fissare i prezzi di vendita, può ipotizzare di conoscere esattamente la domanda di ognuno di essi, data dal vettore b . Il vincolo $\lambda^T A \leq c^T$ rappresenta il vincolo di competitività sul mercato. Infatti, considerando il j -esimo cibo, esso risulterà composto, mediamente, da elementi nutritivi in quantità espresse dal vettore A_j ed avrà costo unitario c_j . L'acquisto di una unità di cibo j , al prezzo c_j , porta dunque un contributo pari ad A_{1j} unità dell'elemento nutritivo 1, A_{2j} dell'elemento 2, ecc. I prezzi dei singoli componenti nutritivi dovranno apparire vantaggiosi al potenziale acquirente. Pertanto, ad esempio, se l'acquirente volesse confrontare il cibo j con il suo equivalente composto da elementi puri, dovrebbe poter osservare un vantaggio α , al massimo, un'equivalenza di prezzo. Da qui nasce il vincolo $\lambda_1 A_{1j} + \lambda_2 A_{2j} + \cdots + \lambda_m A_{mj} \leq c_j$.

8. DUALITÀ LINEARE

In questa coppia di problemi duali idealizzati, la dualità oppone compratore a produttore, minimizzazione del costo a massimizzazione del prezzo di vendita, vincoli di qualità a prezzi associati agli elementi nutritivi, costi dei cibi a vincoli di competitività di mercato. La dualità ci conferma che se entrambi i giocatori, produttore e consumatore, conoscono la strategia ottimale, ci sarà indifferenza tra le due scelte, cibi reali e componenti nutritivi, perfetto equilibrio tra l'interesse del compratore e quello del produttore. Naturalmente l'esempio è fortemente idealizzato e lontano da quanto avviene nei mercati reali, dove gli attori in gioco sono molti più di due, le regole della competizione molto meno elementari, gli interessi meno decisamente contrapposti. Tuttavia già da questo semplice schema si vede come in un gioco di tipo competitivo è fondamentale conoscere la propria strategia ottimale, senza la quale il gioco viene perduto: se il compratore non conosce la dieta ottimale, potrà vedere convenienza nell'acquisto del cibo sintetico; se il venditore non saprà fissare all'ottimo i prezzi, realizzerà margini di guadagno inferiori a quelli teoricamente possibili.

Anche la proprietà degli scarti complementari ha un risvolto economico: scelta una dieta ammissibile ed una politica di prezzi competitiva, queste saranno strategie ottimali per il compratore ed il produttore se e solo se per ogni cibo effettivamente presente in una dieta ottimale (variabile strettamente positiva) il produttore realizzerà un margine nullo (slack nullo nel vincolo di competitività) e per ogni elemento nutritivo sovrabbondante nella dieta (surplus positivo nel primale) il produttore avrà fissato prezzo nullo nel duale. Quindi gli elementi nutritivi sovrabbondanti nella dieta ottima hanno in realtà valore di mercato nullo e gli unici cibi sui quali il produttore può fissare prezzi competitivi sono quelli che l'acquirente non desidera comprare nella dieta ottima!

Se nel problema sono inclusi elementi nutritivi “dannosi” (ad esempio i grassi), nella formulazione del primale si troveranno vin-

8.4. Interpretazione della dualità

coli di \leq :

$$\begin{aligned} & \min c^T x \\ & Ax \geq b \\ & Cx \leq d \\ & x \geq 0. \end{aligned}$$

Il duale di questo problema ha la forma

$$\begin{aligned} & \max \lambda^T b + \mu^T d \\ & \lambda^T A + \mu^T C \leq c^T \\ & \lambda \geq 0 \\ & \mu \leq 0 \end{aligned}$$

Si vede che agli elementi dannosi viene attribuito un prezzo negativo o, equivalentemente, uno sconto sul prezzo complessivo. La proprietà degli scarti complementari in questo caso afferma che se la dieta ottimale contiene un quantitativo inferiore alla soglia massima di un elemento considerato dannoso, lo sconto ad esso associato non potrà che essere nullo. Viceversa si potrà attribuire uno sconto non nullo solo agli elementi dannosi poco costosi, a quelli cioè che nella dieta di minimo costo sono presenti nella quantità massima tollerabile.

Infine, in alcuni casi il modello della dieta prevede limitazioni sia inferiori che superiori sugli elementi nutritivi:

$$\begin{aligned} & \min c^T x \\ & b \leq Ax \leq d \\ & x \geq 0, \end{aligned}$$

8. DUALITÀ LINEARE

il cui duale ha la forma

$$\begin{aligned} \max & \lambda^T b + \mu^T d \\ (\lambda + \mu)^T A & \leq c^T \\ \lambda & \geq 0 \\ \mu & \leq 0. \end{aligned}$$

Si vede facilmente che nel caso in cui $b_i < d_i$, grazie alla complementarietà, in una soluzione ottimale al più uno dei due prezzi λ_i, μ_i può essere non nullo. Quindi si può interpretare la quantità $\lambda + \mu$ come il vettore di prezzi effettivi di ciascun elemento nutritivo, e si vede che il prezzo ottimale potrà essere positivo per gli elementi nutritivi che nella dieta ottimale risultano pari al minimo richiesto, e negativo per quelli pari al massimo.

Duale del problema di miscelazione ottimale

Il modello di miscelazione (ad esempio di rottami o di olii) ha una struttura che può essere rappresentata come segue:

$$\begin{aligned} \min & \sum_j c_j x_j \\ \sum_i A_{ij} x_j & = b_i \end{aligned} \tag{8.4}$$

$$\sum_j x_j = 1 \tag{8.5}$$

$$x_j \geq 0 \tag{8.6}$$

dove c_j sono i costi unitari dei rottami, A_{ij} rappresenta la percentuale di elemento chimico i contenuta in un'unità di rottame j , e la quantità totale prodotta è pari ad un'unità – si veda il capitolo 2.1 per la formulazione esplicita di questo modello.

8.4. Interpretazione della dualità

Per l'interpretazione di questo modello è utile riferirsi ad un esempio concreto. Si consideri l'esempio di miscelazione i cui dati sono riportati nel file seguente:

```
set MATERIALI := A B C D E F G H I;
set ELEMENTI := Pb Zn Sn;

param costo :=
A 4.1 B 4.3 C 5.8 D 6.0 E 7.6
F 7.5 G 7.3 H 6.9 I 7.3
;

param analisi_rich := Pb 40
                      Zn 30
                      Sn 30;

param compos:
      A   B   C   D   E   F   G   H   I  :=
Pb    10   10   40   60   30   30   30   50   20
Zn    10   30   50   30   30   40   20   40   30
Sn    80   60   10   10   40   30   50   10   50;
```

Il problema duale, in forma esplicita, ha la seguente forma:

$$\max 40\lambda_{\text{Pb}} + 30\lambda_{\text{Zn}} + 30\lambda_{\text{Sn}}$$

$$10\lambda_{\text{Pb}} + 10\lambda_{\text{Zn}} + 80\lambda_{\text{Sn}} \leq 4.1 \quad (A)$$

$$10\lambda_{\text{Pb}} + 30\lambda_{\text{Zn}} + 60\lambda_{\text{Sn}} \leq 4.3 \quad (B)$$

...

$$20\lambda_{\text{Pb}} + 30\lambda_{\text{Zn}} + 50\lambda_{\text{Sn}} \leq 7.3 \quad (I)$$

In questo problema, il vincolo $\sum_j x_j = 1$ è stato rimosso, in quanto implicato dal fatto che il problema contiene un'analisi chimica completa dei materiali; infatti, sommando i vincoli relativi alle richieste di elementi chimici si ottiene un vincolo equivalente a quello

8. DUALITÀ LINEARE

rimosso. Questo problema ammette un'interpretazione di tipo economico. Innanzitutto è opportuno osservare che il problema primale, di miscelazione di metalli, richiede la minimizzazione del costo della miscela. I valori assunti dall'obiettivo sono quindi espressi, ad esempio, in Euro. Dal teorema forte di dualità si sa che il valore ottimale del problema duale è identico al valore ottimale del problema primale; è ragionevole pertanto ipotizzare che l'unità di misura dell'obiettivo nel problema duale sia ancora espressa in Euro. Poiché i coefficienti dell'obiettivo del duale sono numeri puri (percentuali di elementi chimici) si deduce che le incognite λ rappresentano valori monetari. Si può interpretare il modello duale come segue: un ipotetico produttore di metalli puri desidera fissare i prezzi di vendita di Piombo, Stagno e Zinco. Sapendo che l'acquirente ha fissato in 40, 30 e 30 le percentuali richieste in una miscela di costo minimo di tali metalli, il produttore ritiene ragionevole prevedere di poter vendere esattamente 40 parti (su 100) di Piombo, 30 di Stagno e 30 di Zinco. L'obiettivo corrisponde quindi alla massimizzazione del ricavo totale del venditore, nel caso in cui abbia fissato il prezzo di vendita dei tre metalli ai valori $\lambda_{\text{Pb}}, \lambda_{\text{Zn}}, \lambda_{\text{Sn}}$.

Il membro di sinistra del primo vincolo

$$10\lambda_{\text{Pb}} + 10\lambda_{\text{Zn}} + 80\lambda_{\text{Sn}} \leq 4.1$$

rappresenta il prezzo di vendita di una miscela al 10% di Piombo, 10% di Zinco, 80% di Stagno: poiché 10, 10, 80 sono i coefficienti del materiale A, si deduce che il membro di sinistra del primo vincolo rappresenta il prezzo al quale il produttore di metalli venderebbe una miscela perfettamente equivalente al materiale A. Il vincolo è interpretabile quindi come un vincolo di concorrenza: il produttore fissa il prezzo dei metalli in modo tale da rendere competitivo (se i prezzi sono tali che nel vincolo valga la relazione di minore stretto) o al più equivalente (se vale l'eguaglianza) l'acquisto di metalli puri rispetto all'acquisto di un'equivalente quantitativo di rottame di tipo A. I vincoli successivi impongono prezzi competitivi per tutti gli

8.4. Interpretazione della dualità

altri rottami. Quindi il duale del problema di miscelazione non è altro che il problema che si potrebbe porre un produttore di elementi chimici puri che desiderasse fissare i prezzi dei singoli elementi in modo tale da massimizzare il proprio ricavo, con vincoli che corrispondono alla necessità di fissare prezzi tali da rendere competitivo l'acquisto di metalli puri rispetto ai rottami equivalenti disponibili sul mercato. Quindi il produttore cerca di convincere l'acquirente di miscele a servirsi dei propri metalli puri, fissando prezzi apparentemente concorrenziali con quelli di mercato. In realtà la teoria della dualità garantisce che il costo della miscela ottimale è esattamente pari al ricavo massimo del produttore: esiste equilibrio perfetto tra domanda e offerta, naturalmente a patto che entrambi i giocatori, il minimizzatore ed il massimizzatore, conoscano la strategia ottimale. Riassumendo, nel problema duale del problema della miscela di costo minimo, ogni variabile del duale è associata ad un elemento chimico (vincolo) e ne rappresenta il prezzo di vendita. Ogni vincolo del duale è associato ad un materiale, ed impone la competitività del prezzo di vendita rispetto al costo del materiale.

Si può osservare che dalla proprietà degli scarti complementari vista precedentemente scende che se nella miscela ottimale un certo materiale compare effettivamente (la variabile corrispondente è diversa da zero), lo slack del vincolo corrispondente nel duale deve necessariamente essere nullo; ma poiché lo slack, o scarto, nei vincoli del duale corrisponde al margine di profitto del produttore, cioè alla differenza tra il prezzo imposto ed il prezzo di mercato, i teoremi di dualità implicano che nell'ottimo, e solo nell'ottimo, per ogni materiale il cui acquisto è conveniente, il produttore deve fissare il prezzo in modo tale che il margine di profitto sia nullo. Risulta quindi abbastanza chiaro che gli obiettivi ottimali saranno identici: il produttore ha la possibilità di fissare prezzi apparentemente concorrenziali solo per i materiali che l'acquirente di miscele non acquisterebbe; per gli altri materiali, il prezzo fissato dal produttore è pari a quello di mercato.

8. DUALITÀ LINEARE

Ritornando al modello generale di miscelazione, con il vincolo relativo alla quantità totale prodotta, si vede abbastanza facilmente che il duale, indicando con λ' le variabili del duale associate ai vincoli sugli elementi chimici e con μ la variabile associata al vincolo $\sum_j x_j = 1$, risulta essere il seguente:

$$\max \sum_i b_i \lambda'_i + \mu \quad (8.7)$$

$$\sum_i A_{ij} \lambda'_i + \mu \leq c_j \quad \forall j \quad (8.8)$$

Questo modello è abbastanza simile al duale del problema della dieta e si può quindi interpretare in modo simile. L'unica differenza apparente consiste nella variabile μ : analizzando gli obiettivi, e sapendo che il valore dell'ottimo di questi due problemi duali deve essere identico, essendo duali di problemi di miscelazione equivalenti fra loro, si deduce che, nell'ottimo, il valore di μ sarà legato al prezzo complessivo di vendita degli elementi chimici per i quali non è stato fissata una richiesta, cioè alla rimanenza. Tuttavia è bene notare che il significato ed il valore numerico delle variabili λ e λ' variano. Per interpretare correttamente il legame tra i due problemi duali e per intuire un'interpretazione economica del modello (8.7-8.8) occorre procedere in modo più formale. Si consideri il problema originale, nel quale l'ultimo vincolo viene espresso in modo tale che risulti esplicitamente il fatto che la somma per righe della matrice A e del vettore b è pari a 1:

$$\begin{aligned} & \min \sum_j c_j x_j \\ & \sum_j A_{ij} x_j = b_i \quad i = 1, \dots, m-1 \\ & \sum_j (1 - \sum_{i=1}^{m-1} A_{ij}) x_j = 1 - \sum_{i=1}^{m-1} b_i \\ & x_j \geq 0 \quad \forall j \end{aligned}$$

8.4. Interpretazione della dualità

Il duale, come si è visto, ha la forma

$$\begin{aligned} \max & \sum_{i=1}^{m-1} b_i \lambda_i + (1 - \sum_{i=1}^{m-1} b_i) \lambda_m \\ & \sum_{i=1}^{m-1} A_{ij} \lambda_i + (1 - \sum_{i=1}^{m-1} A_{ij}) \lambda_m \leq c_j \quad \forall j \end{aligned}$$

da cui, raccogliendo, si ottiene

$$\begin{aligned} \max & \sum_{i=1}^{m-1} b_i (\lambda_i - \lambda_m) + \lambda_m \\ & \sum_{i=1}^{m-1} A_{ij} (\lambda_i - \lambda_m) + \lambda_m \leq c_j \quad \forall j. \end{aligned}$$

Da qui, ponendo

$$\lambda'_i := \lambda_i - \lambda_m, \quad \mu := \lambda_m$$

si ottiene la formulazione del duale (8.7-8.8). Quindi, in quest'ultimo, le variabili hanno il seguente significato: μ è pari al prezzo di un'unità di rimanenza, mentre il prezzo di ogni altro componente chimico è diminuito del prezzo della rimanenza stessa. Quindi i valori che si determinano nel duale, associati ai singoli elementi chimici, non rappresentano il prezzo di vendita di tali elementi, ma il prezzo diminuito del prezzo di vendita della rimanenza. Si vede abbastanza facilmente come anche in questo caso i vincoli del duale non siano altro che limitazioni al prezzo di vendita di composti equivalenti a quelli disponibili sul mercato.

Duale del problema di “mix” di produzione

Il modello del mix di produzione rappresenta un’idealizzazione del problema di decisione consistente nel dover pianificare come allocare risorse produttive (materie prime, ore di manodopera, capacità

8. DUALITÀ LINEARE

produttiva di macchine utensili, ecc.) per poter produrre manufatti diversi, che richiedono quantità differenti di risorse. L'idealizzazione consiste sia nell'ipotesi di linearità sia nell'ipotesi che le quantità prodotte siano variabili indipendenti: si ritiene cioè che il mercato sia in grado di assorbire qualunque quantitativo si decida di produrre. Spesso questo modello viene utilizzato ipotizzando una domanda di prodotto, stimata dagli esperti di analisi di mercato, e la decisione non è tanto *cosa*, ma *come* produrre, nel senso che lo stesso manufatto può essere prodotto in modi diversi, utilizzando macchinari diversi e sfruttando in modo differente, e con profitti differenti, la capacità produttiva.

Indicati con p_1, p_2, \dots, p_n i profitti unitari relativi ai prodotti (od ai modi di produzione) $1, 2, \dots, n$, con $\ell_1, \ell_2, \dots, \ell_m$ i livelli massimi di utilizzo di m risorse (ad esempio ore di lavoro su m macchine), con A_{ij} il consumo per unità di prodotto i relativo alla risorsa j , il problema viene formulato come

$$\begin{aligned} & \max p^T x \\ & Ax \leq \ell \\ & x \geq 0 \end{aligned}$$

il cui duale risulta equivalente al problema

$$\begin{aligned} & \min y^T \ell \\ & y^T A \geq p^T \\ & y \geq 0 \end{aligned}$$

Un'interpretazione del problema duale può essere la seguente: si immagini che il responsabile della produzione stia valutando la possibilità di utilizzare la manodopera e le altre risorse disponibili per effettuare prestazioni all'esterno della fabbrica; si pensa cioè di vendere ore di manodopera ed altre risorse, sottraendole alla produzione. Immaginando di destinare tutte le risorse a questa attività

8.4. Interpretazione della dualità

di prestazione esterna, indicando con y_i il prezzo associato all'affitto di un'unità della risorsa i (ad esempio un'ora/uomo), il profitto totale sarà dato da $\ell^T y$. L'obiettivo di minimizzazione è coerente con il desiderio di un acquirente imporre prezzi competitivi rispetto a quelli di mercato. L'acquirente cerca cioè di determinare i prezzi da associare alle risorse disponibili che rendano il costo totale minimo. Naturalmente l'imprenditore vorrebbe realizzare un profitto da questa riconversione. Per garantirsi contro eventuali perdite rispetto alla situazione attuale, l'acquirente di risorse ragiona nel modo seguente: per ogni unità di prodotto j non realizzata, si liberano risorse in quantità pari a $a_{1j}, a_{2j}, \dots, a_{mj}$. I prezzi unitari di acquisto y_1, y_2, \dots, y_m devono essere tali che il ricavo corrispondente al loro impiego esternamente alla fabbrica deve essere almeno pari al profitto mancato a causa della decisione di non produrre un'unità di j .

In formule

$$a_{1j}y_1 + a_{2j}y_2 + \cdots + a_{mj}y_m \geq p_j.$$

Da qui scende la forma del problema duale sopra indicata.

Esempio 23. Si consideri l'esempio seguente: nel quale alcuni prodotti per essere fabbricati devono passare attraverso 5 fasi di lavorazione e richiedono tempi macchina differenti:

Reparti	Prodotti							Disponibilità
	1	2	3	4	5	6	7	
assemblaggio	2	2	4	7	9	15	18	1700
falegnameria	7	5	9	10	10	12	15	2000
imballaggio	1	1	1	1	2	1	0	500
verifica	1	1	1	2	1	2	2	300
verniciatura	2	2	2	3	3	3	3	1500
Profitto	10	18	20	25	27	28	35	

Eseguendo il simplex con questi dati si ottiene la soluzione ottimale consistente nella produzione di 200 unità del prodotto 2 e 100 del prodotto 5, con un profitto totale pari a 6300. La soluzione

8. DUALITÀ LINEARE

ottimale del duale assegna valore nullo a tutte le variabili tranne che a quella associata alla falegnameria (1.8) e quella associata alla verifica (9). Infatti è facilmente verificabile che in tutti gli altri reparti esiste una capacità residua non sfruttata: la relazione degli scarti complementari impone che nell'ottimo le variabili duali associate siano nulle. Per quanto riguarda l'interpretazione economica del problema duale (di minimo), si osservi che esso consiste nell'assegnazione di un prezzo di vendita alle ore disponibili nei vari reparti. Risulta quindi dalla dualità che il prezzo più competitivo da assegnare alle ore disponibili nei vari reparti è nullo per tutte le lavorazioni, con l'eccezione dei reparti falegnameria e verifica. In altre parole, non sarà conveniente affittare all'esterno le ore di manodopera dei reparti assemblaggio, imballaggio e verniciatura. Per quanto riguarda la soluzione ottima del duale, si ricava, direttamente per sostituzione, che il profitto ricavabile dalla vendita delle ore di manodopera corrispondenti ai diversi prodotti è dato da

prod 1	prod 2	prod 3	prod 4	prod 5	prod 6	prod 7
21.6	18.0	25.2	36.0	27.0	39.6	45.0

Confrontando questi profitti unitari con quelli ricavabili dalla produzione, sembra di poter dedurre che l'affitto di ore di manodopera è conveniente: in realtà l'impressione è falsa. Infatti gli unici prodotti per i quali si ricaverrebbe un profitto maggiore sono in realtà costituiti da lotti che non conviene produrre (nella soluzione ottimale la produzione associata ad essi è nulla). Per i prodotti 2 e 5, che effettivamente compaiono a livello positivo nella produzione ottimale, il margine di profitto si riduce a zero. Ovviamente tutto ciò è ben prevedibile dalla teoria della dualità: i valori ottimali del primale e del duale coincidono. Il fatto che un profitto strettamente maggiore non possa essere ottenuto per i prodotti che compaiono a livello non nullo nella soluzione ottimale del problema del mix di produzione è una banale conseguenza del teorema degli scarti complementari.

8.5 Dualità e teoria dei giochi

La nascita della teoria della dualità lineare (Dantzig & Thapa, 2003) è collegata ai primi sviluppi elementari della *teoria dei giochi* non cooperativi (Morgenstern & Neumann, 1944). Il legame verrà illustrato analizzando un semplicissimo e classico gioco a due persone, la “morra cinese”. I due giocatori, contemporaneamente, scelgono una tra tre mosse possibili: S (Sasso), C (Carta), F (Forbici). Nel confronto, il Sasso batte le Forbici, ma è battuto dalla Carta. Le Forbici battono la Carta; scelte identiche fra i due giocatori portano al pareggio. E’ abbastanza intuitivo (e facilmente dimostrabile) che la strategia ottimale per entrambi i giocatori (immaginando che ciascuno di essi desideri massimizzare la differenza tra il numero vincite e di perdite) sia quella di “scegliere a caso” tra le tre mosse. Questa strategia equivale ad estrarre una delle tre mosse in base ad una legge che attribuisca probabilità 1/3 a ciascuna di esse. La situazione cambia leggermente nel caso in cui le vincite e le perdite siano differenziate. Si consideri ad esempio la seguente matrice di *payoff*:

		carta	forbici	sasso
A =	carta	0	1	-2
	forbici	-3	0	4
	sasso	6	-5	0

Dal segno degli elementi della matrice si deduce che essi rappresentano il guadagno (perdita nel caso negativo) del giocatore associato alle colonne. Il giocatore che sceglie la riga subirà una perdita (guadagno se negativa) pari al valore corrispondente nella matrice. Cioè, se il giocatore R (Righe) scegliesse Carta ed il giocatore C (Colonne) scegliesse Forbici, il giocatore R perderebbe 1 ed il giocatore Colonne guadagnerebbe 1.

Una *strategia*, per un qualunque giocatore, è formalizzabile come un vettore (a tre componenti in questo caso) di numeri non negativi

8. DUALITÀ LINEARE

la cui somma è pari ad 1, cioè ad una densità discreta di probabilità. In altre parole, associando la componente 1 a Carta, la numero 2 a Forbici, la 3 a Sasso, il giocatore Righe dovrà scegliere $x \in \mathbb{R}^3$:

$$\begin{aligned}x_1 + x_2 + x_3 &= 1 \\x_1, x_2, x_3 &\geq 0\end{aligned}$$

Una *strategia deterministica* equivale alla scelta di x con componenti binarie. Analogamente il giocatore Colonne sceglierà la propria strategia, qui indicata con un vettore $y \in \mathbb{R}^3$:

$$\begin{aligned}y_1 + y_2 + y_3 &= 1 \\y_1, y_2, y_3 &\geq 0\end{aligned}$$

Si ipotizzi, per il momento, che il giocatore Colonne conosca la strategia \bar{x} , utilizzata dal giocatore Righe. In base a questa ipotesi, se il giocatore Colonne scegliesse di giocare “Carta”, il suo guadagno atteso sarebbe

$$0\bar{x}_1 - 3\bar{x}_2 + 6\bar{x}_3$$

In generale, in funzione della strategia y utilizzata, il guadagno atteso varrà

$$v_C^* = \sum_{i=1}^3 \sum_{j=1}^3 a_{ij} \bar{x}_i y_j$$

dove con a_{ij} si è indicato il generico elemento della matrice di payoff. Conoscendo la strategia del giocatore Righe, quindi, il giocatore Colonne cercherà di ottimizzare il proprio guadagno risolvendo il problema di programmazione lineare:

$$\begin{aligned}\max_y (\bar{x}^T A) y \\1^T y = 1 \\y \geq 0\end{aligned}$$

8.5. Dualità e teoria dei giochi

Questo è un problema di *PL* standard, a parte il verso dell’ottimizzazione, con un solo vincolo di egualanza. Le soluzioni di base, tra cui cercare quella ottima, sono costituite da un’unica variabile e , pertanto, le soluzioni ottimali sono *strategie pure*, cioè deterministiche, non randomizzate. La strategia ottimale del giocatore risulta quindi essere

$$\arg \max \{ \bar{x}^T A_1, \bar{x}^T A_2, \bar{x}^T A_3 \}$$

A questo punto si analizzi il comportamento del giocatore Righe. Immaginando una politica difensiva (nel gergo della teoria dei giochi si direbbe “conservativa”), il giocatore può cercare di ottimizzare il proprio obiettivo nell’ipotesi peggiore possibile, cioè nell’ipotesi che l’altro giocatore abbia scoperto la strategia di gioco utilizzata. Quindi il giocatore Righe, in questo contesto, cercherà di minimizzare le proprie perdite nel caso peggiore possibile:

$$\begin{aligned} \min_x \max & \{ x^T A_1, x^T A_2, x^T A_3 \} \\ & 1^T x = 1 \\ & x \geq 0 \end{aligned}$$

Questo non è un problema di *PL*, ma si può vedere facilmente che risulta equivalente al seguente problema di *PL*, grazie ai risultati di equivalenza presentati nella sezione 3.3:

$$\begin{aligned} z_R^* &= \min_{x \in \mathbb{R}^3, z \in \mathbb{R}} z \\ z &\geq x^T A_j & j = 1, 2, 3 \\ 1^T x &= 1 \\ x &\geq 0 \end{aligned}$$

La strategia risultante è ottimale nel senso minimax (cioè minimizza la massima perdita possibile). Se si seguisse un ragionamento analogo per il giocatore Colonne (ipotizzando che il giocatore

8. DUALITÀ LINEARE

Righe conosca la strategia dell'altro e cercando di massimizzare il minimo guadagno), si arriverebbe al modello seguente:

$$\begin{aligned} z_C^* &= \max_{y \in \mathbb{R}^3, w \in \mathbb{R}} w \\ w &\geq a_i^T y && i = 1, 2, 3 \\ 1^T y &= 1 \\ y &\geq 0 \end{aligned}$$

Si vede facilmente che i due problemi formano una coppia di problemi duali. Quindi, per i teoremi di dualità si deduce che essi avranno valore ottimale coincidente (una soluzione ottimale esiste sicuramente per entrambi i problemi in quanto entrambi hanno un insieme ammissibile non vuoto). Quindi si può definire un unico *valore del gioco* pari al valore ottimale v^* comune ai due problemi duali. In base alla convenzione adottata, se il valore del gioco è nullo, il gioco è equilibrato, se è positivo, il giocatore colonne avrà un vantaggio e viceversa se negativo. Con i dati dell'esempio considerato, la strategia ottimale dei due giocatori ed il valore ottimale del gioco risultano pari a

$$\begin{aligned} x^T &= [0,59 \quad 0,28 \quad 0,13] \\ y^T &= [0,34 \quad 0,42 \quad 0,24] \\ v^* &= -0,06 \end{aligned}$$

La Figura 8.1 riporta il grafico delle perdite del giocatore Righe in una simulazione di una sequenza di giochi nel corso della quale ciascun giocatore adotta la strategia appena vista. Come si nota dalla figura, anche se nel transitorio il giocatore Righe subisce delle perdite (positive), a lungo andare i suoi guadagni cumulati tendono ad avvicinarsi alla retta con pendenza $-0,06$, pari al valore del gioco, riportata sul grafico della figura.

8.5. Dualità e teoria dei giochi



Figura 8.1: Simulazione di gioco

Il metodo del simplesso duale

Lo studio della teoria della dualità permette di vedere con ottica differente il metodo del simplesso e di interpretarne i passi principali in modo nuovo. Si osserva anche che in molti testi la teoria della dualità viene presentata prima dell'algoritmo del simplesso, ed è utilizzata per giustificare lo sviluppo del metodo.

Il metodo del simplesso può essere visto come un algoritmo che, iniziando da una soluzione di base ammissibile per il problema, genera una sequenza finita di basi ammissibili fino a determinare l'ottimo, o a dimostrare che il problema è illimitato. Si consideri un problema di *PL* standard

$$\begin{aligned} \min c^T x \\ Ax = b \\ x \geq 0. \end{aligned}$$

e sia A_B una qualsiasi base (ammissibile o meno). E' sempre possibile associare a questa base una soluzione, non necessariamente ammissibile, per il problema di *PL*:

$$\bar{x}_B = A_B^{-1} b.$$

In modo del tutto analogo, si può associare alla stessa base un vettore

$$\bar{\lambda}_B^T = c_B^T A_B^{-1}$$

9. IL METODO DEL SIMPLEXO DUALE

compatibile con le soluzioni ammissibili del duale. La base A_B si dirà *ammissibile* (per il primale) o \mathcal{P} -ammissibile se

$$\bar{x}_B \geq 0;$$

si dirà ammissibile per il duale o \mathcal{D} -ammissibile se

$$\bar{\lambda}_B^T A_N \leq c_N^T$$

(sarebbe $\bar{\lambda}_B^T A \leq c^T$, ma per le componenti associate a B la relazione è banalmente verificata).

Si può vedere immediatamente che la \mathcal{D} -ammissibilità di una base corrisponde alla non-negatività dei coefficienti di costo ridotto. Con queste definizioni, l'algoritmo del simplex si può descrivere nel seguente modo:

1. Sia B una base ammissibile per il problema primale;
2. Se B è ammissibile anche per il duale, allora è ottimale;
3. Altrimenti effettuare un'operazione di pivot che modifichi la base in modo tale che:
 - a) la nuova base sia ancora ammissibile per il primale
 - b) il costo non aumenti

Si noti che nel corso di tutte le iterazioni del metodo del simplex, viene mantenuta l'identità dei costi primale/duale:

$$c^T x = c_B^T x_B = c_B^T A_B^{-1} b = \lambda_B^T b$$

e che, pertanto, grazie al teorema forte di dualità l'ottimalità si avrà non appena anche la soluzione corrente del duale sarà ammissibile. Il metodo del simplex quindi genera coppie di soluzioni primali/duali con ugual costo con la garanzia che, delle due, la soluzione

9.1. Il metodo del simplesso duale: formulazione matriciale

di base del primale resti sempre ammissibile; esse saranno ottimali se e solo se sono ammissibili per i rispettivi problemi.

In modo totalmente simmetrico, è possibile costruire un algoritmo di tipo duale, basato cioè sul seguente schema:

1. Sia B una base ammissibile per il problema duale;
2. Se B è ammissibile anche per il primale, allora è ottimale;
3. Altrimenti effettuare un'operazione di pivot che modifichi la base in modo tale che:
 - a) la nuova base sia ancora ammissibile per il duale
 - b) il costo non diminuisca

Un algoritmo di questo tipo, detto algoritmo del simplesso duale, produce una sequenza di soluzioni di base in genere non ammissibili per il problema da risolvere, ma sempre ammissibili per il duale. La richiesta di non diminuzione del costo da un'iterazione all'altra nasce dal fatto che se la base \mathcal{D} -ammissibile corrente non è ottimale, in virtù del teorema debole di dualità occorrerà cercare un'altra soluzione di base di costo non inferiore.

9.1 Il metodo del simplesso duale: formulazione matriciale

Si consideri una base A_B ammissibile per il duale di un problema di *PL* standard; posto $\lambda_B^T = c_B^T A_B^{-1}$ sia

$$\lambda_B^T A \leq c^T$$

o, equivalentemente,

$$c_B^T A_B^{-1} A_N \leq c_N^T$$

9. IL METODO DEL SIMPLEXO DUALE

ovvero

$$\hat{c}^T \geq 0.$$

Naturalmente se la base A_B risultasse anche ammissibile per il primale:

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

sarebbe ottimale. Si ipotizzi pertanto che almeno per un indice i si abbia

$$e_i^T A_B^{-1}b < 0.$$

La soluzione corrente del duale soddisfa esattamente i vincoli associati alle variabili di base:

$$\lambda_B^T A_B = c_B^T$$

e, grazie alla relazione degli scarti complementari, è per questo motivo che alcune variabili di base del primale possono essere non nulle. Volendo pertanto generare una nuova soluzione ammissibile del duale che riduca l'intervallo (gap) di dualità si potrà cercare una nuova soluzione $\bar{\lambda}$ che soddisfi

$$\bar{\lambda}^T A_B \leq c_B^T$$

per la quale la diseguaglianza valga strettamente in corrispondenza della i -esima variabile di base del primale: in questo modo, forzando, se possibile, l' i -esimo vincolo del duale ad avere slack non nullo, si obbligherà la variabile corrispondente del primale a valere 0. Si cerca pertanto una modifica della soluzione corrente secondo lo schema

$$\bar{\lambda} = \lambda_B + d$$

in modo tale che

$$\bar{\lambda}^T A_B = c_B^T - \alpha e_i^T$$

con $\alpha \geq 0$ (possibilmente $\alpha > 0$) in modo tale da rendere se possibile non attivo l' i -esimo vincolo del duale. Sostituendo, si ottiene

$$\lambda_B^T A_B + d^T A_B = c_B^T - \alpha e_i^T$$

9.1. Il metodo del simplex duale: formulazione matriciale

da cui

$$d^T = -\alpha e_i^T A_B^{-1}. \quad (9.1)$$

Quindi per cercare di forzare a 0 la i -esima variabile negativa nella soluzione di base del primale occorre cercare un passo α positivo lungo la direzione (9.1). Le condizioni alle quali occorre ricondursi sono, come si è visto in precedenza, la non diminuzione del costo della soluzione corrente del duale ed il mantenimento della \mathcal{D} -ammissibilità. Per quanto riguarda il costo, si ha

$$\bar{\lambda}^T b = \lambda_B^T b - \alpha e_i^T A_B^{-1} b;$$

essendo $\alpha \geq 0$ e $e_i^T A_B^{-1} b < 0$, la non diminuzione è garantita; per il soddisfacimento dei vincoli del duale si ha invece che quelli corrispondenti alle colonne di A_B sono soddisfatti per costruzione; per gli altri occorre invece scegliere α in modo tale da garantire che

$$\bar{\lambda}^T A_j \leq c_j \quad \forall j \in N$$

e cioè

$$(\lambda_B^T - \alpha e_i^T A_B^{-1}) A_j \leq c_j \quad \forall j \in N;$$

se

$$e_i^T A_B^{-1} A_j \geq 0$$

l'ammissibilità del vincolo j è garantita per qualunque scelta di α . Se invece

$$e_i^T A_B^{-1} A_j < 0$$

occorre scegliere α in modo tale che

$$0 \leq \alpha \leq -\frac{c_j - \lambda_B^T A_j}{e_i^T A_B^{-1} A_j}.$$

Da qui scende pertanto il criterio per la scelta del passo nel metodo del simplex duale:

$$\alpha = \min_j \left\{ -\frac{\hat{c}_j}{e_i^T A_B^{-1} A_j} : e_i^T A_B^{-1} A_j < 0 \right\}. \quad (9.2)$$

9. IL METODO DEL SIMPLEXO DUALE

Naturalmente se per nessun indice j vale $e_i^T A_B^{-1} A_j < 0$, il passo α potrà essere scelto arbitrariamente grande, con conseguente aumento del costo della soluzione del duale. Il problema duale pertanto risulta illimitato superiormente e, di conseguenza, il primale non ammissibile. Nel caso in cui il passo α nella (9.2) risulti finito e strettamente positivo, almeno un vincolo non attivo nella soluzione λ_B diventerà attivo. Sarà a questo punto possibile effettuare un'operazione di pivot, utilizzando esattamente la stessa tecnica già vista per il simplexso primale, per far entrare in base la colonna di A corrispondente a tale vincolo al posto di quella associata all' i -esima componente di base.

Riassumendo, quindi, il metodo del simplexso duale parte da una base ammissibile per il duale e non per il primale; individua una variabile uscente di base associata ad una componente negativa della soluzione di base del primale; determina poi la variabile entrante in base all'osservazione del primo vincolo del duale che diviene attivo effettuando uno spostamento lungo la direzione $-e_i^T A_B^{-1}$. Se nessun vincolo del duale diviene attivo lungo tale direzione, il problema (primale) risulta inammissibile. Altrimenti viene effettuata un'operazione di pivot e la base corrente viene aggiornata. Come nel metodo primale, anche in quello duale è facile dimostrare la finitezza nel caso in cui ad ogni iterazione il passo α sia strettamente positivo: infatti in questo caso il costo associato alla soluzione di base corrente del duale aumenta strettamente e, grazie alla finitezza delle basi, certamente l'algoritmo termina in un numero finito di operazioni di pivot. Passi degeneri $\alpha = 0$ si possono tuttavia presentare nel caso in cui qualche coefficiente di costo ridotto sia nullo; tale situazione è duale rispetto a quella delle soluzioni degeneri del primale. E' possibile però ottenere un algoritmo del simplexso duale a terminazione garantita utilizzando tecniche anticiclaggio, del tipo di quella di Bland. E' anche possibile sviluppare un metodo prima fase per l'inizializzazione del simplexso duale, un metodo cioè in grado di procurare, se esiste,

9.1. Il metodo del simplesso duale: formulazione matriciale

una soluzione di base ammissibile per il duale. Questo metodo non verrà sviluppato in questa sede – può essere un interessante e non banale esercizio per il lettore. Al di là del desiderio di non appesantire troppo questa parte della trattazione, la scelta di non esaminare troppo a fondo questi aspetti del metodo del simplesso duale sta anche nel fatto che il principale campo di applicazione del metodo non è quello della risoluzione “da zero” di un problema di *PL*, quanto quello della ri-ottimizzazione di un problema, già risolto con altre tecniche, nel quale a seguito di una variazione del termine noto o dell’aggiunta di uno o più vincoli, si sia persa l’ammissibilità del primale senza aver compromesso quella del duale. Su questo utilizzo del metodo duale si tornerà più avanti, parlando di analisi di sensitività e di ottimizzazione lineare intera.

Esempio 24. Si consideri il problema

$$\begin{aligned} \min & \begin{bmatrix} -1 & 1 & 0 & 2 & 1 \end{bmatrix} x \\ & \begin{bmatrix} -1 & 0 & -1 & 1 & 1 \\ 0 & 2 & 1 & 1 & -2 \end{bmatrix} x = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \\ & x \geq 0 \end{aligned}$$

La base formata dalle prime due colonne non è ammissibile per il primale; tuttavia la soluzione di base del duale

$$\lambda_B^T = \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & \frac{1}{2} \end{bmatrix}$$

è ammissibile per il duale, come si vede facilmente per sostituzione; il costo associato a tale soluzione è 2. In corrispondenza della base $B = \{1,2\}$ le componenti del primale valgono -1 e 1 rispettivamente. Per effettuare un passo di pivot con il metodo del simplesso duale occorre pertanto effettuare una modifica alla soluzione

9. IL METODO DEL SIMPLEXO DUALE

corrente del tipo

$$\begin{aligned}\bar{\lambda}^T &= \left[\begin{array}{cc} 1 & \frac{1}{2} \end{array} \right] - \alpha \left[\begin{array}{cc} 1 & 0 \end{array} \right] \left[\begin{array}{cc} -1 & 0 \\ 0 & 2 \end{array} \right]^{-1} \\ &= \left[\begin{array}{cc} 1 & \frac{1}{2} \end{array} \right] - \alpha \left[\begin{array}{cc} -1 & 0 \end{array} \right] \\ &= \left[\begin{array}{cc} 1 + \alpha & \frac{1}{2} \end{array} \right].\end{aligned}$$

La scelta di $\alpha \geq 0$ deve tener conto dei vincoli del duale:

$$\left[\begin{array}{cc} 1 + \alpha & \frac{1}{2} \end{array} \right] \left[\begin{array}{ccc} -1 & 1 & 1 \\ 1 & 1 & -2 \end{array} \right] \leq \left[\begin{array}{ccc} 0 & 2 & 1 \end{array} \right]$$

o, equivalentemente, utilizzando la formula (9.2),

$$\begin{aligned}e_1^T A_B^{-1} A_N &= \left[\begin{array}{cc} 1 & 0 \end{array} \right] \left[\begin{array}{cc} -1 & 0 \\ 0 & 2 \end{array} \right]^{-1} \left[\begin{array}{ccc} -1 & 1 & 1 \\ 1 & 1 & -2 \end{array} \right] \\ &= \left[\begin{array}{ccc} 1 & -1 & -1 \end{array} \right] \\ \hat{c}^T &= \left[\begin{array}{ccc} 0 & 2 & 1 \end{array} \right] - \left[\begin{array}{cc} 1 & \frac{1}{2} \end{array} \right] \left[\begin{array}{ccc} -1 & 1 & 1 \\ 1 & 1 & -2 \end{array} \right] \\ &= \left[\begin{array}{ccc} \frac{1}{2} & \frac{1}{2} & 1 \end{array} \right]\end{aligned}$$

da cui

$$\alpha = \min \left\{ -\frac{\frac{1}{2}}{-1}, -\frac{1}{-1} \right\} = \frac{1}{2};$$

la nuova soluzione ammissibile del duale sarà quindi $\left[\begin{array}{cc} \frac{3}{2} & \frac{1}{2} \end{array} \right]$, con costo pari a $\frac{5}{2}$. La variabile x_1 esce di base, mentre entra in base la variabile che ha causato il passo α e cioè x_4 . La soluzione di base corrispondente a $B = \{2, 4\}$ è data da

$$\begin{aligned}\bar{x}_B &= \left[\begin{array}{cc} 0 & 1 \\ 2 & 1 \end{array} \right]^{-1} \left[\begin{array}{c} 1 \\ 2 \end{array} \right] \\ &= \left[\begin{array}{c} \frac{1}{2} \\ 1 \end{array} \right]\end{aligned}$$

9.2. Il metodo del simplesso duale: sviluppo mediante dizionario

che risulta ammissibile e, pertanto, ottimale.

Utilizzando la notazione introdotta per descrivere il dizionario in questo paragrafo, si possono mettere a confronto i due algoritmi:

SIMPLESSO PRIMALE	SIMPLESSO DUALE
INIZIALIZZAZIONE	INIZIALIZZAZIONE
Base B ammissibile: $A_B^{-1}b \geq 0$	Base B \mathcal{D} -ammissibile: $\hat{c}^T \geq 0$
OTTIMALITÀ \mathcal{D} -ammissibilità: $\hat{c}^T \geq 0$	OTTIMALITÀ \mathcal{P} -ammissibilità: $A_B^{-1}b \geq 0$
VARIABILE ENTRANTE $x_s: s \in N, \hat{c}_s < 0$	VARIABILE USCENTE $x_r: r \in B, \bar{x}_r = e_i^T A_B^{-1}b < 0$
VARIABILE USCENTE $x_i: e_i^T A_B^{-1}A_s > 0$	VARIABILE ENTRANTE $x_s: e_i^T A_B^{-1}A_s < 0$
$i \in \arg \min_{i \in B: e_i^T A_B^{-1}A_s > 0} \frac{e_i^T A_B^{-1}b}{e_i^T A_B^{-1}A_s}$	$s \in \arg \min_{j \in N: e_i^T A_B^{-1}A_s < 0} \frac{\hat{c}_j}{-e_i^T A_B^{-1}A_s}$

Come si può notare, vi è una certa simmetria tra i due metodi. Tale simmetria viene confermata dalla proprietà seguente:

Teorema 39. *Le operazioni di pivot compiute dall'algoritmo del simplesso duale applicato ad un problema di PL standard sono identiche a quelle che compirebbe l'algoritmo del simplesso se applicato al problema duale.*

La dimostrazione, abbastanza meccanica, viene lasciata come esercizio.

9.2 Il metodo del simplesso duale: sviluppo mediante dizionario

Le stesse operazioni viste nel paragrafo precedente possono essere effettuate direttamente sul dizionario del metodo del simplesso. Il motivo per cui vengono ripresentate sta nel fatto che, pur essendo

9. IL METODO DEL SIMPLEXO DUALE

decisamente meno elegante della formulazione matriciale, la rappresentazione tramite dizionario è più semplice. Si consideri un dizionario per un problema di *PL* standard:

$$\begin{aligned} z &= c_B^T A_B^{-1} b + \hat{c}_N^T x_N \\ x_B &= A_B^{-1} b - A_B^{-1} A_N x_N \end{aligned}$$

e si ipotizzi che la base A_B sia ammissibile per il duale. Sia cioè

$$\hat{c}_N^T = c_N^T - c_B^T A_B^{-1} A_N \geq 0.$$

Naturalmente, se fosse $A_B^{-1} b \geq 0$, il dizionario rappresenterebbe una soluzione ottimale del problema. Altrimenti si immagini di compiere un'operazione di pivot, consistente, ad esempio, nello scambio tra la variabile x_r , con $r \in B$, con la variabile x_s , con $s \in N$. Si consideri, ad esempio, l'equazione del dizionario corrispondente alla variabile di base x_r :

$$x_r = \bar{x}_r + \sum_{j \in N} \bar{\alpha}_{rj} x_j;$$

dove con $\bar{\alpha}_{rj}$ si è indicato il coefficiente di x_j nell'equazione del dizionario associata ad x_r . Ipotizzando che $\bar{\alpha}_{rs} \neq 0$ ed effettuando il pivot, l'equazione assume la forma

$$\begin{aligned} x_s &= \frac{-\bar{x}_r - \sum_{j \in N \setminus \{s\}} \bar{\alpha}_{rj} x_j + x_r}{\bar{\alpha}_{rs}} \\ &= -\frac{\bar{x}_r}{\bar{\alpha}_{rs}} + \frac{x_r}{\bar{\alpha}_{rs}} - \sum_{j \in N \setminus \{s\}} \frac{\bar{\alpha}_{rj}}{\bar{\alpha}_{rs}} x_j. \end{aligned}$$

Dopo l'operazione di pivot, l'equazione del costo assume la forma

$$z = c_B^T A_B^{-1} b + \sum_{j \in N \setminus \{s\}} \hat{c}_j x_j + \hat{c}_s \frac{-\bar{x}_r - \sum_{j \in N \setminus \{s\}} \bar{\alpha}_{rj} x_j + x_r}{\bar{\alpha}_{rs}}.$$

9.2. Il metodo del simplesso duale: sviluppo mediante dizionario

Il costo associato alla nuova soluzione di base diviene:

$$z \leftarrow c_B^T A_B^{-1} b - \hat{c}_s \frac{\bar{x}_r}{\bar{\alpha}_{rs}}$$

ed i nuovi coefficienti di costo ridotto risultano i seguenti:

$$\begin{aligned}\hat{c}_r &\leftarrow \frac{\hat{c}_s}{\bar{\alpha}_{rs}} \\ \hat{c}_j &\leftarrow \hat{c}_j - \hat{c}_s \frac{\bar{\alpha}_{rj}}{\bar{\alpha}_{rs}} \quad j \in N \setminus \{s\}.\end{aligned}$$

Per imporre la \mathcal{D} -ammissibilità del nuovo dizionario occorre richiedere che tutti i nuovi coefficienti di costo ridotto siano non negativi. Per imporre la non negatività del coefficiente di costo ridotto della variabile appena uscita di base, x_r , si ha che necessariamente deve essere $\bar{\alpha}_{rs} > 0$. Per la non negatività di tutti gli altri coefficienti di costo ridotto, si osservi che se $\bar{\alpha}_{rj} \leq 0$ questa è garantita, grazie alla non negatività dei coefficienti di costo ridotto prima dell'operazione di pivot e dal fatto che, come si è appena visto, deve essere $\bar{\alpha}_{rs} > 0$. Se invece $\bar{\alpha}_{rj} > 0$, la \mathcal{D} -ammissibilità sarà mantenuta a patto che

$$\hat{c}_j - \hat{c}_s \frac{\bar{\alpha}_{rj}}{\bar{\alpha}_{rs}} \geq 0 \quad \forall j \in N : \bar{\alpha}_{rj} > 0$$

ovvero

$$\frac{\hat{c}_s}{\bar{\alpha}_{rs}} \leq \frac{\hat{c}_j}{\bar{\alpha}_{rj}} \quad \forall j \in N : \bar{\alpha}_{rj} > 0$$

da cui segue la regola per la scelta della variabile *entrante* in base:

Una volta decisa la variabile uscente di base, x_r , la variabile entrante in base, x_s , deve essere scelta in modo tale che:

$$s \in \arg \min_{j \in N : \bar{\alpha}_{rj} > 0} \left\{ \frac{\hat{c}_j}{\bar{\alpha}_{rj}} \right\}$$

9. IL METODO DEL SIMPLEXO DUALE

Imponendo poi la non decrescita del costo si ottiene

$$-\hat{c}_s \frac{\bar{x}_r}{\bar{\alpha}_{rs}} \geq 0$$

che, grazie al fatto che $\hat{c}_s \geq 0$ e che s deve essere scelto in modo tale che $\bar{\alpha}_{rs} > 0$, conduce alla seguente regola:

La variabile uscente di base, x_r , deve essere scelta in modo tale che

$$\bar{x}_r < 0.$$

Esempio 25. Si consideri un problema di PL standard rappresentato mediante il dizionario seguente:

$$\begin{aligned} z &= -8 + 2x + y + 4u + t \\ v &= -1 + x - y - u + t \\ w &= 1 - x + 2y + u - t. \end{aligned}$$

Come si nota, la base rappresentata dalle variabili v e w risulta non ammissibile per il primale, ma, essendo i coefficienti di costo ridotto non negativi, ammissibile per il duale. Poiché $v = -1$ la variabile v uscirà di base nel corso dell'operazione di pivot. Per decidere quale sia la variabile entrante in base, occorre considerare le variabili fuori base con coefficiente positivo, e cioè x e t . Confrontando i rapporti

$$\frac{2}{1}, \frac{1}{1}$$

si vede che il minimo è quello corrispondente a t . Lo scambio avverrà quindi fra v e t . Si ottiene il dizionario:

$$\begin{aligned} z &= -7 + x + 2y + 5u + v \\ t &= 1 + v - x + y + u \\ w &= y - v \end{aligned}$$

che, essendo anche ammissibile per il primale, risulta ottimale.

Analisi di sensitività

10.1 Introduzione

Con il termine di “analisi di sensitività” si intende una serie di tecniche, di notevole importanza applicativa, mediante le quali è possibile ottenere informazioni su quando e di quanto può variare la soluzione di un problema di *PL* variando uno o più dati del problema. Tramite l’analisi di sensitività è possibile rispondere a domande del tipo “cosa accadrebbe se un certo coefficiente venisse perturbato?” senza dover risolvere dall’inizio un nuovo problema di *PL*.

Un’analisi di questo tipo, nota anche con il termine di *analisi di post-ottimalità*, è estremamente importante per un numero di ragioni. Innanzitutto in molti casi i dati del problema (cioè la matrice A , il vettore dei costi c , il termine noto b) non possono essere considerati come dati esatti, ma sono desunti da osservazioni sperimentali: è importante quindi controllare che la soluzione sia “stabile”, nel senso che sarebbe poco accettabile una soluzione che presentasse forti variazioni in seguito a piccoli cambiamenti nei dati. Inoltre in molti casi i dati del problema sono in qualche modo *controllabili* dall’utente, il quale può, entro certi limiti, variarli: si pensi, nel caso della dieta, a possibili variazioni nelle richieste minime b , oppure, in un problema di trasporto, a variazioni nella domanda o nella capacità produttiva. Lo scopo dell’analisi è controllare se ta-

10. ANALISI DI SENSITIVITÀ

li variazioni portano ad una modifica significativa della soluzione ottimale.

L'analisi di sensitività verrà presentata con riferimento al seguente esempio:

Esempio 26. Tra i maggiori utilizzatori dei modelli di programmazione lineare e del metodo del simplexso, sicuramente si trova l'industria petrolifera. L'esempio qui riportato deriva da un modello estremamente semplificato di problemi di miscela reali nel campo della raffinazione della benzina. Si consideri il problema di miscellare in opportune proporzioni alcuni prodotti (nell'esempio i primi due si riferiscono a benzine, il terzo a nafta) in modo tale da rispettare alcuni vincoli sulla qualità del prodotto risultante. Tra gli elementi caratteristici della qualità di una benzina vi sono: il numero di ottani, la pressione dei vapori, la volatilità. Ipotizzando che in una miscela questi tre componenti si comportino in modo lineare, si consideri il problema di determinare la composizione di un'unità di benzina di costo minimo utilizzando 3 tipi di materie prime, ciascuna delle quali caratterizzata dalla seguente composizione:

	X	Y	Z
Costo	10	20	12
Ottani	125	100	75
Pressione	2	2	4
Volatilità	6	8	10

e si immagini di voler determinare la composizione ottimale di una unità di prodotto sotto le condizioni seguenti:

- numero di ottani minimo: 100;
- Pressione massima: 3;
- Volatilità minima: 8.

10.1. Introduzione

I dati sono desunti, con modifiche, da un caso abbastanza realistico. Il problema di *PL* risultante ha dunque la forma seguente:

	X	Y	Z	s_1	s_2	s_3	
min	10X	+	20Y	+	12Z		
Ottani	125X	+	100Y	+	75Z	- s_1	= 100
Pressione	2X	+	2Y	+	4Z	+ s_2	= 3
Volatilità	6X	+	8Y	+	10Z	- s_3	= 8
Totalle	X	+	Y	+	Z		= 1

oltre, naturalmente, ai vincoli di non negatività su tutte le variabili (slack compresi). Risolvendo con il metodo del simplesso, si giunge ad un dizionario del tipo:

$$z = 11 + 9s_2 + 5s_3$$

$$X = \frac{1}{2} - \frac{1}{2}s_2 - \frac{1}{2}s_3$$

$$Y = s_2 + \frac{1}{2}s_3$$

$$Z = \frac{1}{2} - \frac{1}{2}s_2$$

$$s_1 = -\frac{25}{2}s_3$$

e si ottengono le seguenti informazioni:

$$B : \{X, Y, Z, s_1\}$$

$$A_B^{-1} = \begin{bmatrix} 0 & \frac{1}{2} & -\frac{1}{2} & 3 \\ 0 & -1 & \frac{1}{2} & -1 \\ 0 & \frac{1}{2} & 0 & -1 \\ -1 & 0 & -\frac{25}{2} & 200 \end{bmatrix}; \quad x_B = \begin{bmatrix} \frac{1}{2} \\ 0 \\ \frac{1}{2} \\ 0 \end{bmatrix}$$

$$\lambda_B^T = [0 \quad -9 \quad 5 \quad -2]; z = 11$$

10. ANALISI DI SENSITIVITÀ

Si consideri un problema di *PL* in forma standard

$$\begin{aligned} \min_x c^T x \\ Ax = b \\ x \geq 0 \end{aligned}$$

ed il corrispondente duale

$$\begin{aligned} \max_{\lambda} \lambda^T b \\ \lambda^T A \leq c^T \end{aligned}$$

e, supponendo che entrambi abbiano soluzione ottimale, si indichino con x^*, λ^*, z^* rispettivamente la soluzione ottimale del primale, la soluzione ottimale del duale ed il valore ottimale della funzione obiettivo. Si pone il problema di analizzare le variazioni nella soluzione ottimale in seguito a:

- cambiamenti del vettore b ;
- cambiamenti del vettore c ;
- aggiunta di una nuova variabile;
- cambiamenti della matrice A ;
- aggiunta di un nuovo vincolo.

Nell'analisi che segue, si considererà ogni cambiamento singolarmente; l'effetto combinato di più cambiamenti può essere analizzato in modo analogo.

10.2 Sensitività sul termine noto

Se il vettore dei termini noti b viene modificato, indicando con \tilde{b} il vettore modificato e con Δ la variazione, si avrà

$$\tilde{b} = \tilde{b}(\Delta) = b + \Delta.$$

10.2. Sensitività sul termine noto

Si indichino con $\tilde{x} = \tilde{x}(\Delta), \tilde{\lambda} = \tilde{\lambda}(\Delta), \tilde{z} = \tilde{z}(\Delta)$ rispettivamente un ottimo del problema modificato, un ottimo del duale ed il valore ottimale della coppia primale/duale. Naturalmente $x^* = \tilde{x}(0)$. In seguito alla modifica effettuata, la regione ammissibile del problema *duale* non cambia; pertanto λ^* è sicuramente soluzione *ammissibile* per il duale modificato, cioè per il problema

$$\begin{aligned} \max_{\lambda} & \lambda^T(b + \Delta) \\ & \lambda^T A \leq c^T. \end{aligned}$$

Quindi λ^* , pur non essendo necessariamente una soluzione ottimale, resta ammissibile per il duale e pertanto

$$\begin{aligned} \tilde{z}(\Delta) &= \tilde{\lambda}(\Delta)(b + \Delta) \\ &\geq \lambda^*(b + \Delta) \end{aligned}$$

si può quindi immediatamente ottenere una limitazione alla variazione nel costo ottimale dovuta alla perturbazione Δ :

$$\tilde{z}(\Delta) \geq z^* + \lambda^{*T} \Delta. \quad (10.1)$$

Se si considera il caso particolare in cui la variazione riguarda un unico termine noto $\Delta = \delta e_i$, si ottiene

$$\tilde{z}(\delta) \geq z^* + \lambda^{*T} \delta e_i.$$

da cui

$$\tilde{z}(\delta) \geq z^* + \lambda_i^* \delta$$

In generale, data una funzione $f : S \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, un vettore $v \in \mathbb{R}^n$ si dice *sottogradiente* di f in x se

$$f(y) \geq f(x) + v^T(y - x) \quad \forall y \in S.$$

Si può dimostrare che, nel caso di funzioni convesse, un sottogradiente esiste sempre; inoltre, nel caso differenziabile, coincide

10. ANALISI DI SENSITIVITÀ

con il gradiente. Dal punto di vista geometrico, il sottogradiente identifica la normale ad un iper piano di supporto del grafico della funzione nel punto x .

Per quanto visto, la componente i -esima di una qualsiasi soluzione ottima del duale rappresenta quindi un sottogradiente (in questo caso una sottoderivata) della funzione costo ottimale rispetto alla variazione dell' i -esimo termine noto; nei punti ove esiste la derivate prima, il suo valore coincide con quello della componente dell'ottimo del duale.

Si indichi con B una base ottimale per il problema primale alla quale sono associati coefficienti di costo ridotto non negativi $\hat{c}^T = c_N^T - c_B^T A_B^{-1} A_N \geq 0$. Come si nota dalla formula dei coefficienti di costo ridotto e come si deduce facilmente dal fatto che la modifica di b non cambia l'ammissibilità del duale, anche dopo la variazione questi coefficienti di costo ridotto restano non negativi. La base B resterà pertanto ottimale se e solo se sarà ammissibile. Poiché il sistema di equazioni $Ax = b$ è soddisfatto da ogni soluzione di base per costruzione, l'ammissibilità, e l'ottimalità, della base B saranno garantite se e solo se

$$A_B^{-1}(b + \Delta) \geq 0.$$

Se questa condizione vale, allora un ottimo di base del primale sarà dato da

$$\tilde{x}_B(\Delta) = x_B^* + A_B^{-1}\Delta$$

con

$$\begin{aligned} \tilde{z}(\Delta) &= z^* + c_B^T A_B^{-1}\Delta \\ &= z^* + \lambda_B^{*T} \Delta, \end{aligned}$$

dove con λ_B^* si è indicata la soluzione "di base" del duale associata a B . La disequazione (10.1) nel caso in cui la base ottima non varia diviene un'equazione. Si osservi però che, mentre la disequazione è valida per qualsiasi soluzione ottima del duale, l'equazione è valida

esclusivamente se si utilizza la soluzione di base del duale complementare associata ad una base con coefficienti di costo ridotto non negativi. Proseguendo l'analisi già iniziata e relativa al cambiamento di un solo elemento del termine noto $\Delta = \delta e_i$, si osserva che, nel caso in cui la base ottima non vari, si ottiene

$$\frac{\tilde{z}(\delta) - z^*}{\delta} = \lambda_{B_i}^*.$$

Un caso in cui sicuramente la base ottima non varia, almeno per perturbazioni δ sufficientemente piccole, è quello in cui la base B non è degenere. Infatti se $A_B^{-1}b > 0$, necessariamente esisterà un intervallo, non necessariamente finito, $[\delta_\ell, \delta_h]$, con $\delta_\ell < 0 < \delta_h$, tale che $A_B^{-1}b + A_B^{-1}e_i\delta \geq 0$ per ogni $\delta \in [\delta_\ell, \delta_h]$. In questo caso, non solo si può immediatamente dedurre la soluzione ottimale per un intervallo di variazioni possibili, ma è anche possibile fornire un'interpretazione geometrica dell'ottimo del duale. Vale infatti in questo caso

$$\begin{aligned}\lambda_{B_i}^* &= \frac{\tilde{z}(\delta) - z^*}{\delta} \\ &= \lim_{\delta \rightarrow 0} \frac{\tilde{z}(\delta) - z^*}{\delta} \\ &= \left. \frac{dz^*}{d\delta} \right|_0.\end{aligned}$$

quindi:

Teorema 40. *Le componenti della soluzione ottimale del duale di un problema di PL, nel caso in cui la base non sia degenere, rappresentano il gradiente della funzione di obiettivo nel punto di ottimo rispetto al termine noto.*

E' importante anche osservare che nel caso non degenere la soluzione ottimale del duale è unica; in questo caso l' i -esima componente della soluzione ottimale del duale rappresenta la derivata

10. ANALISI DI SENSITIVITÀ

del valore ottimale della coppia primale/duale rispetto a variazioni dell' i -esimo termine noto.

Si può osservare che, nel caso in cui la variazione del termine noto comporti un cambio della base ottimale, non è necessario ripetere l'intero metodo del simplex; la perdita di ottimalità a seguito di una variazione del termine noto avviene mantenendo l'ammisibilità del duale. Sarà pertanto possibile utilizzare il dizionario finale del metodo del simplex, opportunamente modificato nel termine noto e nel costo, come dizionario iniziale per il metodo del simplex *duale*. In particolare, se il dizionario finale del problema era

$$\begin{aligned} z &= c_B^T A_B^{-1} b + \hat{c}^T x_N \\ x_B &= A_B^{-1} b - A_B^{-1} A_n x_N \end{aligned}$$

dopo la modifica da b a $b + \Delta$, il dizionario si trasforma in

$$\begin{aligned} z &= c_B^T A_B^{-1} (b + \Delta) + \hat{c}^T x_N \\ x_B &= A_B^{-1} (b + \Delta) - A_B^{-1} A_n x_N \end{aligned}$$

dove, come si vede, cambiano solo i termini noti delle equazioni. Se una componente di base diviene negativa, il dizionario si trova nella forma adatta a poter proseguire col simplex duale.

Esempio 27. Nell'esempio sopra considerato, si immagini di variare le richieste relative al numero di ottani, alla pressione, alla volatilità. Ad esempio, variando di δ unità la richiesta relativa al numero di ottani, la base corrente risulterà ottimale a patto che

$$\left[\begin{array}{c} \frac{1}{2} \\ 0 \\ \frac{1}{2} \\ 0 \end{array} \right] + \left[\begin{array}{cccc} 125 & 100 & 75 & -1 \\ 2 & 2 & 4 & 0 \\ 6 & 8 & 10 & 0 \\ 1 & 1 & 1 & 0 \end{array} \right]^{-1} \left[\begin{array}{c} \delta \\ 0 \\ 0 \\ 0 \end{array} \right] \geq 0$$

10.2. Sensitività sul termine noto

cioè

$$\begin{bmatrix} \frac{1}{2} \\ 0 \\ \frac{1}{2} \\ -\delta \end{bmatrix} \geq 0,$$

da cui si deduce che la base corrente resta ottimale per qualsiasi valore negativo di δ . La soluzione resta invariata per qualsiasi diminuzione del numero di ottani richiesto: infatti la miscela ottimale resta $X = Z = 50\%$; naturalmente lo slack relativo al numero di ottani cambia, riflettendo l'eccesso di ottani di una simile miscela rispetto al minimo richiesto.

Se invece si suppone di aumentare il numero di ottani, necessariamente la base ottimale deve cambiare. Per determinare la nuova soluzione ottimale si può aggiornare il dizionario finale del metodo del simplex:

$$\begin{aligned} z &= 11 + 9s_2 + 5s_3 \\ X &= \frac{1}{2} - \frac{1}{2}s_2 - \frac{1}{2}s_3 \\ Y &= s_2 + \frac{1}{2}s_3 \\ Z &= \frac{1}{2} - \frac{1}{2}s_2 \\ s_1 &= -\delta - \frac{25}{2}s_3 \end{aligned}$$

Dall'ultima equazione di questo dizionario, ammissibile per il problema duale, si vede immediatamente che il problema non ammette soluzioni ammissibili per $\delta > 0$, essendo il duale illimitato.

Esempio 28. Si immagini ora di modificare la richiesta relativa alla pressione, che passa quindi da 3 a $3 + \delta$. La base attuale resta

10. ANALISI DI SENSITIVITÀ

ottimale a patto che

$$\begin{bmatrix} \frac{1}{2} \\ 0 \\ \frac{1}{2} \\ 0 \end{bmatrix} + \begin{bmatrix} 125 & 100 & 75 & -1 \\ 2 & 2 & 4 & 0 \\ 6 & 8 & 10 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \delta \\ 0 \\ 0 \end{bmatrix} \geq 0,$$

cioè che

$$\begin{bmatrix} \frac{1}{2} + \frac{1}{2}\delta \\ -\delta \\ \frac{1}{2} + \frac{1}{2}\delta \\ 0 \end{bmatrix} \geq 0.$$

In questo caso, risolvendo il sistema di disequazioni nella variabile δ , si ottiene l'insieme di variazioni possibili: $-1 \leq \delta \leq 0$. Quindi variando la richiesta di pressione massima tra 2 e 3 la base ottimale non cambia. Cambia naturalmente la composizione della miscela ottimale, con l'inserimento a valori positivi anche del componente Y , e con un costo totale dato da

$$\begin{bmatrix} 10 & 20 & 12 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2} + \frac{1}{2}\delta \\ -\delta \\ \frac{1}{2} + \frac{1}{2}\delta \\ 0 \end{bmatrix} = 11 - 9\delta$$

Aumentando la richiesta di pressione massima, ponendo $\delta > 0$, si può determinare il nuovo ottimo iniziando dal dizionario seguente, che risulta ammissibile per il duale ma non per il primale (seconda

10.2. Sensitività sul termine noto

equazione):

$$\begin{aligned} z &= 11 - 9\delta + 9s_2 + 5s_3 \\ X &= \frac{1}{2} + \frac{1}{2}\delta - \frac{1}{2}s_2 - \frac{1}{2}s_3 \\ Y &= -\delta + s_2 + \frac{1}{2}s_3 \\ Z &= \frac{1}{2} + \frac{1}{2}\delta - \frac{1}{2}s_2 \\ s_1 &= -\frac{25}{2}s_3 \end{aligned}$$

per arrivare, dopo un'operazione di pivot, al dizionario finale:

$$\begin{aligned} z &= 11 + 9Y + \frac{1}{2}s_3 \\ X &= \frac{1}{2} - \frac{1}{2}Y - \frac{1}{4}s_3 \\ s_2 &= \delta + Y - \frac{1}{2}s_3 \\ Z &= \frac{1}{2} - \frac{1}{2}Y + \frac{1}{4}s_3 \\ s_1 &= -\frac{25}{2}s_3 \end{aligned}$$

Questo dizionario è ottimale per qualsiasi scelta della costante $\delta \geq 0$ e corrisponde alla base $B = \{X, s_2, Z, s_1\}$. L'ottimo del duale associato a tale base è dato da

$$\lambda^T = [\ 10 \ 0 \ 12 \ 0 \] \left[\begin{array}{cccc} 125 & 0 & 75 & -1 \\ 2 & 1 & 4 & 0 \\ 6 & 0 & 10 & 0 \\ 1 & 0 & 1 & 0 \end{array} \right]^{-1} = [\ 0 \ 0 \ \frac{1}{2} \ 7 \].$$

Se invece δ viene scelto inferiore a -1 , il dizionario precedente risulta ammissibile per il duale, ma non ammissibile per il primale

10. ANALISI DI SENSITIVITÀ

a causa sia della prima che della terza equazione. Da entrambe le equazioni si deduce immediatamente che il duale è illimitato e, di conseguenza, il primale è inammissibile. Il grafico del costo ottimale in funzione dei valori possibili per il requisito massimo di pressione del vapore risulta pertanto quello indicato in Figura 10.1.

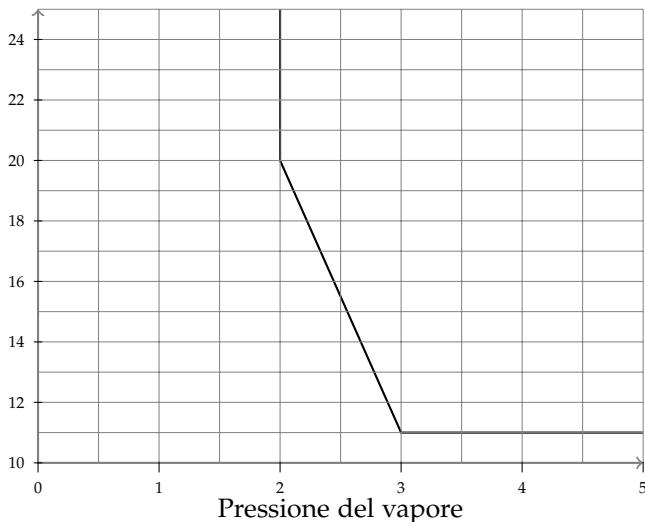


Figura 10.1: Andamento del costo ottimale in funzione della pressione del vapore

Si può anche osservare che, nel caso della variazione del requisito di pressione massima, il grafico del costo ottimale presenta due tratti lineari con pendenza, rispettivamente, pari a -9 e 0 . Questi sono anche i valori della seconda componente dell'ottimo di base del duale. Tuttavia occorre ricordare che in casi degeneri, come questo, il duale possiede in genere infinite soluzioni ottimali, solo una delle quali rappresenta la pendenza della funzione di costo.

10.3 Sensitività sul vettore dei costi

In questo paragrafo viene analizzato l'effetto prodotto sulla soluzione da variazioni nei costi associati alle variabili. Questa situazione è, in un certo senso, duale della precedente: la variazione di uno o più coefficienti di costo non modifica la regione ammissibile del problema primale. La soluzione ottimale del primale risulta pertanto ammissibile anche dopo la variazione, anche se, naturalmente, può cessare di essere ottimale. Se la soluzione ottimale conosciuta è di base, e ad essa sono associati coefficienti di costo ridotto non negativi, non essendo variata l'ammissibilità, l'ottimalità sarà garantita dalla non negatività dei coefficienti di costo ridotto. Dall'espressione dei coefficienti di costo ridotto si vede che se viene variato il costo di una variabile non appartenente alla base ottimale, questa variazione implica un cambiamento *esclusivamente* del coefficiente di costo ridotto della variabile in questione (che può diventare competitiva rispetto alle variabili attualmente in base). Viceversa, se varia il costo di una variabile appartenente alla base ottimale, si ha in generale una variazione in *tutti* i coefficienti di costo ridotto delle variabili non di base. Occorre pertanto distinguere i due casi.

Variazione del costo di una variabile non di base

I coefficienti di costo ridotto, se con B ed N si indicano rispettivamente la base ottimale e gli indici delle variabili non di base, sono dati dall'espressione

$$\begin{aligned}\hat{c}^T &= c_N^T - c_B^T A_B^{-1} A_N \\ &= c_N^T - \lambda_B^T A_N.\end{aligned}$$

Se c_j , con $j \in N$, subisce un incremento δ , il coefficiente di costo ridotto corrispondente diviene

$$c_j + \delta - \lambda_B^T A_j = \hat{c}_j + \delta.$$

10. ANALISI DI SENSITIVITÀ

L'ottimo cambia se e solo se

$$\delta < \lambda_B^T A_j - c_j$$

cioè se e solo se

$$\delta < -\hat{c}_j$$

che, come ovvio, è una quantità negativa.

Variazione del costo di una variabile di base

Se il costo c_j di una variabile appartenente alla base ottimale varia di una quantità δ , i coefficienti di costo ridotto divengono

$$c_N^T - (c_B^T + \delta e_j^T) A_B^{-1} A_N$$

dove, al solito, e_j rappresenta il j -esimo versore. I nuovi coefficienti di costo ridotto sono dati da

$$\hat{c}^T - \delta e_j^T A_B^{-1} A_N.$$

Il calcolo dei nuovi coefficienti di costo ridotto risulta particolarmente semplice avendo a disposizione il dizionario finale del metodo del simplex: gli elementi del vettore $-e_j^T A_B^{-1} A_N$ sono semplicemente i coefficienti delle variabili non di base nell'equazione del dizionario corrispondente alla variabile della quale si è variato il costo. Nel caso della variazione anche di un solo coefficiente di costo di una variabile appartenente alla base ottimale possono variare tutti i coefficienti di costo ridotto; al contrario di quanto avviene variando il costo di una variabile fuori base, nel caso del cambiamento del costo di una variabile di base può darsi che la base ottima cambi anche in seguito ad una *diminuzione* del costo. Questa apparente contraddizione viene eliminata osservando che in questo caso il cambio di base non provocherà l'uscita di base della variabile alla quale è stato diminuito il costo, ma di un'altra.

10.3. Sensitività sul vettore dei costi

Immaginando, come si è fatto nel caso della variazione di un termine noto, di variare con continuità il costo di una variabile, si può caratterizzare la funzione costo ottimo in dipendenza di tale variazione. Immaginando per il momento che la variabile il cui costo viene variato sia di base, si avrà che, nel caso in cui tutti i coefficienti di costo ridotto siano strettamente positivi, la soluzione ottimale non cambierà per cambiamenti sufficientemente piccoli del costo. La soluzione ottimale corrente resterà pertanto ottimale ed il costo varierà nel modo seguente:

$$\tilde{z}(\delta) = z^* + (A_B^{-1}b)_j \delta = z^* + x_{B_j}^* \delta.$$

Il costo ottimale varierà dunque linearmente, con pendenza pari al valore della j -esima variabile di base. Nel momento in cui, a seguito della variazione, almeno uno dei coefficienti di costo ridotto si annulla, proseguendo nella perturbazione si perderà l'ottimalità (ma non l'ammissibilità) della soluzione corrente. Occorrerà dunque determinare un nuovo ottimo, cosa che si può fare iniziando con il metodo del simplesso da un dizionario ricavato da quello ottimale prima della modifica dei costi. Il valore $z^* + (A_B^{-1}b)_j \delta$, anche nel caso in cui B non sia più una base ottimale, fornirà una stima per eccesso della funzione costo ottimale. Si può dimostrare in modo relativamente semplice che la funzione costo ottimo risulta lineare a tratti e *concava*.

L'analisi appena fatta riguarda la variazione di un'unica componente del vettore dei costi. Naturalmente variazioni più complesse possono essere analizzate seguendo la stessa traccia: modificando di un vettore Δ_B il costo delle variabili di base e di Δ_N quello delle variabili fuori base, i coefficienti di costo ridotto divengono

$$\begin{aligned}\hat{c}^T(\Delta) &= c_N + \Delta_N - (c_B + \Delta_B)^T A_B^{-1} A_N \\ &= \hat{c}^T + \Delta_N - \Delta_B^T A_B^{-1} A_N.\end{aligned}$$

10. ANALISI DI SENSITIVITÀ

Esempio 29. Se il costo del materiale X viene variato e diviene pari a $10 + \delta$, i coefficienti di costo ridotto risultano

$$\left[9 - \frac{1}{2}\delta \quad 5 - \frac{1}{2}\delta \right],$$

quindi la soluzione ottimale resterà invariata fino a quando $\delta \leq 10$. Per valori superiori necessariamente l'ottimo deve cambiare, a patto che il problema non divenga illimitato. Ad esempio, se il costo di X diventasse pari a 22 ($\delta = 12$), ricordando che, a seguito del cambio del costo di X il costo della base corrente viene modificato in $11 + \frac{\delta}{2}$, il dizionario diviene:

$$\begin{aligned} z &= 17 + 3s_2 - s_3 \\ X &= \frac{1}{2} - \frac{1}{2}s_2 - \frac{1}{2}s_3 \\ Y &= s_2 + \frac{1}{2}s_3 \\ Z &= \frac{1}{2} - \frac{1}{2}s_2 \\ s_1 &= -\frac{25}{2}s_3 \end{aligned}$$

e, con un'operazione di pivot, si giunge a

$$\begin{aligned} z &= 17 + 3s_2 + \frac{2}{25}s_1 \\ X &= \frac{1}{2} - \frac{1}{2}s_2 + \frac{1}{25}s_1 \\ Y &= s_2 - \frac{1}{25}s_1 \\ Z &= \frac{1}{2} - \frac{1}{2}s_2 \\ s_3 &= -\frac{2}{25}s_1 \end{aligned}$$

10.3. Sensitività sul vettore dei costi

da dove si vede come un cambio della base ottimale non porta necessariamente ad un cambio della soluzione ottimale: la combinazione di X e Z al 50% resta ottimale anche per valori del costo di X superiori.

Se si decidesse invece di variare il costo unitario di Y si otterrebbero i coefficienti di costo ridotto

$$\begin{bmatrix} 9 + \delta & 5 + \frac{1}{2}\delta \end{bmatrix}$$

da cui risulta che per diminuzioni di almeno 9 unità, la base ottimale deve cambiare. E' intuitibile che, a seguito della diminuzione di costo, non sarà Y la variabile destinata ad uscire di base. Infatti osservando il dizionario nel caso generale:

$$z = 11 + (9 + \delta)s_2 + \left(5 + \frac{1}{2}\delta\right)s_3$$

$$X = \frac{1}{2} - \frac{1}{2}s_2 - \frac{1}{2}s_3$$

$$Y = s_2 + \frac{1}{2}s_3$$

$$Z = \frac{1}{2} - \frac{1}{2}s_2$$

$$s_1 = -\frac{25}{2}s_3$$

si vede immediatamente che in caso di valori di δ inferiori a -9 il dizionario non è più ottimale ed è necessario effettuare un'operazione di pivot scambiando X o Z con s_2 oppure, nel caso $\delta \leq -10$, in alternativa s_1 con s_3 . In tutti i casi Y resta in base. Ad esempio,

10. ANALISI DI SENSITIVITÀ

se $\delta = -9.5$, scambiando X con s_2 , si ottiene il dizionario

$$\begin{aligned} z &= \frac{21}{2} + X + \frac{3}{4}s_3 \\ s_2 &= -2X + 1 - s_3 \\ Y &= 1 - 2X - \frac{1}{2}s_3 \\ Z &= X + \frac{1}{2}s_3 \\ s_1 &= -\frac{25}{2}s_3 \end{aligned}$$

che corrisponde ad una soluzione in cui la miscela è costituita al 100% da Y .

10.4 Aggiunta di una nuova variabile

In certi casi è possibile che una nuova variabile si renda disponibile. È opportuno a questo punto chiedere se tale variabile sia o meno conveniente, cioè se la presenza di questa nuova possibilità modifichi o meno la scelta ottimale corrente. Il problema è equivalente all'analisi dell'effetto del cambiamento del costo di una variabile non di base: la nuova variabile infatti può essere considerata come una variabile non appartenente alla base. Il costo ridotto della nuova variabile x_{n+1} , se con c_{n+1} si indica il costo e con A_{n+1} la colonna dei coefficienti di x_{n+1} nella matrice dei vincoli, sarà

$$\hat{c}_{n+1} = c_{n+1} - \lambda_B^{\star T} A_{n+1};$$

pertanto la nuova variabile sarà competitiva rispetto a quelle che attualmente formano la base ottimale se e solo se

$$c_{n+1} < \lambda_B^{\star T} A_{n+1}.$$

In questo caso per determinare il nuovo ottimo, basta ricorrere al metodo del simplex, inizializzato da un dizionario ottenuto

10.4. Aggiunta di una nuova variabile

da quello ottimale prima della variazione. In tale dizionario si dovranno inserire i coefficienti della nuova variabile, oltre che nella definizione del costo, in ciascuna equazione: poiché il dizionario è definito da

$$x_B = A_B^{-1}b - A_B^{-1}A_N x_N$$

i coefficienti della nuova variabile nel dizionario saranno dati da

$$-A_B^{-1}A_{n+1}.$$

Pertanto il dizionario con il quale si può iniziare la risoluzione del problema dopo l'aggiunta della nuova variabile sarà

$$\begin{aligned} z &= c_B^T A_B^{-1} b + \hat{c}^T x_N + (c_{n+1} - c_B^T A_B^{-1} A_{n+1}) x_{n+1} \\ x_B &= A_B^{-1} b - A_B^{-1} A_N x_N - A_B^{-1} A_{n+1} x_{n+1} \end{aligned}$$

Esempio 30. Si supponga che, nell'esempio utilizzato finora in questo capitolo, si renda disponibile un nuovo prodotto, W , caratterizzato da:

costo: 20, Ottani:90, Pressione:2, Volatilità:10

Questo nuovo prodotto sarà conveniente se e solo se

$$\hat{c}_W = c_W - c_B^T A_B^{-1} A_W < 0,$$

cioè se e solo se risulta negativa la quantità

$$20 - \begin{bmatrix} 0 & -9 & 5 & -2 \end{bmatrix} \begin{bmatrix} 90 \\ 2 \\ 10 \\ 1 \end{bmatrix} = -10.$$

In questo caso l'aggiunta della nuova variabile alla base provoca tendenzialmente una diminuzione del costo totale. Volendo determinare la nuova soluzione ottimale, occorre procurarsi i coefficienti

10. ANALISI DI SENSITIVITÀ

di W da inserire nel dizionario:

$$A_B^{-1} A_W = \begin{bmatrix} 0 & \frac{1}{2} & -\frac{1}{2} & 3 \\ 0 & -1 & \frac{1}{2} & -1 \\ 0 & \frac{1}{2} & 0 & -1 \\ -1 & 0 & -\frac{25}{2} & 200 \end{bmatrix} \begin{bmatrix} 90 \\ 2 \\ 10 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ 0 \\ -15 \end{bmatrix}$$

da cui si ottiene il dizionario

$$z = 11 + 9s_2 + 5s_3 - 10W$$

$$X = \frac{1}{2} - \frac{1}{2}s_2 - \frac{1}{2}s_3 + W$$

$$Y = s_2 + \frac{1}{2}s_3 - 2W$$

$$Z = \frac{1}{2} - \frac{1}{2}s_2$$

$$s_1 = -\frac{25}{2}s_3 + 15W$$

e, con un passo di pivot,

$$z = 11 + 4s_2 + \frac{5}{2}s_3 + 5Y$$

$$X = \frac{1}{2} - \frac{1}{4}s_3 - \frac{1}{2}Y$$

$$W = -\frac{1}{2}Y + \frac{1}{2}s_2 + \frac{1}{4}s_3$$

$$Z = \frac{1}{2} - \frac{1}{2}s_2$$

$$s_1 = -\frac{35}{4}s_3 - \frac{15}{2}Y + \frac{15}{2}s_2.$$

Si vede che, in realtà, avendo effettuato un'operazione di pivot a livello 0, il costo ottimale non è cambiato, e W è entrata in base, ma a livello 0.

10.5 Sensitività sulla matrice dei coefficienti

Nel caso in cui venga variato un coefficiente della matrice dei vincoli A , occorre distinguere i due casi relativi a cambiamenti nelle colonne della base ottimale o nelle rimanenti colonne.

Modifica di una colonna non di base

Supponendo di modificare la colonna non di base A_j di un vettore Δ , si ottengono i nuovi coefficienti di costo ridotto

$$c_j - \lambda_B^{*T} (A_j + \Delta) = \hat{c}_j - \lambda_B^{*T} \Delta.$$

Come nel caso della variazione del costo, la modifica di uno o più coefficienti di una variabile che non appartiene alla base ottimale provoca la perdita di ottimalità solamente se il coefficiente di costo ridotto relativo a tale variabile diviene negativo.

Modifica di una colonna di base

In questo caso la situazione è in apparenza più complessa, poiché la variazione potrebbe provocare la perdita di ammissibilità della base per il primale, per il duale, o entrambe; potrebbe anche capitare che la base cessi di essere tale, cioè divenga una matrice non invertibile. In realtà è abbastanza semplice determinare una nuova soluzione ottimale procedendo nel modo seguente. In primo luogo si inserisce nel dizionario una variabile nuova, x_{n+1} , i cui coefficienti sono quelli variati; tale variabile viene inserita utilizzando la tecnica appena vista per le variabili fuori base. A questo punto occorre far uscire di base la variabile x_j alla quale sono stati cambiati i coefficienti. Questo può essere fatto, ad esempio, facendo crescere il costo di tale variabile fino a che non sarà necessario effettuare un'operazione di pivot. Se si modifica il costo di una variabile di base, ad esempio quella di indice j , i coefficienti di costo ridotto

10. ANALISI DI SENSITIVITÀ

diventano

$$\hat{c}^T = c_N^T - \lambda_B^{\star T} A_N - \Delta e_j^T A_B^{-1} A_N.$$

Almeno un coefficiente di costo ridotto diventerà negativo per Δ sufficientemente grande, se

$$e_j^T A_B^{-1} A_N \not\leq 0$$

In questo caso viene effettuata un'operazione di pivot che fa uscire di base la variabile x_j ; la variabile può essere eliminata definitivamente dal dizionario. Altrimenti, se

$$e_j^T A_B^{-1} A_N \leq 0$$

aumentando il costo nessun coefficiente di costo ridotto diventa negativo; in questo caso si può vedere che la variabile x_j non può uscire di base, pena la perdita di ammissibilità. La situazione corrisponde quindi al caso in cui l'eliminazione della variabile x_j provoca la perdita di ammissibilità del problema.

10.6 Aggiunta di un nuovo vincolo

L'aggiunta di vincoli in problemi di *PL* è un'operazione molto comune. Generalmente si procede all'aggiunta di vincoli per correggere una soluzione ottimale ritenuta non soddisfacente: è bene ricordare che la modellizzazione di un problema reale è un procedimento di tipo iterativo, che, partendo da una rappresentazione molto approssimata e semplificata di una situazione concreta, procede per raffinamenti successivi fino a giungere ad una rappresentazione che costituisca un ragionevole compromesso tra semplicità di rappresentazione e risoluzione, e rappresentatività. Molto spesso, una volta risolto un problema di *PL*, si giudica insoddisfacente la soluzione ottimale e, per migliorare la qualità della soluzione trovata, si introducono vincoli aggiuntivi: ad esempio, nei problemi di dieta spesso è necessario introdurre limiti sulle quantità dei singoli cibi,

10.6. Aggiunta di un nuovo vincolo

nei problemi di miscelazione si possono correggere soluzioni che tendono a privilegiare un unico fornitore, e così via. In altre situazioni, può capitare che un problema risulti illimitato: in tali casi il modello è sicuramente poco vincolato e può essere utile aggiungere vincoli che impediscono a determinate variabili di assumere valori arbitrariamente grandi. Infine, e forse questo è il caso più significativo, come si vedrà in seguito, molti algoritmi per la risoluzione di problemi di ottimizzazione lineare *intera* procedono per risoluzioni successive di problemi di *PL* che differiscono l'uno dall'altro semplicemente per l'aggiunta di un vincolo.

Naturalmente se la soluzione ottimale determinata soddisfa il vincolo aggiunto, continuerà ad essere ottimale. In caso contrario, sicuramente la soluzione ottimale varierà. Se la soluzione ottimale non soddisfa il vincolo aggiunto, quest'ultimo viene denominato *taglio*. Se il vincolo aggiunto è una disequazione del tipo

$$\alpha^T x \leq \beta$$

si potrà direttamente aggiungere il vincolo al dizionario:

$$\begin{aligned} z &= c^T \bar{x} + \hat{c}_N x_N \\ x_B &= \bar{x}_B - A_B^{-1} A_N x_N \\ s &= \beta - \alpha^T x \end{aligned}$$

e l'ultima equazione può essere immediatamente ricondotta alla forma necessaria per il dizionario eliminando le variabili di base x_B per sostituzione:

$$s = \beta - \alpha_B^T \bar{x}_B + (\alpha_B^T A_B^{-1} A_N - \alpha_N^T) x_N.$$

Se, come si è ipotizzato, il vincolo aggiunto è un taglio, il termine noto risultante dopo la sostituzione sarà negativo. I coefficienti di costo ridotto, non essendo cambiati, saranno ancora non negativi. Pertanto il nuovo dizionario può essere ri-ottimizzato mediante

10. ANALISI DI SENSITIVITÀ

l'algoritmo del simplex duale. Se il vincolo da aggiungere fosse un'equazione, anziché una disequazione, lo si potrebbe, ad esempio, ricondurre ad una coppia di disequazioni; in alternativa si potrebbe aggiungere il vincolo ed una variabile artificiale e quindi procedere con il metodo due fasi.

10.7 Lettura dell'output dei codici di programmazione lineare

Lo scopo di questo paragrafo è quello di vedere come, utilizzando le basi teoriche fornite dallo studio del metodo del simplex e della dualità, si possa leggere ed interpretare in modo non banale l'output fornito dai codici di calcolo per la programmazione lineare.

Problema della dieta

Si consideri il classico problema della dieta che, con il linguaggio AMPL, può essere codificato nel modo seguente:

```
set NUTR;    # elementi nutritivi
set CIBO;

param costo {CIBO} >= 0;
param qta_min {CIBO} >= 0. default 0;
param qta_max {j in CIBO} >= qta_min[j], default +infinity;
# quantita' minime e massime di ciascun cibo nella dieta - default 0 e
# infinito rispettivamente

param rich_min {NUTR} >= 0, default 0;
param rich_max {NUTR} >= 0, default +infinity;
# richieste minime e massime per i singoli elementi nutritivi

param compos {CIBO,NUTR} >= 0;
```

10.7. Lettura dell'output dei codici di programmazione lineare

```
var Qta {j in CIBO} >= qta_min[j], <= qta_max[j];  
minimize costo_totale: sum {j in CIBO} costo[j] * Qta[j];  
subject to dieta {i in NUTR}:  
    rich_min[i] <= sum {j in CIBO} compos[j,i] * Qta[j] <= rich_max[i];
```

Supponiamo di utilizzare un file di dati sui cibi disponibili, i prezzi al consumo, le caratteristiche nutritive composto ad esempio come segue:

```
set NUTR := Proteine Vit-A Vit-C VitB1  
                  VitB2 Niacina Calcio Ferro Sodio ;  
set CIBO := Hamburger McLean BigMac PatateFritte PolloFritto  
                  Miele ChefSalad InsalataVerde UovoMcMuffin Wheaties  
                  Vaniglia Latte SuccoArancia SuccoPompelmo SuccoMela;  
  
param      costo  :=  
Hamburger      0.59  
McLean        1.79  
BigMac        1.65  
PatateFritte   0.68  
PolloFritto    1.56  
Miele          0.00  
ChefSalad     2.69  
InsalataVerde 1.96  
UovoMcMuffin   1.36  
Wheaties       1.09  
Vaniglia       0.63  
Latte          0.56  
SuccoArancia   0.88  
SuccoPompelmo  0.68  
SuccoMela      0.68  
;  
  
param:  
      rich_min  rich_max :=  
Proteine      55      .  
Sodio         0       3000  
Vit-A        100     .  
Vit-C        100     .  
VitB1        100     .  
VitB2        100     .  
Niacina       100     .  
Calcio        100     .
```

10. ANALISI DI SENSITIVITÀ

```
Ferro      100      . ;  
  
param compos :  
Calorie Proteine Grassi Sodio Vit-A Vit-C VitB1 :=  
Hamburger 255    12     9   490    4    4   20  
McLean     320    22    10   670    10   10   25  
BigMac     500    25    26   890     6    2   30  
PatateFritte 220    3    12   110     0   15   10  
PolloFritto 270    20    15   580     0    0    8  
Miele       45     0     0    0     0    0    0  
ChefSalad  170    17     9   400    100   35   20  
InsalataVerde 50     4     2    70    90   35   6  
UovoMcMuffin 280    18    11   710    10    0   30  
Wheaties    90     2     1    220    20   20   20  
Vaniglia    105    4     1    80     2    0    2  
Latte       110    9     2   130    10    4    8  
SuccoArancia 80     1     0    0     0   120   10  
SuccoPomelo  80     1     0    0     0   100   4  
SuccoMela   90     0     0    5     0    2    2  
  
:           # continuazione tabella  
  
VitB2 Niacina  Calcio Ferro  :=  
  
Hamburger  10    20     10    15  
McLean     20    35     15    20  
BigMac     25    35     25    20  
PatateFritte 0    10     0     2  
PolloFritto 8    40     0     6  
Miele       0     0     0     0  
ChefSalad  15    20     15     8  
InsalataVerde 6    2     4     8  
UovoMcMuffin 20    20     25    15  
Wheaties    20    20     2     20  
Vaniglia    10    2     10     0  
Latte       30    0     30     0  
SuccoArancia 0    0     0     0  
SuccoPomelo  2    2     0     0  
SuccoMela   0    0     0     4 ;
```

Eseguendo il metodo del simplesso, si ottiene una soluzione ottimale di valore 6.126053691. Richiedendo la visualizzazione di alcuni parametri, otteniamo il seguente risultato:

10.7. Lettura dell'output dei codici di programmazione lineare

```
costo_totale = 6.12605

:      Qta.lb    Qta      Qta.ub    Qta.rc    :=
BigMac      0      0      Infinity   0.784998
ChefSalad    0      0      Infinity   0.519093
Hamburger    0      5.29919  Infinity   0
InsalataVerde 0      0.535076  Infinity   0
Latte        0      1.44149  Infinity   0
McLean       0      0      Infinity   0.880097
Miele         0      0      Infinity   0
PatateFritte 0      0      Infinity   0.543075
PolloFritto   0      0      Infinity   1.52974
SuccoArancia 0      0      Infinity   0.064
SuccoMela    0      0      Infinity   0.532466
SuccoPompelmo 0      0.380781  Infinity   0
UovoMcMuffin 0      0      Infinity   0.561258
Vaniglia     0      0      Infinity   0.484319
Wheaties     0      0.811574  Infinity   0
;

:      dieta.lb  dieta.body  dieta.ub    dieta      :=
Calcio       100     100      Infinity   0.0138042
Ferro        100     100      Infinity   0.0338561
Niacina      100     124.047  Infinity   0
Proteine     55      80.7079  Infinity   0
Sodio        0       3000     3000      -0.000298066
Vitamina-A  100     100      Infinity   0.0157422
Vitamina-C  100     100      Infinity   0.0068
VitaminaB1  100     138.481  Infinity   0
VitaminaB2  100     116.44   Infinity   0
```

Nella prima parte del listato, sotto la colonna Qta compare il valore ottimale dell'incognita Qta; le colonne Qta.lb e Qta.ub contengono il limiti inferiori e superiori imposti sulle incognite; l'ultima colonna, Qta.rc contiene i coefficienti di costo ridotto (rc: reduced costs) associati a ciascuna variabile.

Si noti che, come naturale, le variabili con valore non nullo hanno coefficiente di costo ridotto 0. Viceversa, le variabili alle quali è associato un coefficiente di costo ridotto positivo sono nulle. Possono esistere variabili a livello 0 con coefficiente di costo ridotto 0 (nel caso dell'esempio, il Miele). Questo è comprensibile: il Miele ha prezzo di vendita 0; non è detto che un cibo a prezzo nullo abbia

10. ANALISI DI SENSITIVITÀ

coefficiente di costo ridotto nullo: la formula è, ricordiamo,

$$\hat{c}_j = c_j - \sum_{i=1}^m a_{ij} \lambda_i,$$

quindi una variabile potrebbe essere “competitiva” rispetto ad altre non tanto a causa del prezzo, quanto grazie all’apporto di contributi nutritivi.

Nel caso del Miele, oltre ad essere $c_j = 0$ si vede che tutti i coefficienti a_{ij} sono nulli (l’unica eccezione sono le Calorie, ma nel modello non viene imposto un vincolo su di esse).

Per quanto riguarda i vincoli, la seconda parte del listato di output riporta le colonne `dieta.1b`, `dieta.body`, `dieta.ub` che corrispondono, in modo simile a quanto visto per le variabili, al limite inferiore, al valore attuale ed al limite superiore per ciascun vincolo. Cioè se nel problema è presente un vincolo del tipo

$$l \leq \sum_j a_j x_j \leq u$$

nella colonna con suffisso `1b` comparirà l , in quella con suffisso `ub` si avrà u , mentre nella colonna `body` verrà riportato il valore $\sum_j a_j \bar{x}_j$, dove \bar{x} è la soluzione (ottimale) corrente. Dal listato si vede, in particolare, che nella dieta di costo minimo alcune sostanze nutritive compaiono ad un livello superiore a quello minimo richiesto (ad esempio la `VitaminaB1`).

L’ultima colonna, `dieta`, rappresenta il valore della variabile ottimale del duale determinata dall’algoritmo risolutivo.

Dai teoremi di dualità si ottiene qualche interpretazione interessante. Ad esempio, il costo ottimale (minimo) della dieta, pari a €6.126053691, può essere facilmente ottenuto sommando i prodotti tra i livelli delle variabili ed il prezzo associato ai cibi:

$$5.29919 \times 0.59 + 0.535076 \times 1.96 + 1.44149 \times 0.56 + \dots$$

10.7. Lettura dell'output dei codici di programmazione lineare

Però, avendo studiato l'interpretazione economica del problema duale, possiamo immaginare di associare un costo agli elementi nutritivi: il prezzo "ottimale" è dato proprio dalla soluzione ottima del duale. In particolare potremmo spiegare la cifra totale di €6.126053691 come

$$100 \times 0.0138042 + 100 \times 0.0338561 - 3000 \times 0.000298066 + \\ 100 \times 0.0157422 + 100 \times 0.0068.$$

In altre parole, il prezzo totale è spiegabile in funzione di un prezzo associato alle richieste fatte: ogni unità (delle 100 richieste) di Calcio è costata €0.0138042, ogni unità di Ferro, €0.0338561; la Niacina non è costata nulla (è infatti sovrabbondante nella dieta). Osservando i valori della soluzione del duale, scopriamo che il costo unitario maggiore è quello del Ferro: infatti, dalla tabella dei dati nutrizionali, vediamo che il Ferro è assente in molti cibi e presente in quantità ridotta in altri. Occorre prestare attenzione al dato relativo al Sodio: a differenza degli altri elementi nutritivi, sul Sodio è imposto un limite *superiore*. Questo significa che tale elemento è considerato dannoso per la dieta. Nell'interpretazione economica, la presenza di un vincolo di \leq nel primale provoca un vincolo di segno sulla variabile corrispondente del duale, ma questa volta il vincolo è di *non positività*. Si può facilmente interpretare questo vincolo: essendo il Sodio un elemento considerato "dannoso", il produttore avrà la possibilità di venderlo solo ad un prezzo "negativo" (in altri termini, farà uno sconto sul prezzo totale a chi acquista anche del Sodio).

Dobbiamo tenere presente anche il fatto che i vincoli di questo problema sono "doppi": ciascuno corrisponde a 2 diseguaglianze. Usando la trasformazione in un problema standard, si otterrebbero quindi m vincoli di \geq ed m vincoli di \leq . Tuttavia le implementazioni efficienti del metodo del simplex lo evitano questa duplicazione di vincoli. Evidentemente in ogni soluzione ammissibile al più uno dei due vincoli può risultare attivo. Grazie alla proprietà degli scar-

10. ANALISI DI SENSITIVITÀ

ti complementari, in corrispondenza di una soluzione ottimale si potrà avere, nel duale, al più una variabile non nulla, associata al vincolo attivo. Se il vincolo attivo corrisponde alla disequazione col segno di \geq , tale variabile del duale sarà vincolata ad essere non negativa; viceversa per l'altra disequazione. Nell'output prodotto dai più comuni algoritmi di *PL*, ai vincoli con doppia limitazione, inferiore e superiore, viene quindi associata un'unica variabile del duale.

Ad esempio, nel caso del **Calcio** o del **Ferro**, lo slack non nullo è quello “destro” (ovviamente, essendo il limite superiore pari a $+\infty$): quindi la variabile duale associata al vincolo di \leq sarà nulla (e non viene stampata). Il valore 0.0138042 si riferisce quindi al vincolo **Ferro** ≥ 100 . Per quanto riguarda la **Niacina**, nessuno dei due vincoli è “attivo” (cioè entrambe le variabili, di slack e di surplus, sono diverse da 0): per la complementarietà entrambe le variabili del duale, nell'ottimo, saranno nulle. Per il **Sodio** invece, il vincolo ha la forma

$$0 \leq \text{Sodio} \leq 3000;$$

essendo il **Sodio** contenuto nella dieta ottimale pari a 3000, la variabile di surplus (vincolo di \geq) è diversa da zero e, quindi, la variabile del duale ad essa associata è nulla. Il valore -0.000298066 si riferisce pertanto al vincolo

$$\text{Sodio} \leq 3000.$$

Ricordando l'interpretazione del valore delle variabili del duale come derivate (nel caso tali derivate esistano) del costo ottimale rispetto a variazioni nei termini noti, si può anche osservare quanto segue: un aumento del contenuto di **Calcio** richiesto pari ad 1 unità (cioè pari a 101 in luogo di 100) porterebbe ad un aumento del costo della dieta *almeno* pari a 0.0138042 – esattamente pari a tale valore nel caso la base ottimale non cambi, ma questo è un evento

10.7. Lettura dell'output dei codici di programmazione lineare

difficilmente determinabile a partire dall'analisi dell'output, a meno che non sia disponibile la matrice di base. Un aumento della richiesta di Niacina, se non troppo elevato, non porta a modifiche del costo ottimale.

Un ragionamento simile può essere fatto anche nel primale, ricordando che, data la simmetria tra primale e duale, le variabili del primale ed i coefficienti di costo ridotto giocano lo stesso ruolo degli slack nel duale e dell'ottimo del duale.

Riprendendo i risultati visti prima, si osserva, ad esempio, che il cibo BigMac compare a livello 0 nella dieta ottimale; un eventuale aumento della richiesta minima, ad esempio l'imposizione di un limite inferiore pari ad 1, provocherebbe un aumento tendenziale del costo almeno pari a 0.784998 (sarà esattamente pari a questo valore nel caso in cui la base ottimale resti invariata). Se si modificano i dati, si ottiene in effetti il seguente risultato:

```
costo = 6.91105

:      Qta.lb   Qta    Qta.ub   Qta.rc   :=
BigMac     1       1    Infinity  0.784998
ChefSalad   0       0    Infinity  0.519093
Hamburger   0   3.31206    Infinity  0
InsalataVerde  0   0.46314    Infinity  0
Latte       0   1.24551    Infinity  0
McLean      0       0    Infinity  0.880097
Miele        0       0    Infinity  0
PatateFritte 0       0    Infinity  0.543075
PolloFritto  0       0    Infinity  1.52974
SuccoArancia 0       0    Infinity  0.064
SuccoMela   0       0    Infinity  0.532466
SuccoPompelmo 0   0.369458  Infinity  0
UovoMcMuffin 0       0    Infinity  0.561258
Vaniglia    0       0    Infinity  0.484319
Wheaties    0   1.3307   Infinity  0
;

:      dieta.lb  dieta.body  dieta.ub   dieta      :=
Calcio     100      100    Infinity  0.0138042
```

10. ANALISI DI SENSITIVITÀ

```
Ferro      100      100      Infinity   0.0338561
Niacina    100     129.52     Infinity   0
Proteine    55      80.8378     Infinity   0
Sodio       0      3000      3000     -0.000298066
Vitamina-A 100      100      Infinity   0.0157422
Vitamina-C 100      100      Infinity   0.0068
VitaminaB1 100     137.076     Infinity   0
VitaminaB2 100     125.618     Infinity   0
;
```

Il costo totale è salito a €6.91105, con un aumento di €0.785 che, a meno di arrotondamenti (dovuti solo al numero di cifre significative presentate in output, non a imprecisioni interne), è esattamente il valore del coefficiente di costo ridotto associato a BigMac. E' importante osservare che il costo ridotto *non* è il costo: l'ingresso di un'unità di BigMac nella dieta non provoca un aumento di costo di €1.65 (il prezzo di BigMac): il valore del coefficiente di costo ridotto è spiegabile tenendo conto del fatto che la presenza di 1 unità di BigMac permette di risparmiare in altri cibi. Confrontando le due tabelle ottimali:

```
:        dieta_1    dieta_2    :=
BigMac    0          1
ChefSalad 0          0
Hamburger 5.29919   3.31206
InsalataVerde 0.535076 0.46314
Latte     1.44149   1.24551
McLean    0          0
Miele     0          0
PatateFritte 0        0
PolloFritto 0        0
SuccoArancia 0        0
SuccoMela  0        0
SuccoPompelmo 0.380781 0.369458
UovoMcMuffin 0        0
Vaniglia   0        0
Wheaties   0.811574  1.3307
;
```

10.7. Lettura dell'output dei codici di programmazione lineare

si vede che la nuova dieta corrisponde ad una redistribuzione non banale dei cibi, con molti cibi presenti in quantità minore nella seconda dieta (a causa della presenza, con un'unità di BigMac, di varie componenti nutritive che prima erano fornite da altri cibi). Tuttavia qualche cibo ha aumentato il proprio livello (si veda, in particolare, il cibo Wheaties). Questo aumento è probabilmente spiegabile osservando che tra gli elementi nutritivi ne appare uno, il già citato Sodio, sul quale è imposto (ed è attivo) un vincolo di \leq : il cibo nuovo BigMac contiene molto Sodio e pertanto, per compensare, sarà necessario introdurre in maggiore quantità cibi poveri di tale elemento.

Si può quindi notare come il valore numerico corrispondente ad un coefficiente di costo ridotto non sia una quantità di banale valutazione: esso corrisponde ad un reale “indice di competitività” tra un prodotto ed un insieme di alternative.

Parte III

Ottimizzazione di flussi su reti

Introduzione alla teoria dei grafi ed alle reti di flusso

In questa parte verranno presentati alcune definizioni e proprietà tratte dalla teoria algoritmica dei grafi; non verranno presentate analisi astratte approfondite sulla teoria vera e propria dei grafi, ma verranno semplicemente introdotti alcuni concetti e proprietà importanti per comprendere l'utilità della rappresentazione mediante grafo, con particolare riferimento alla modellizzazione di problemi di flusso su rete.

11.1 Terminologia e definizioni principali

Si definisce *grafo* una coppia di insiemi

$$G = \langle V, E \rangle$$

dove V è un insieme di cardinalità finita detto insieme dei *vertici* o *nodi* del grafo e

$$E \subseteq V \times V$$

è un insieme, detto degli *archi lati*, i cui elementi sono coppie di nodi. Un grafo può essere quindi definito elencando l'insieme dei nodi e un insieme di coppie di nodi, detti archi.

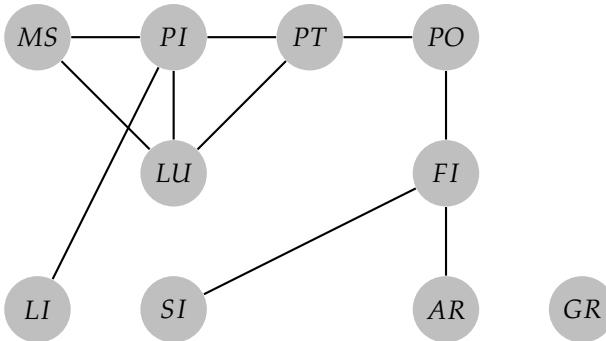


Figura 11.1: Rappresentazione grafica di un grafo

Esempio 31. Un esempio di grafo è il seguente:

$$V = \{FI, PI, MS, LI, GR, SI, AR, PO, PT, LU\}$$

$$E = \left\{ \begin{array}{l} (FI, PO), (FI, AR), (FI, SI), (PO, PT), (PT, PI), \\ (PT, LU), (LU, MS), (PI, LI), (PI, LU), (PI, MS) \end{array} \right\}$$

i cui nodi rappresentano alcune città della Toscana, mentre gli archi sono associati a collegamenti autostradali diretti tra coppie di città.

Uno dei motivi per i quali i grafi sono diventati un importantissimo strumento di modellazione è il fatto che essi ammettono una rappresentazione grafica (da cui il nome) molto intuitiva. Ad ogni nodo si associa un simbolo (spesso un cerchio o un quadrato) ed ogni arco viene rappresentato collegando con una linea i due nodi che lo rappresentano.

Esempio 31 (cont.).

Nell'esempio precedente, una possibile rappresentazione grafica è illustrata nella figura 11.1

Un arco è un sottoinsieme di nodi di cardinalità 2. In questo senso, l'arco (FI, PO) e l'arco (PO, FI) sono lo stesso arco. Nei mo-

11.1. Terminologia e definizioni principali

delli di flusso è invece molto più frequente il caso in cui gli archi siano orientati: in un arco orientato è importante distinguere tra il primo ed il secondo nodo. In altri termini $(FI, PO) \neq (PO, FI)$. Si può pensare agli archi orientati come a strade a senso unico, in cui è ben evidente il nodo di partenza ed il nodo di arrivo. La definizione di *grafo orientato*, o *digrafo*, differisce da quella di grafo solo per il fatto che gli archi sono coppie *ordinate* di nodi. In questo volume verranno considerati quasi esclusivamente grafi orientati e, pertanto, per brevità si parlerà semplicemente di grafo, intendendo in realtà un digrafo. Per rappresentare graficamente un grafo (orientato) si raffigureranno gli archi mediante frecce che vanno dal nodo origine dell'arco al nodo destinazione.

Esempio 32. La Figura 11.2 riporta un esempio di grafo orientato, una porzione di albero genealogico. Gli archi connettono un genitore con ciascun figlio.

I grafi (orientati) utilizzati per la trattazione dei problemi di flusso sono in genere grafi *semplici*, dove il termine semplice sta ad indicare grafi privi di archi multipli e di cappi. *Archi multipli* sono archi distinti fra loro che connettono la stessa coppia di nodi; un *cappio* è un arco del tipo (a, a) , dove a è un nodo. Graficamente, gli archi multipli si possono rappresentare come



mentre i cappi sono rappresentabili nel modo seguente:



11. INTRODUZIONE ALLA TEORIA DEI GRAFI ED ALLE RETI DI FLUSSO

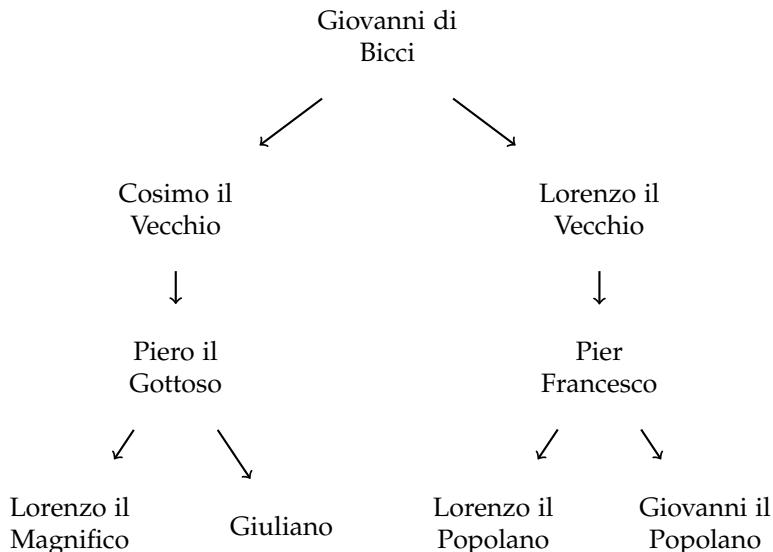


Figura 11.2: Rappresentazione grafica di un di-grafo

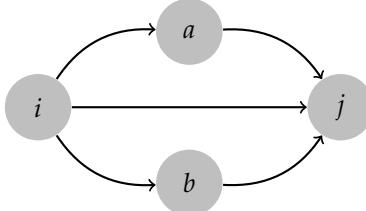
Si osservi che questi



non sono archi multipli. La restrizione ai grafi semplici porta ad alcune semplificazioni nel trattamento dei problemi di flusso e nella notazione; d'altra parte, anche se si potrebbe eliminare questa restrizione, essa non è particolarmente svantaggiosa, dal momento

11.1. Terminologia e definizioni principali

che sia gli archi multipli che i cappi possono essere trasformati in grafi equivalenti semplici mediante l'aggiunta di nodi e archi fittizi:



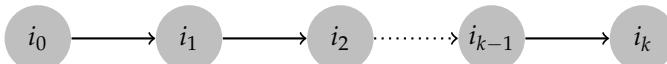
e, per i cappi,



Dato un grafo si dice che un suo arco (i, j) *incide* sui nodi i e j . I nodi i e j si dicono *adiacenti* se sono connessi direttamente da un arco. Il *grado* di un nodo è pari al numero di archi incidenti sul nodo; nel caso orientato si possono definire il *grado uscente* ed il *grado entrante* come il numero di archi orientati uscenti oppure entranti nel nodo stesso.

In un grafo orientato, si dice *cammino* dal nodo i_0 al nodo i_k una successione finita di archi

$$(i_0, i_1), (i_1, i_2), \dots, (i_{k-1}, i_k)$$



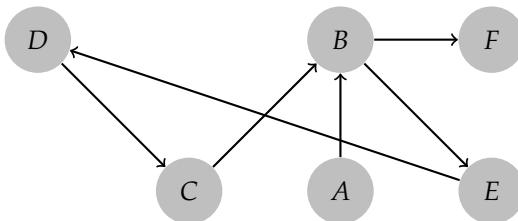
11. INTRODUZIONE ALLA TEORIA DEI GRAFI ED ALLE RETI DI FLUSSO

tale che il nodo terminale di un arco coincide con il nodo iniziale dell'arco successivo. Si dice invece *catena* una successione finita di archi che può essere trasformata in un cammino cambiando eventualmente il verso a qualche arco, cioè invertendo il nodo terminale dell'arco con il nodo iniziale. Il grafo seguente:

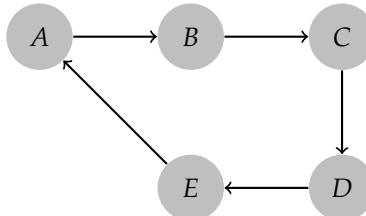


è una catena.

Un cammino od una catena si dicono *semplici* se nessun nodo o arco viene ripetuto. Un esempio di cammino *non semplice* è il seguente A,B,E,D,C,B,F:

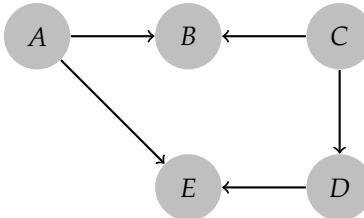


Un *circuito* è un cammino, costituito da almeno due archi distinti, nel quale il nodo iniziale i_0 coincide con quello finale i_k :



mentre un *ciclo* è simile ad un circuito, ma l'orientamento degli archi non segue necessariamente un unico verso:

11.1. Terminologia e definizioni principali



Una catena è quindi costituita da un insieme di archi orientati caratterizzati dal fatto che modificando l'orientamento di alcuni di essi si ottiene un cammino orientato.

Un grafo si dice *connesso* quando tra ogni coppia di nodi esiste almeno una catena; un grafo *orientato* si dice *fortemente connesso* se tra ogni coppia di nodi esiste almeno un cammino. Ad esempio, il grafo in Figura 11.3 è connesso, ma non in senso forte, come si

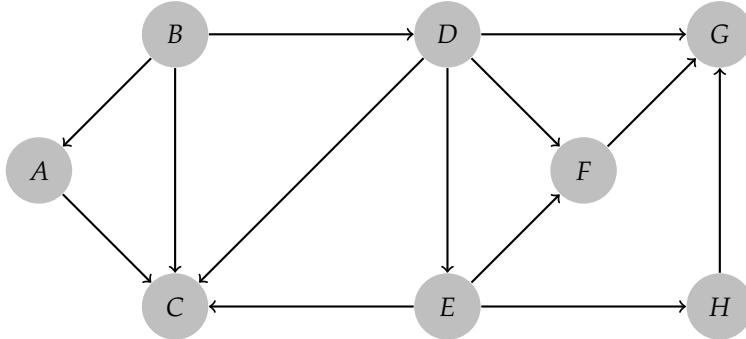


Figura 11.3: Un grafo connesso, ma non fortemente connesso

può vedere, ad esempio, dal nodo C , che ha grado uscente pari a zero e, pertanto, da C nessun nodo può essere raggiunto con un cammino orientato. Vi possono essere coppie di nodi sconnesse fra loro anche se il grado, uscente o entrante, non è zero: ad esempio, il nodo A ha grado entrante pari al grado uscente e pari ad 1 ed il

11. INTRODUZIONE ALLA TEORIA DEI GRAFI ED ALLE RETI DI FLUSSO

nodo E ha grado uscente 3 e grado entrante 1; tuttavia i nodi A ed E non sono connessi strettamente fra loro.

Per rappresentare un grafo si possono usare molte strutture dati differenti – uno stesso algoritmo può avere tempi di esecuzione notevolmente differenti in funzione del tipo di struttura dati utilizzata nella memorizzazione di un grafo. Una struttura elementare è quella vista nella definizione di grafo, una lista di archi e nodi. Una differente struttura è, ad esempio, la *lista di adiacenza*, in cui ad ogni nodo viene associata la lista degli archi uscenti dal nodo e la lista degli archi entranti. Nell'esempio di grafo sopra rappresentato, le liste di adiacenza sono date da

Nodo	Archi uscenti	Archi entranti
A	(A, C)	(B, A)
B	$(B, A), (B, C), (B, D)$	
C		$(A, C), (B, C), (D, C), (E, C)$
D	$(D, C), (D, E), (D, F), (D, G)$	(B, D)
...

Si può anche rappresentare la stessa informazione mediante la *matrice di adiacenza*, un matrice binaria $|V| \times |V|$ il cui elemento di posto (i, j) è pari ad 1 se e solo se nel grafo esiste l'arco (i, j) . Nell'esempio, omettendo gli elementi pari a zero, la matrice di incidenza ha la forma seguente:

	A	B	C	D	E	F	G	H
A				1				
B	1		1	1				
C								
D				1	1	1	1	
E					1			1
F						1		
G							1	
H								1

11.1. Terminologia e definizioni principali

La stessa matrice può essere usata anche per rappresentare grafi non orientati: in tale caso essa sarà simmetrica. Una rappresentazione molto utile sia nella pratica che nell'analisi teorica degli algoritmi su grafi è quella basata sulla *matrice di incidenza*. Tale matrice, adatta a rappresentare unicamente i grafi orientati e privi di cappi, è costituita da $|V|$ righe e $|E|$ colonne: ogni riga corrisponde ad un nodo ed ogni colonna è associata ad un arco. Dato un nodo v ed un arco e , l'elemento della matrice A di incidenza, corrispondente alla riga v ed alla colonna e è così definito:

$$A_{v,e} = \begin{cases} +1 & \text{se } e = (v, w) \text{ per qualche } w \in V \\ -1 & \text{se } e = (u, v) \text{ per qualche } u \in V \\ 0 & \text{altrimenti.} \end{cases}$$

Quindi nella matrice di incidenza, considerata la colonna associata ad un arco $e = (i, j)$, in tale colonna si troverà un elemento pari a $+1$ in corrispondenza della riga associata al nodo i , -1 in corrispondenza della riga associata al nodo j e 0 in tutte le altre posizioni.

La matrice di incidenza del grafo di pagina 257 è la seguente:

	AC	BA	BC	BD	DC	DE	DF	DG	EC	EF	EH	FG	HG
A	+	–											
B		+	+	+									
C	–	–	–	–								–	
D			–	+	+	+	+	+					
E				–					+	+	+		
F					–				–			+	
G						–					–	–	
H							–				–		+

Come si vede, in ogni colonna della matrice di incidenza esiste un unico elemento pari ad 1 ed un unico elemento pari a -1 .

Sommendo le righe di una matrice di incidenza si ottiene quindi il vettore nullo – le righe della matrice di incidenza, pertanto,

non sono linearmente indipendenti fra loro. In particolare una riga qualunque può essere cancellata senza che il contenuto informativo della matrice venga alterato.

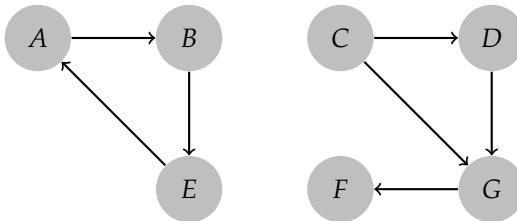
Come da un grafo si può passare alla matrice di incidenza, così da una matrice di incidenza si può ricostruire il grafo.

11.2 Alberi

Dato un grafo $G = \langle V, E \rangle$, un *sottografo* di G è un grafo $G' = \langle V', E' \rangle$ con

$$\begin{aligned} V' &\subseteq V \\ E' &\subseteq E \\ E' &\subseteq V' \times V' \end{aligned}$$

Una *componente连通* G' di un grafo G è un sottografo di G connesso e *massimale*, cioè è un sottografo di G connesso e tale che non esiste nessun sottografo connesso contenente propriamente G' . Chiaramente se un grafo è connesso esso contiene un'unica componente connessa, cioè il grafo stesso. Un grafo con n nodi e nessun arco ha invece n componenti connesse. Il grafo seguente



ha due componenti connesse, corrispondenti ai nodi A, B, E e C, D, F, G rispettivamente.

Definizione 22. Un *albero* è un grafo connesso e aciclico.

Esempio 33. Il grafo in Figura 11.4 è un albero.

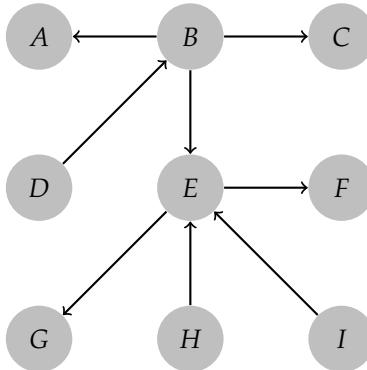


Figura 11.4: Un albero

Definizione 23. Dato un grafo $G = \langle V, E \rangle$, un suo sottografo che sia un albero e che abbia lo stesso insieme di nodi, $T = \langle V, E' \rangle$, con $E' \subseteq E$ si dice *albero di supporto* per G o *albero ricoprente*

Esempio 34. Nella Figura 11.5 viene rappresentato un grafo ed evidenziato un suo albero di supporto.

Dato un grafo orientato si dice *grado* di un nodo il numero di archi incidenti su quel nodo. In un grafo si chiama *foglia* un nodo di grado 0 oppure 1. Nell'albero disegnato in Figura 11.5, i nodi A, C, D, F, G, H, I sono foglie, tutte di grado 1.

Alcune proprietà degli alberi risulteranno utili per ulteriori caratterizzazioni:

Teorema 41. Si consideri un albero $T = \langle V, E \rangle$ con almeno 2 nodi. Il grafo ottenuto eliminando un qualsiasi arco (i, j) è composto da due componenti connesse, ciascuna delle quali è un albero.

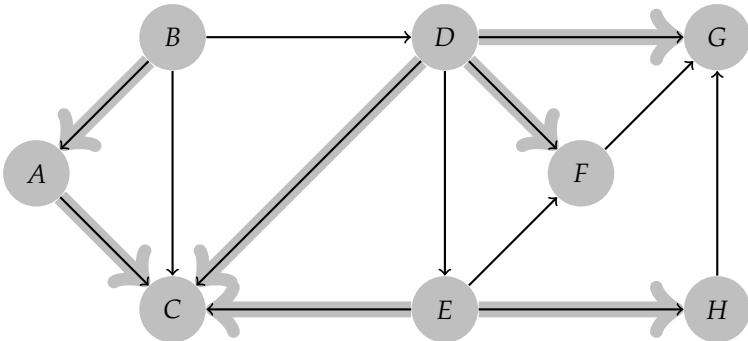


Figura 11.5: Un grafo ed un suo albero di supporto

Dimostrazione. Se da un albero si elimina un arco qualsiasi (i,j) , l'albero si sconnette: infatti se rimanesse connesso dovrebbe esistere una catena tra i e j diversa dall'arco (i,j) ; ma in questo caso, rimettendo l'arco al suo posto si genererebbe un ciclo, cosa impossibile essendo T un albero. Quindi la rimozione di un arco provoca la formazione di almeno due componenti connesse. Le componenti connesse sono in realtà esattamente due, poiché, nuovamente, la reinserzione dell'arco (i,j) genera un grafo connesso; ma l'inserzione di un arco può al massimo connettere fra loro due componenti. Quindi si è dimostrato che eliminando un arco da un albero si generano sempre esattamente due componenti connesse. Il fatto che siano entrambe acicliche, e quindi alberi, è banale, essendo aciclico ogni sottografo di un grafo aciclico. \square

Teorema 42. *Un albero $T = \langle V, E \rangle$ con almeno 2 nodi ha almeno due foglie.*

Dimostrazione. Un albero con $m = 2$ nodi ha evidentemente due foglie. Ipotizziamo che la proprietà valga per un generico grafo con $m \geq 2$ nodi e dimostriamolo per un grafo con $m + 1$ nodi. Si scelga

un arco qualunque e si elimini dal grafo. Per quanto appena visto, il grafo risultante ha due componenti connesse, entrambe alberi. Se una delle due componenti ha un solo nodo, questo è una foglia; l'altra componente avrà, per ipotesi di induzione, almeno due foglie. Rimettendo al suo posto l'arco eliminato si perde al più una foglia e quindi il teorema vale. Se invece entrambe le componenti connesse hanno più di un nodo, dopo l'eliminazione dell'arco si hanno in totale almeno 4 foglie. Di nuovo, la reimmissione dell'arco tolto può al massimo provocare la perdita di due foglie. \square

Teorema 43. *Un albero con m nodi ha $m - 1$ archi.*

Dimostrazione. Il teorema vale banalmente per $m = 1$ e per $m = 2$. Si ipotizzi il teorema valido per un generico albero con $m - 1$ nodi. Si consideri un albero con m nodi. Grazie alla proprietà precedente, questo albero ha sicuramente almeno una foglia. Esisterà un unico arco incidente sulla foglia. Disconnettendo tale arco, l'albero si sconnette in due componenti, una delle quali è un albero con $m - 1$ nodi e l'altra è un albero ridotto ad 1 nodo. L'albero con $m - 1$ nodi per ipotesi di induzione ha $m - 2$ archi; la componente costituita da un solo nodo non ha archi. Rimettendo a posto l'arco eliminato si rigenera l'albero e si possono contare in totale $(m - 2) + 1 = m - 1$ archi. \square

E' possibile caratterizzare gli alberi in molti modi diversi:

Teorema 44. *Si consideri un grafo $T = \langle V, E \rangle$ e sia $|V| = m$, $|E| = n$. Le seguenti affermazioni sono equivalenti:*

1. *T è un albero, cioè è connesso e privo di cicli;*
2. *T è connesso ed ha $n = m - 1$ archi;*
3. *T è privo di cicli ed ha $n = m - 1$ archi;*

11. INTRODUZIONE ALLA TEORIA DEI GRAFI ED ALLE RETI DI FLUSSO

4. *T è connesso, ma l'eliminazione di un suo arco qualsiasi provoca la sconnessione in due componenti;*
5. *T è aciclico, ma l'aggiunta di un qualsiasi arco provoca la formazione di uno ed un solo ciclo;*
6. *Esiste una ed una sola catena che connette ciascuna coppia di nodi di T.*

Dimostrazione. Si può iniziare mostrando che 1, 2 e 3 sono equivalenti:

$1 \implies 2$ deriva immediatamente dal teorema 43

$2 \implies 3$ se per assurdo T , connesso con $m - 1$ archi, possedesse un ciclo, allora sarebbe possibile eliminare un arco appartenente al ciclo senza perdere la connessione. Si genererebbe così un grafo connesso, aciclico, con $m - 2$ archi. Ma questo è in contrasto con il teorema 43

$3 \implies 1$ infatti se T non fosse connesso, allora si potrebbero aggiungere archi che connettono a due a due le componenti del grafo, senza creare cicli, fino ad arrivare ad un'unica componente. A questo punto si avrebbe un grafo connesso e aciclico, cioè un albero, ma con più di $m - 1$ archi, il che è assurdo.

Quindi le proposizioni 1, 2, 3 sono equivalenti.

$1 \implies 4$ è il teorema 41

$4 \implies 5$ se esistesse un ciclo, l'eliminazione di un arco appartenente al ciclo non provocherebbe la sconnessione; aggiungendo un arco qualunque sicuramente si perde l'aciclicità, dato che i due nodi connessi dall'arco aggiunti possono essere raggiunti attraverso due catene diverse. Il fatto che si generi un unico ciclo è dovuto al fatto che se se ne generassero due, i due

cicli avrebbero in comune l'arco aggiunto. Eliminando l'arco aggiunto, resterebbero due diverse catene che connettono gli estremi di tale arco: il grafo quindi risulterebbe ciclico;

5 \implies 6 si supponga che T sia aciclico: nessuna coppia di nodi può essere connessa da più di una catena. Infatti se una coppia di nodi fosse connessa da due catene, queste formerebbero un ciclo. Inoltre, tra ogni coppia di nodi esiste certamente una catena (e quindi il grafo è connesso): infatti se tra due nodi i e j non esistesse una catena, l'aggiunta di un arco (i, j) al grafo non provocherebbe la nascita di un ciclo, contrariamente all'ipotesi.

6 \implies 1 Se tra ogni coppia di nodi esiste una catena, per definizione il grafo è connesso; se inoltre tale catena è unica, il grafo è anche aciclico.

Si è quindi dimostrato che 1, 4, 5, 6 sono equivalenti fra loro, e questo completa la dimostrazione del teorema. \square

Le prime tre caratterizzazioni si possono interpretare dicendo che date due delle tre proprietà: connessione, aciclicità, numero di archi pari al numero di nodi meno uno, la terza è una conseguenza.

11.3 Flussi su reti – reti di flusso

Una *rete di flusso* è una struttura basata su un grafo orientato che permette la modellazione di flussi a regime stazionario. Si può pensare, come esempio di applicazione, allo studio di reti di flusso di veicoli in un'area urbana, al flusso dell'acqua potabile nella rete idrica di una città, al flusso di informazione lungo una rete di telecomunicazione, al flusso di materiali in un impianto di produzione; modelli di flusso vengono anche usati per rappresentare problemi di connessione, problemi di copertura di servizi, ecc. Nelle reti di

flusso generalmente si associa ad ogni arco una quantità variabile che rappresenta il flusso sull'arco; ai nodi è associata una quantità che rappresenta la domanda netta di flusso o l'offerta netta di flusso (un nodo può essere produttore, consumatore o semplicemente di transito per il flusso). Sugli archi è in genere definita una quantità positiva che rappresenta il massimo flusso passante lungo l'arco stesso. Formalmente, si definisce *rete di flusso* una struttura del tipo

$$N = \langle V, E, b, cap, c \rangle$$

dove:

- $\langle V, E \rangle$ è un grafo orientato;
- $b : V \rightarrow \mathbb{R}$: ad ogni nodo della rete $v \in V$ viene associata una quantità b_v , detta *bilancio*, che rappresenta la differenza fra flusso uscente dal nodo e flusso entrante;
- $cap : E \rightarrow \mathbb{R}^+$: ad ogni arco della rete è associata una *capacità massima*, una quantità cioè pari al massimo flusso che, nell'unità di tempo, può attraversare l'arco.
- $c : E \rightarrow \mathbb{R}$: ad ogni arco è associato un costo unitario, corrispondente al passaggio di un'unità di flusso dall'arco.

Si definisce *problema del flusso di costo minimo* il seguente:

$$\begin{aligned} & \min_{f \in \mathbb{R}^{|E|}} \sum_{(i,j) \in E} c_{ij} f_{ij} \\ & \sum_{k:(i,k) \in E} f_{ik} - \sum_{h:(h,i) \in E} f_{hi} = b_i \quad \forall i \in V \quad (11.1) \\ & f_{ij} \leq cap_{ij} \quad \forall (i,j) \in E \\ & f_{ij} \geq 0 \quad \forall (i,j) \in E. \end{aligned}$$

La (11.1) rappresenta la *legge di conservazione del flusso*, o *legge di Kirchoff*: la differenza fra flusso uscente e flusso entrante in ciascun

nodo i deve essere pari ad una quantità prefissata b_i . Il problema del flusso di costo minimo è dunque un problema di programmazione lineare. Per rappresentare il problema in forma matriciale, si osservi che la matrice di incidenza contiene, nella riga corrispondente ad un generico nodo i , elementi pari a $+1$ associati agli archi uscenti e pari a -1 per gli archi entranti. Il vincolo di conservazione del flusso può pertanto essere espresso come

$$Af = b$$

dove A è la matrice di incidenza del grafo, f il vettore dei flussi, b il vettore dei bilanci associati ad ogni nodo. Se con c si indica il vettore dei costi associati agli archi, l'intero problema può quindi essere rappresentato come

$$\begin{aligned} & \min c^T f \\ & Af = b \\ & f \leq cap \\ & f \geq 0. \end{aligned}$$

Basi e soluzioni di base nei problemi di flusso

Sia dato un grafo $G = \langle V, E \rangle$ con $|V| = m$ e $|E| = n$, e sia A la matrice di incidenza (nodi–archi). Si consideri un problema di flusso di costo minimo

$$\begin{aligned} & \min c^T f \\ & Af = b \\ & f \geq 0 \end{aligned}$$

in cui si è supposto che la capacità minima di ogni arco sia 0 e quella massima sia $+\infty$; il vettore b rappresenta il bilancio di ogni nodo, pari alla differenza tra flusso uscente da ciascun nodo e flusso entrante. Il generico problema con vincoli sul flusso sia inferiori che

superiori può essere trattato come estensione, non elementare, del problema qui considerato. Si è già visto che la somma delle righe della matrice di incidenza A è pari al vettore nullo. Si ipotizza quindi che $\sum_{i \in V} b_i = 0$, poiché se così non fosse il sistema sarebbe certamente privo di soluzioni.

Se i termini noti hanno somma paria a zero, segue che sicuramente almeno una delle equazioni $Af = b$ è ridondante e può essere eliminata senza che l'insieme ammissibile del problema vari. Questo significa anche che, ammesso che esistano soluzioni ammissibili, non potranno certamente esistere soluzioni di base. Il rango della matrice di incidenza non può essere superiore a $|V| - 1$. Scelto comunque un nodo v , sia $A_{\setminus v}$ la matrice di incidenza da cui è stata cancellata la riga corrispondente a v e sia $b_{\setminus v}$ il vettore b modificato in modo analogo. Ricordando che la matrice A ha $|V|$ righe ed $|E|$ colonne, vale la seguente proprietà fondamentale:

Teorema 45. *La matrice A di incidenza di un grafo $G = \langle V, E \rangle$ ha rango pari a $|V| - 1$ se e solo se il grafo è connesso.*

La dimostrazione di questo importante teorema si basa su altre proprietà, altrettanto fondamentali:

Teorema 46. *Un grafo è connesso se e solo se possiede un albero di supporto.*

Dimostrazione. Dato un grafo connesso è possibile eliminare un arco alla volta, mantenendo sempre la connessione; dopo un numero finito di iterazioni si ottiene un grafo connesso nel quale è impossibile eliminare un arco senza perdere la connessione: questa condizione è necessaria e sufficiente perché il grafo risultante sia un albero.

La dimostrazione dell'implicazione in senso opposto è banale: se un grafo contiene un albero di supporto, è certamente connesso. \square

Teorema 47. *La matrice di incidenza di un albero $T = \langle V, E_T \rangle$ con $|V| \geq 2$ ha rango pari a $|V| - 1$; inoltre tale matrice, privata di una riga qualsiasi, può essere scritta in forma triangolare con elementi non nulli sulla diagonale.*

Dimostrazione. Sia A_T la matrice di incidenza di incidenza dell’albero T ; per una proprietà degli alberi, A_T ha $|V|$ righe e $|V| - 1$ colonne. Sia $v \in V$ un nodo qualsiasi. La procedura per la costruzione della sotto-matrice $|V| - 1 \times |V| - 1$ triangolare è la seguente. Per una proprietà degli alberi, essendo $|V| \geq 2$ sicuramente esistono almeno due foglie in T . Sia $w \neq v$ una foglia di T , distinta da v . Essendo w una foglia esisterà uno ed un solo arco incidente in w ; se si inserisce il nodo w nella prima riga della matrice di incidenza e l’arco incidente in w come prima colonna della matrice si ottiene una struttura del tipo:

$$A_T = \begin{array}{|c|c|} \hline \pm 1 & 0^T \\ \hline p & A_{T'} \\ \hline \end{array}$$

dove con p si è indicato un vettore che ha tutte le componenti pari a zero tranne una in corrispondenza del secondo nodo su cui incide l’arco associato alla prima colonna. La matrice $A_{T'} \in \mathbb{R}^{|V|-1 \times |V|-2}$ è la matrice di incidenza del sottografo di T ottenuto trascurando il nodo w e l’arco incidente su w . Tale sottografo, essendo ottenuto da un albero per eliminazione di una foglia e dell’unico arco incidente su tale foglia è, a sua volta, un albero, ma con $|V| - 1$ nodi. A tale albero T' può essere applicato lo stesso ragionamento e, cioè, si può individuare una foglia non coincidente con v ed associarla alla prima riga della matrice $A_{T'}$, cioè alla seconda riga di A_T . Si vede pertanto che dopo $|V| - 1$ iterazioni si giungerà ad una matrice le cui prime $|V| - 1$ righe formano una matrice triangolare.

Al termine di questa procedura si saranno trascritte $|V| - 1$ righe della matrice di incidenza dell’albero corrispondenti ad una matrice triangolare. Per costruzione, tutti gli elementi sulla diagonale di

tale matrice sono non nulli. Poiché il determinante di una matrice triangolare è pari al prodotto degli elementi sulla diagonale, segue che questa matrice $|V| - 1 \times |V| - 1$ è invertibile e, pertanto, il rango della matrice dell'albero è pari a $|V| - 1$. \square

La costruzione della matrice triangolare corrisponde quindi ad una visita del grafo a partire dai nodi foglia verso il nodo (arbitrario) v , detto *radice*.

Teorema 48. *Se la matrice di incidenza di un grafo ha rango $|V| - 1$, allora gli archi corrispondenti alle colonne di una base della matrice di incidenza di G (privata di una riga qualsiasi) formano un albero di supporto per G .*

Dimostrazione. Si consideri una sottomatrice di A di rango $|V| - 1$ e siano $A_{e_1}, A_{e_2}, \dots, A_{e_{|V|-1}}$ le sue colonne, corrispondenti ad un sottografo di G . In tale sottografo non può esistere alcun ciclo; infatti, se in un grafo qualsiasi esistesse un ciclo *orientato*, cioè un circuito, come ad esempio

$$\mathcal{C} = \{v_0, v_1, \dots, v_k, v_{k+1} = v_0\},$$

composto dagli archi $(v_0, v_1), \dots, (v_k, v_0)$, osservando che la colonna della matrice di incidenza corrispondente ad un arco (i, j) può essere scritta come $e_i - e_j$, dove e_i è l' i -esimo versore, sommando le colonne della matrice di incidenza del ciclo si otterebbe:

$$(e_{v_0} - e_{v_1}) + (e_{v_1} - e_{v_2}) + \dots + (e_{v_{k-1}} - e_{v_k}) + (e_{v_k} - e_{v_0}) = 0$$

Si è quindi dimostrato che le colonne della matrice di incidenza associate ad un ciclo sono linearmente dipendenti (una loro combinazione lineare con coefficienti pari ad 1 fornisce il vettore nullo). Se il ciclo invece non fosse un circuito basterebbe scegliere un verso di percorrenza qualsiasi e moltiplicare per $+1$ gli archi percorsi "in avanti" e per -1 (il che equivale in un certo senso a cambiarne

il verso) gli archi percorsi all’indietro. Si otterrebbe anche in questo caso una combinazione lineare non banale che genera il vettore nullo.

Si è quindi dimostrato che se un sottoinsieme delle colonne di una matrice di incidenza è formato da vettori linearmente indipendenti, il grafo corrispondente è aciclico; nel caso di un sottoinsieme di $|V| - 1$ colonne tale grafo, essendo aciclico e composto da $|V| - 1$ archi, sarà un albero di supporto. \square

Grazie a questi teoremi si può ora dimostrare la proprietà principale (teorema 45):

Dimostrazione. Dal teorema 46 segue che, se G è connesso, allora possiede un albero di supporto; dal teorema 47 scende che tale albero di supporto corrisponde ad una matrice di incidenza, sottomatrice della matrice di incidenza A del grafo, di rango $|V| - 1$. Quindi si è dimostrata la condizione sufficiente del teorema.

Per quanto riguarda la condizione necessaria, se il rango della matrice A è pari a $V - 1$, il teorema 48 garantisce l’esistenza di un albero di supporto. Ma grazie al teorema 46 segue che il grafo G è connesso. \square

Dai teoremi appena dimostrati segue quindi che vi è corrispondenza biunivoca tra alberi di supporto di un grafo connesso e matrici $(|V| - 1, |V| - 1)$ invertibili nella matrice di incidenza, privata di una riga qualsiasi.

Il teorema 47, assieme al teorema 48, implica un’altra importante proprietà dei problemi di flusso a costo minimo, e cioè che se il termine noto b nella (11.1) è *intero*, allora ogni soluzione di base del problema di *PL* corrispondente sarà intera. Infatti, poiché ogni soluzione di base si ottiene risolvendo un sistema di equazioni del tipo

$$A_B f = b$$

11. INTRODUZIONE ALLA TEORIA DEI GRAFI ED ALLE RETI DI FLUSSO

e poiché A_B può essere portata in forma triangolare superiore con elementi sulla diagonale tutti pari a ± 1 , risolvendo il sistema partendo dall'ultima equazione si vede immediatamente che tutte le componenti di f saranno combinazioni lineari a coefficienti interi di d .

Esempio 35. Si consideri il problema di flusso a costo minimo caratterizzato dal grafo nella figura 11.6, in cui la quantità indicata sugli archi rappresenta il costo associato all'arco.

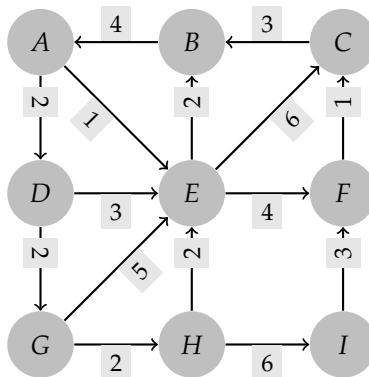


Figura 11.6: Un problema di flusso su reti a costo minimo

Si suppone che il vettore dei bilanci b sia definito da $b(A) = +8, b(E) = -3, b(I) = -5$ e $b(v) = 0$ per ogni altro nodo v . Come si vede, il grafo è connesso. Il sottografo evidenziato nella Figura 11.7 è un albero di supporto.

La sottomatrice della matrice di incidenza corrispondente a tale

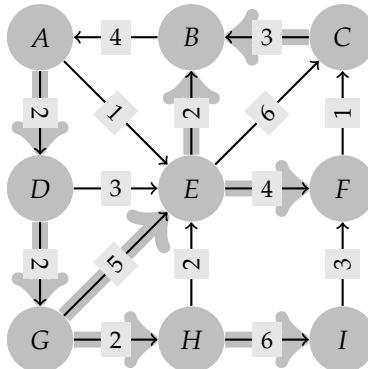


Figura 11.7: Un grafo connesso ed un suo albero di supporto

albero è data da

	AD	DG	CB	EB	EF	GE	GH	HI
A	+1	0	0	0	0	0	0	0
B	0	0	-1	0	0	0	0	0
C	0	0	+1	-1	0	0	0	0
D	-1	+1	0	0	0	0	0	0
E	0	0	0	+1	+1	-1	0	0
F	0	0	0	0	-1	0	0	0
G	0	-1	0	0	0	+1	+1	0
H	0	0	0	0	0	0	-1	+1
I	0	0	0	0	0	0	0	-1

Eliminando una riga da questa matrice, ad esempio la riga corrispondente ad A , si ottiene una matrice quadrata. Il teorema 47 garantisce che questa matrice è una base. Tale teorema afferma che è possibile trasformare in forma triangolare superiore tale sottomatrice. La tecnica per ottenere tale trasformazione, indicata dal procedimento per induzione usato nella dimostrazione, consiste nell'iniziare ad elencare i nodi partendo da una foglia e aggiungendo

l'unico arco su di essa incidente. E' possibile iniziare da una qualsiasi foglia, purché non coincidente con il nodo associato alla riga che si è eliminata dalla matrice di incidenza.

Ad esempio, si può iniziare con la foglia I e l'arco (H, I) incidente su di essa; immaginando di eliminare tale foglia e l'arco corrispondente si otterrebbe un albero con un nodo in meno. A questo punto si prosegue in modo ricorsivo, analizzando il sottoalbero privo della foglia I . Anche in questo sottoalbero esisterà almeno una foglia (non coincidente con A); si può, ad esempio, scegliere H , che ora, grazie all'eliminazione di I , è una foglia del grafo riordotto. Proseguendo si potrebbe ottenere la seguente sequenza di foglie: H, F, C, B, E, G, D . Si ottiene quindi la matrice riordinata e triangolare:

	HI	GH	EF	CB	EB	GE	DG	AD
I	-1	0	0	0	0	0	0	0
H	+1	-1	0	0	0	0	0	0
F	0	0	-1	0	0	0	0	0
C	0	0	0	+1	0	0	0	0
B	0	0	0	-1	-1	0	0	0
E	0	0	+1	0	+1	-1	0	0
G	0	+1	0	0	0	+1	-1	0
D	0	0	0	0	0	0	+1	-1

Osservando poi che il vettore dei bilanci, riordinato coerentemente con le righe della matrice di incidenza, ha la forma

$$b^T = [-5 \ 0 \ 0 \ 0 \ 0 \ 0 -3 \ 0 \ 0]$$

risolvendo il sistema

$$\begin{array}{|c|cccccccc|} \hline & HI & GH & EF & CB & EB & GE & DG & AD \\ \hline I & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ H & +1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ F & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ C & 0 & 0 & 0 & +1 & 0 & 0 & 0 & 0 \\ B & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 \\ E & 0 & 0 & +1 & 0 & +1 & -1 & 0 & 0 \\ G & 0 & +1 & 0 & 0 & 0 & +1 & -1 & 0 \\ D & 0 & 0 & 0 & 0 & 0 & 0 & +1 & -1 \\ \hline \end{array} \quad \begin{bmatrix} f_{HI} \\ f_{GH} \\ f_{EF} \\ f_{CB} \\ f_{EB} \\ f_{GE} \\ f_{DG} \\ f_{AD} \end{bmatrix} = \begin{bmatrix} -5 \\ 0 \\ 0 \\ 0 \\ 0 \\ -3 \\ 0 \\ 0 \end{bmatrix}$$

si determina immediatamente la soluzione di base corrispondente:

$$\begin{aligned}
 f(HI) &= 5 \\
 f(GH) &= 0 + f(HI) = 5 \\
 f(EF) &= 0 \\
 f(CB) &= 0 \\
 f(EB) &= 0 \\
 f(GE) &= 3 + f(EB) + f(EF) = 3 \\
 f(DG) &= f(GE) + f(GH) = 8 \\
 f(AD) &= f(DG) = 8
 \end{aligned}$$

che è una soluzione ammissibile di base, rappresentata in Figura 11.8. Il costo di questa soluzione è pari a $8 \cdot 2 + 8 \cdot 2 + 5 \cdot 2 + 3 \cdot 5 + 5 \cdot 6 = 87$.

Possono esistere altre soluzioni di base ammissibili, alcune delle quali con costo inferiore a quello della soluzione qui determinata. Non è detto che un albero di supporto corrisponda sempre ad una base *ammissibile*. Ad esempio, sostituendo l'arco (H, I) con l'arco (I, F) , si otterrebbe un nuovo albero, ma non una soluzione ammissibile.

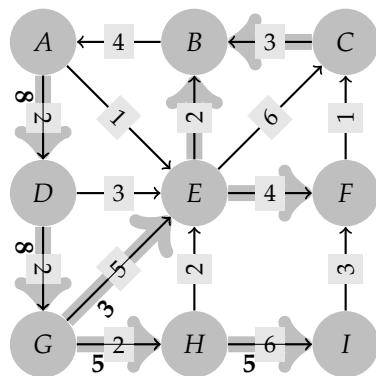


Figura 11.8: Soluzione ammissibile di base

L'algoritmo del simplex su reti

La struttura particolare della matrice dei coefficienti, nonché le forti proprietà che caratterizzano la matrice di incidenza di un grafo, rendono possibile l'implementazione di versioni del metodo del simplex specializzate per i problemi di flusso su reti. La presentazione di una versione del metodo del simplex specializzata per i problemi di flusso su reti è utile non solo per illustrare un'importante variante del metodo classico, ma anche per fornire spunti di riflessione sul metodo stesso del simplex. Si vedrà anche come alcuni classici algoritmi di ottimizzazione su grafi altro non sono che una versione semplificata dell'algoritmo del simplex su reti (Bazaraa et al., 1990; Ahuja, Magnanti, & Orlin, 1993)

Per poter meglio seguire lo sviluppo dell'algoritmo del simplex su reti, ricordiamo qui brevemente i passi del metodo standard del simplex:

1. Sia B una base ammissibile;
2. si determini $\bar{x}_B = A_B^{-1}b$, cioè si risolva il sistema

$$A_B x_b = b \tag{12.1}$$

3. si determini $\bar{\lambda}_B^T = c_B^T A_B^{-1}$, cioè si risolva il sistema

$$\lambda^T A_B = c_B^T \tag{12.2}$$

12. L'ALGORITMO DEL SIMPLEXO SU RETI

4. si calcolino i coefficienti di costo ridotto

$$\hat{c}^T = c_N^T - \bar{\lambda}_B^T A_N \quad (12.3)$$

5. se $\hat{c} \geq 0$ l'algoritmo termina: \bar{x}_B è una soluzione ottimale.

Altrimenti, sia $j : \hat{c}_j < 0$

6. si calcolino i coefficienti $d = -A_B^{-1}A_j$, cioè si risolva il sistema

$$A_B d = -A_j \quad (12.4)$$

7. se $d \geq 0$ il problema è illimitato; altrimenti:

8. si determini $\bar{i} \in \arg \min\left\{-\frac{x_{B_i}}{d_i} : d_i < 0\right\}$

9. si scambino la variabile di base \bar{i} e la variabile fuori base j

Come si vede dallo schema del metodo del simplex, alcune operazioni fondamentali, quali ad esempio la risoluzione di sistemi lineari la cui matrice di coefficienti è la base corrente, si ripetono spesso. Nel seguito illustreremo come le operazioni del metodo del simplex possono essere specializzate per il problema di flusso su reti a costo minimo seguente:

$$\min c^T f \quad (12.5)$$

$$Af = b$$

$$f \geq 0$$

dove, dato un grafo $G = \langle V, E \rangle$, c ed f sono vettori in $\mathbb{R}^{|E|}$, A è la matrice di incidenza nodi-archi del grafo, $b \in \mathbb{R}^{|V|}$ è il vettore delle offerte nette di ciascun nodo, cioè la differenza tra flusso totale uscente ed entrante in ciascun nodo. Si può sempre supporre che il grafo G sia connesso, poiché, in caso contrario, il problema si potrebbe decomporre in tanti sotto-problemi indipendenti uno dall'altro, uno per ciascuna componente connessa del grafo. Il modello

12.1. Duale del problema di flusso su reti a costo minimo

di flusso sopra riportato non prevede capacità massime ai flussi sugli archi, né capacità minime diverse da zero. Sarebbe possibile presentare una versione del metodo del simplex per reti con capacità limitata superiormente o per il caso di limite inferiore diverso da zero per i flussi, ma per far questo occorrerebbe prima presentare la versione del metodo del simplex per problemi con variabili limitate: tale presentazione esula dagli scopi di questo volume.

12.1 Duale del problema di flusso su reti a costo minimo

Il problema (12.5) è in forma standard e, pertanto, il suo duale è di scrittura immediata:

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

Poiché A rappresenta la matrice di incidenza di un grafo, ogni sua colonna ha solo due elementi non nulli, in corrispondenza delle righe associate ai due nodi estremi di ciascun arco. Si può quindi scrivere il duale esplicitamente come

$$\begin{aligned} \max \quad & \sum_{(i,j) \in E} b_{ij} \lambda_{ij} \\ \text{s.t.} \quad & \lambda_i - \lambda_j \leq c_{ij} \quad \forall (i,j) \in E \end{aligned}$$

Nel caso particolare della formulazione del problema del cammino di costo minimo come problema di PL ,

$$\begin{aligned} \min \quad & c^T f \\ \text{s.t.} \quad & Af = e_s - e_t \\ & f \geq 0, \end{aligned}$$

il duale ha la forma

$$\begin{aligned} \max \lambda_s - \lambda_t \\ \lambda_i - \lambda_j \leq c_{ij} \quad \forall (i, j) \in E \end{aligned}$$

Questo problema ammette un'importante interpretazione: se le incognite λ sono interpretate come "potenziali" (ad esempio, elettrici) associati ai nodi, l'obiettivo si riduce alla massimizzazione della differenza di potenziale tra sorgente e terminale, mentre i vincoli corrispondono alla richiesta che, agli estremi di ogni arco, la differenza di potenziale non superi una data soglia. Nel caso delle reti elettriche, le differenze di potenziale assumono il nome di *tensioni*, ed il problema diviene equivalente a quello di determinare la tensione massima tra gli estremi di un circuito, avendo imposto vincoli sulle tensioni massime ai capi di ogni ramo. Si può notare come esista un parallelo tra la dualità nella programmazione lineare e la dualità, nelle reti elettriche, tra correnti e tensioni. Nel caso delle reti elettriche, il problema del cammino di costo minimo può essere visto come il problema di determinare il percorso di minima resistenza incontrato da un flusso unitario di corrente tra due poli (sorgente e terminale). La somma delle resistenze incontrate lungo tale cammino è pari alla massima differenza di potenziale applicabile agli estremi del circuito.

Un'altra possibile interpretazione nasce da un modello "fisico" del problema: si immagini di costruire una rappresentazione del grafo mediante corde di lunghezza proporzionale al costo, collegate fra loro in corrispondenza dei nodi. Si immagini poi di distanziare fra loro al massimo i due nodi corrispondenti a sorgente e destinazione: durante tale operazione, la distanza tra gli estremi di ciascuna spezzata di corda non potrà mai superare la lunghezza della corda stessa. In questo caso la variabile λ_i rappresenta la coordinata, in un sistema di riferimento mono-dimensionale, del nodo i . Si vede facilmente, anche pensando al modello fisico, che gli archi "in tensione" nella soluzione ottimale sono quelli che corrispon-

12.2. Determinazione di una soluzione di base

dono al cammino di costo minimo: questa è un'altra illustrazione delle relazioni di scarti complementari: sugli archi del cammino minimo, ove passa un flusso non nullo, i vincoli del duale saranno attivi: $\lambda_i - \lambda_j = c_{ij}$. Un'altra caratteristica che si può osservare in questa coppia primale/duale è che nel duale le incognite appaiono sempre attraverso la loro differenza; sarà quindi sempre possibile aumentarle o diminuirle tutte della stessa quantità senza perdere né l'ammissibilità né l'ottimalità: si può pensare che una variabile sia ridondante e possa essere fissata ad un valore arbitrario. Questo fatto nel paragone elettrico corrisponde all'ipotesi di fissare un riferimento arbitrario, massa o terra, e misurare le tensioni rispetto a questo valore. Si può sempre immaginare di fissare arbitrariamente (ad esempio, a zero) il valore di una variabile qualsiasi; le altre variabili duali, in questo caso, assumono un valore pari alla differenza con il valore di riferimento. L'arbitrarietà di una variabile del duale è conseguenza del fatto che nel primale un vincolo qualunque può essere eliminato, in quanto ridondante.

Ci proponiamo ora di illustrare le operazioni fondamentali del metodo del simplesso su grafi. Innanzitutto è fondamentale ricordare che le basi della matrice di incidenza (privata di una qualsiasi riga) sono in corrispondenza biunivoca con gli alberi di supporto del grafo G (che abbiamo ipotizzato essere connesso) e che le basi nei problemi di flusso possono essere sempre scritte in forma triangolare (inferiore) con elementi non nulli sulla diagonale.

L'illustrazione del metodo del simplesso su reti sarà basata sull'esempio di rete di flusso in Figura 12.1.

12.2 Determinazione di una soluzione di base

Data una base B è sempre possibile risolvere in un unico modo il sistema $A_B x_B = b$ e determinare una soluzione di base. Nei problemi la cui matrice dei coefficienti è una matrice di incidenza di un grafo connesso, le basi della matrice dalla quale è stata

12. L'ALGORITMO DEL SIMPLEX SU RETI

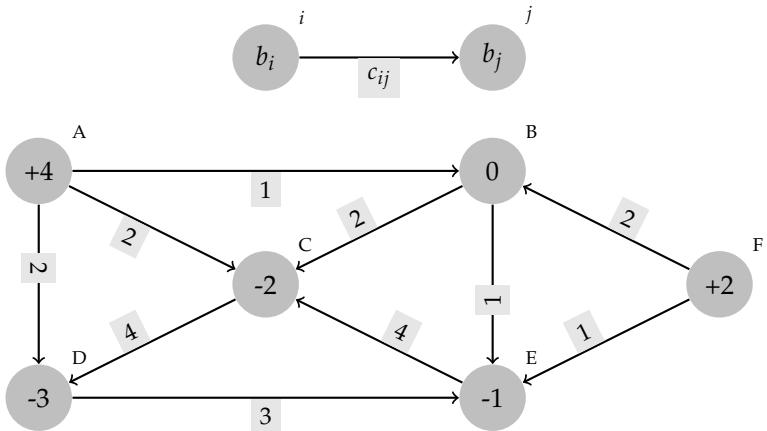
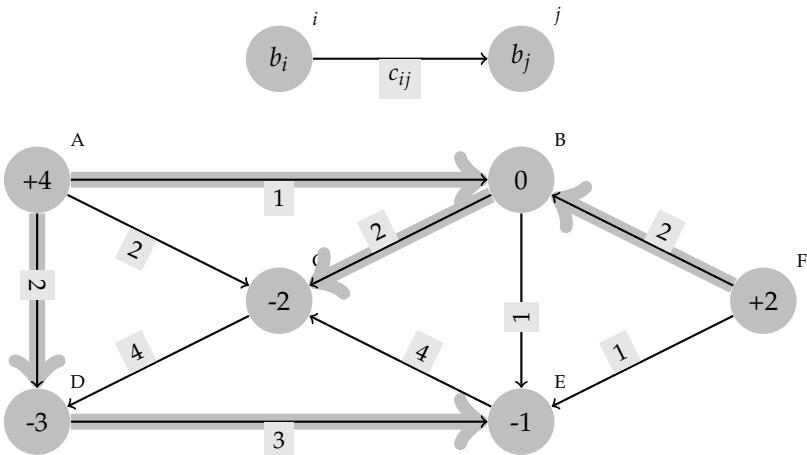


Figura 12.1: Problema di flusso su rete a costo minimo

eliminata una riga sono associate in maniera biunivoca agli alberi di supporto del grafo. Poiché tali basi, come si è visto, possono sempre essere portate in forma triangolare (ad es., inferiore), la soluzione del sistema sopra indicato diviene particolarmente facile e si riduce ad una procedura di sostituzione. Nell'esempio precedente, si consideri la base costituita dalle colonne associate agli archi $(A, B), (A, D), (B, C), (D, E), (F, B)$. riportata di seguito:

12.2. Determinazione di una soluzione di base



Scegliendo, ad esempio, come riga da eliminare dalla matrice di incidenza quella associata al nodo A , si può rappresentare la base come segue:

	FB	BC	AB	DE	AD
F	+1				
C		-1			
B	-1	+1	-1		
E				-1	
D				+1	-1

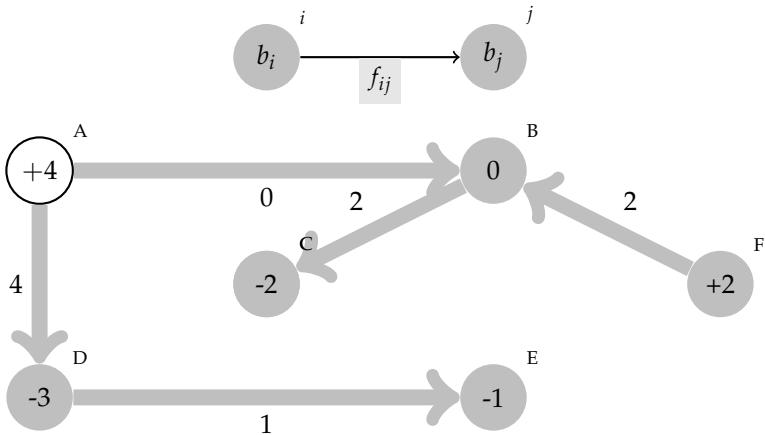
Ricordando che il vettore dei termini noti per il problema in esame nasce dal bilancio imposto sui nodi, pari a $+4$ per il nodo A , 0 per B , -2 per C , -3 per D , -1 per E e $+2$ per F , la soluzione di base si

12. L'ALGORITMO DEL SIMPLEX SU RETI

può trovare partendo dall'ultima riga della matrice:

$$\begin{aligned} f_{FB} &= 2 \\ -f_{BC} &= -2 \Rightarrow f_{BC} = 2 \\ f_{BC} - f_{FB} - f_{AB} &= 0 \Rightarrow f_{AB} = 0 \\ -f_{DE} &= -1 \Rightarrow f_{DE} = 1 \\ f_{DE} - f_{AD} &= -3 \Rightarrow f_{AD} = 4 \end{aligned}$$

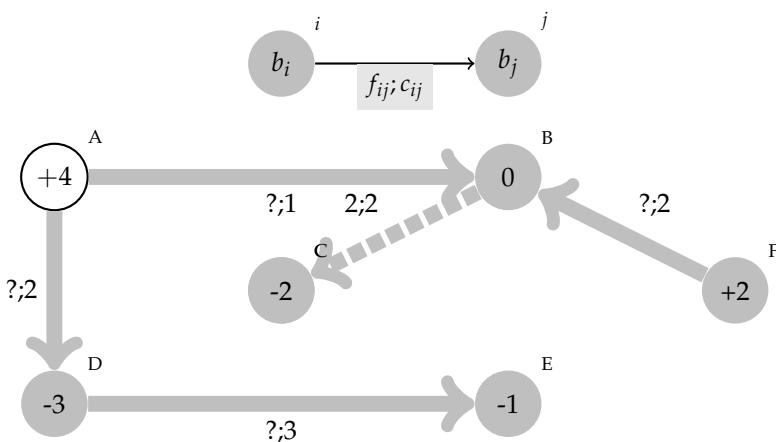
Considerando la rappresentazione grafica dell'albero associato alla base corrente, si può però notare che le operazioni appena viste hanno un'interpretazione naturale in termini di flussi. Infatti, riportando solo gli archi associati all'albero e lasciando colorato in bianco il nodo A associato alla riga eliminata dalla matrice di incidenza, si ottiene il seguente grafo:



Il nodo in bianco viene detto *radice* dell'albero di supporto. Per una nota proprietà, ciascun albero non banale contiene almeno due foglie. Quindi, anche nell'eventualità in cui il nodo radice fosse una foglia, ne rimarrebbe sicuramente almeno un'altra. La procedura

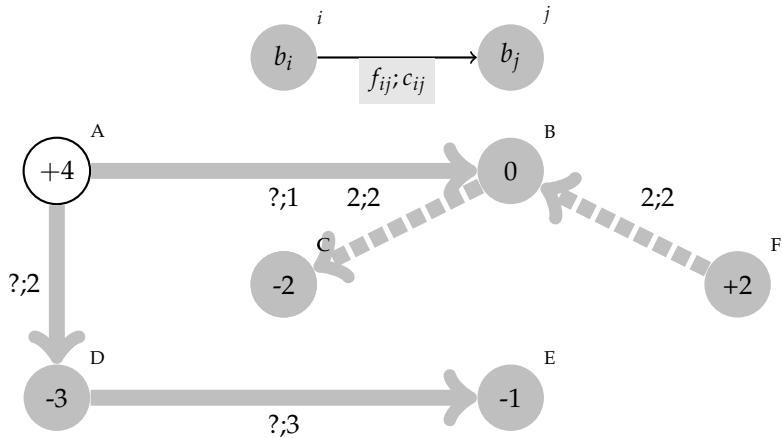
12.2. Determinazione di una soluzione di base

algebrica appena vista per la determinazione della soluzione di base altro non è che una visita dell'albero di supporto a partire dalle foglie verso la radice. Nell'esempio esistono tre foglie (diverse dalla radice): i nodi C , E ed F . Iniziando dal nodo C , per soddisfare il bilancio di tale nodo, l'unico arco incidente su C dovrà avere flusso pari a 2. A questo punto, immaginando di eliminare il nodo C e l'arco (B, C) dal grafo e modificando lo sbilancio del nodo B in modo da comprendere anche il flusso destinato a C si otterebbe il seguente grafo:

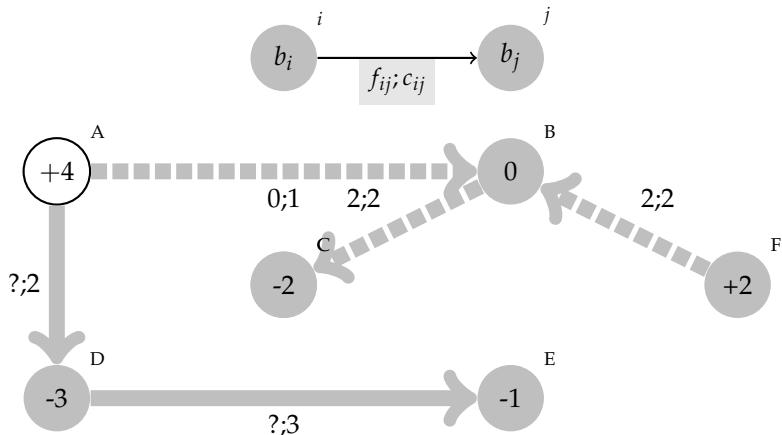


Proseguendo con la foglia F , si vede che l'unico flusso ammissibile in grado di soddisfare il bilancio di tale nodo è dato da $f_{FB} = 2$. Modificando di conseguenza il bilancio del nodo B si ottiene il grafo seguente:

12. L'ALGORITMO DEL SIMPLEX SU RETI

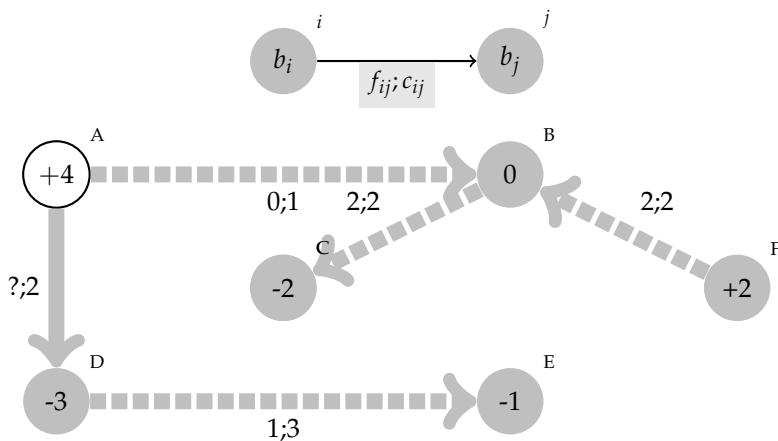


A questo punto il nodo B è diventato una foglia e si può procedere in modo analogo assegnando il flusso necessario sull'unico arco incidente (A, B) :



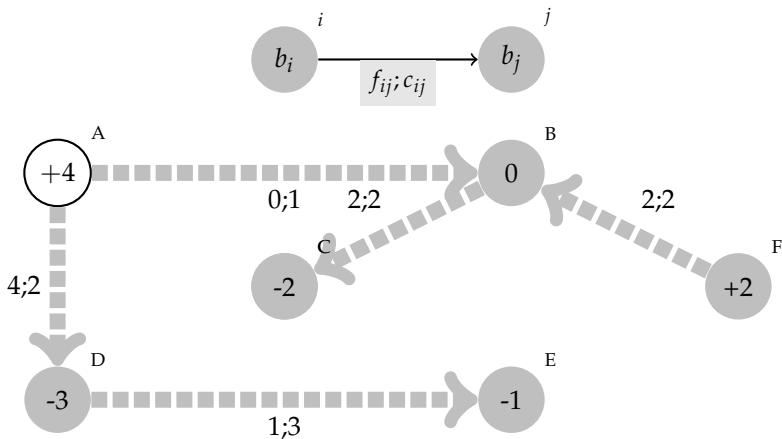
12.2. Determinazione di una soluzione di base

Immaginando di eliminare dal grafo anche l'arco (A, B) , si ottiene un albero con una sola foglia differente dalla radice e, pertanto, è obbligatorio procedere dal nodo E , il cui bilancio, -1 , può essere soddisfatto da un flusso di valore 1 sull'arco (D, E) :



Il nuovo bilancio sul nodo D richiede l'invio, lungo l'arco (A, D) , di 4 unità di flusso:

12. L'ALGORITMO DEL SIMPLEX SU RETI



A questo punto i flussi sono stati assegnati, nell'unico modo possibile, agli archi dell'albero corrispondente alla base corrente. La soluzione di base corrente fornisce una distribuzione ammissibile di flusso il cui costo è pari a 19 unità. Si osservi che i valori dei flussi sono ovviamente identici a quelli calcolati col metodo algebrico, e che il metodo grafico prevede l'esplorazione del grafo, a partire da ciascuna foglia, lungo i cammini che portano alla radice. Naturalmente, poiché la scelta della radice è arbitraria, allo stesso risultato si sarebbe arrivati utilizzando, come nodo radice, un altro nodo qualsiasi. Ad esempio, scegliendo la radice in C , il nodo iniziale sarebbe stato necessariamente o il nodo E oppure il nodo F ; scegliendo ad esempio il nodo E , si sarebbe attribuito il flusso $f_{DE} = 1$ con conseguente variazione del bilancio sul nodo D pari a -4. Successivamente si sarebbe assegnato flusso 4 sull'arco (A, D) , con modifica del bilancio di A da +4 a zero, e così via.

Volendo formalizzare la procedura per la valutazione della soluzione di base associata ad un albero si può pensare al seguente

12.3. Determinazione della soluzione complementare del duale

schema:

```
Dati un albero  $\langle V, T \rangle$ , un nodo radice  $r \in V$ , bilanci sui nodi  
b;  
repeat  
    Sia  $v$  una foglia,  $v \neq r$ ;  
    Sia  $e \in T$  l'unico arco incidente in  $v$ ;  
    Sia  $w$  l'estremo di  $e$  diverso da  $v$ ;  
    if  $e = (v, w)$  (arco uscente) then  
         $f_{vw} := b_v$ ;  
         $b_w := b_w + b_v$ ;  
    else  
         $e = (w, v)$  (arco entrante);  
         $f_{vw} := -b_v$ ;  
         $b_w = b_w - b_v$ ;  
    end  
     $V := V \setminus \{v\}$  ;  
     $T := T \setminus \{e\}$ ;  
until  $V = \{r\}$  ;
```

Algoritmo 2: Flusso di base

E' importante tenere presente il fatto che questa procedura determina l'unica soluzione di base associata all'albero individuato e non è detto che tale soluzione sia ammissibile: possono esistere archi ai quali la procedura assegna flusso negativo. Sarà compito dell'inizializzazione dell'algoritmo trovare un albero, se esiste, al quale è associata una soluzione di base *ammissibile*.

12.3 Determinazione della soluzione complementare del duale

Anche la soluzione del duale associata alla base corrente può essere determinata dalla risoluzione di un sistema di equazioni con matrice dei coefficienti triangolare: $\lambda_B^T A_B = c_B^T$. Come nel caso del-

12. L'ALGORITMO DEL SIMPLEX SU RETI

la determinazione dei flussi sugli archi, anche la valutazione dei potenziali da associare ai nodi può essere effettuata sia in modo algebrico, sia sfruttando le caratteristiche dell'albero associato alla base. Poiché, come noto, nei problemi di flusso un'equazione è ridondante e può essere eliminata, così, nel duale, una variabile è ridondante, e può essere scelta a piacere. Infatti, nei vincoli e nell'obiettivo del duale del problema di flusso a costo minimo, le variabili duali non compaiono mai singolarmente, ma sempre attraverso differenze. E' naturale, anche se non obbligatorio, associare al nodo radice dell'albero il valore duale pari a zero. A questo punto il sistema di equazioni da risolvere nell'esempio in esame diviene:

$$\lambda_{F,C,B,E,D} \begin{array}{|c|ccccc|} \hline & FB & BC & AB & DE & AD \\ \hline F & +1 & & & & \\ C & & -1 & & & \\ B & -1 & +1 & -1 & & \\ E & & & & -1 & \\ D & & & & +1 & -1 \\ \hline \end{array} = \begin{bmatrix} 2 \\ 3 \\ 1 \\ 2 \\ 2 \end{bmatrix}$$

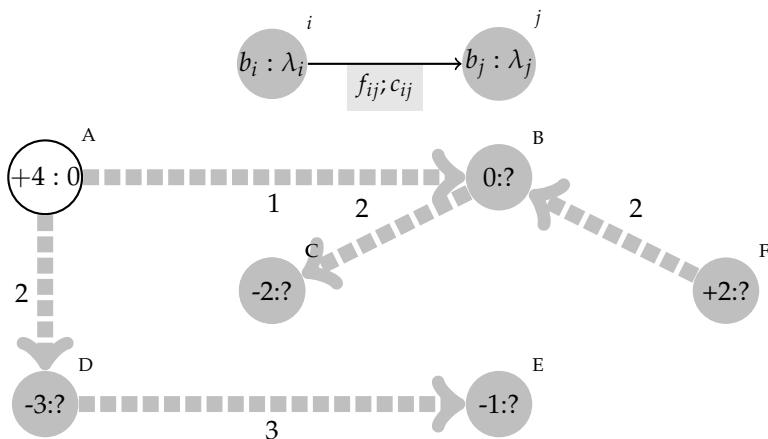
Le equazioni che si possono risolvere immediatamente sono quelle associate alle colonne della matrice di incidenza nelle quali compare un solo elemento non nullo, cioè le colonne più a destra della matrice. Tali colonne corrispondono ad archi incidenti sul nodo radice, al quale, come si è detto, è associata una riga che è stata eliminata dalla matrice di incidenza. In questo caso tali archi sono (A, D) e (A, B) . Essendo la matrice in forma triangolare inferiore, il sistema può essere risolto per sostituzione partendo dall'ultima colonna. Iniziando da (A, D) e proseguendo secondo l'ordine delle colonne della matrice di incidenza, si ottiene la seguente soluzione

12.3. Determinazione della soluzione complementare del duale

del sistema:

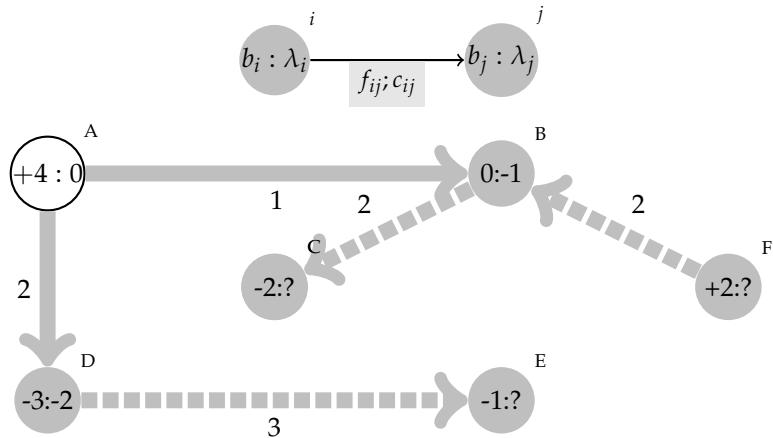
$$\begin{aligned}
 \lambda_D(-1) &= 2 & \Rightarrow \lambda_D &= -2 \\
 \lambda_D - \lambda_E &= 3 & \Rightarrow \lambda_E &= -5 \\
 \lambda_B(-1) &= 1 & \Rightarrow \lambda_B &= -1 \\
 \lambda_B - \lambda_C &= 2 & \Rightarrow \lambda_C &= -3 \\
 -\lambda_B + \lambda_F &= 2 & \Rightarrow \lambda_F &= 1
 \end{aligned}$$

Anche qui si può osservare come queste operazioni possano essere effettuate direttamente sul grafo. Infatti, indicando all'interno dei nodi il valore associato alla soluzione del duale, si ha la seguente situazione iniziale:

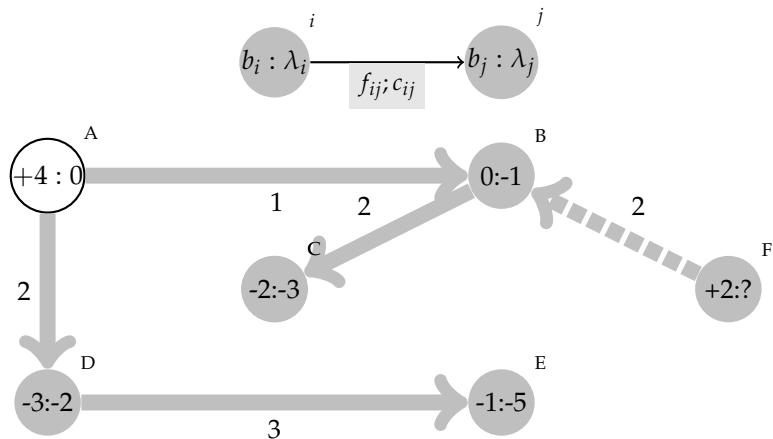


Imponendo la condizione $\lambda_i - \lambda_j = c_{ij}$ sugli archi uscenti dalla radice, si ottiene (in due iterazioni):

12. L'ALGORITMO DEL SIMPLEX SU RETI

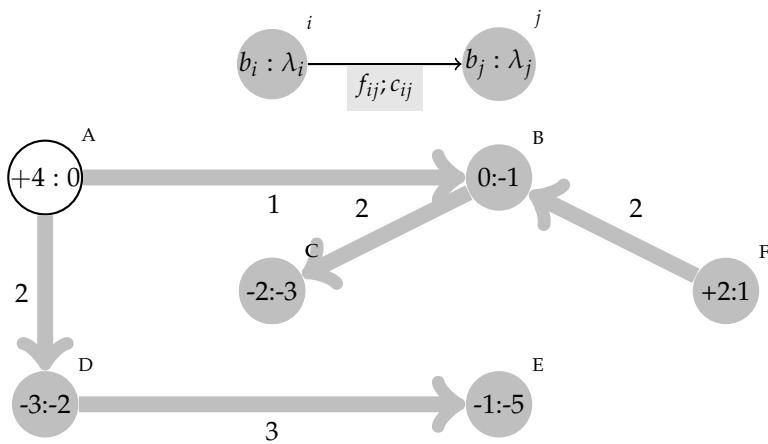


Proseguendo dai nodi ai quali è appena stato assegnato il valore duale ai loro immediati vicini si ottiene:



e, infine,

12.3. Determinazione della soluzione complementare del duale



Un possibile schema algoritmico per la determinazione della soluzione duale (non necessariamente D-ammissibile) associata ad un

12. L'ALGORITMO DEL SIMPLEX SU RETI

albero è il seguente:

```

Dati un albero  $\langle V, T \rangle$ , un nodo radice  $r \in V$ , costi sugli archi
 $c_{ij}$ ;
Sia  $R := \{r\}$ ;
Sia  $W := \{r\}$ ;
Si ponga  $\lambda_r$  pari ad un valore arbitrario (ad es. 0);
repeat
    Sia  $v \in R$  ;
    forall  $w \in V : w \notin W, (v, w) \in T$  do
         $\lambda_w := \lambda_v - c_{vw}$ ;
         $R := R \cup \{w\}$ ;
         $W := W \cup \{w\}$ ;
    end
    forall  $w \in V : (w, v) \in T$  do
         $\lambda_w := \lambda_v + c_{vw}$ ;
         $R := R \cup \{w\}$ ;
         $W := W \cup \{w\}$ ;
    end
     $R := R \setminus \{v\}$ ;
until  $R = \emptyset$  ;

```

Algoritmo 3: Soluzione duale associata ad un albero

Nell'algoritmo, l'insieme R contiene i nodi dai quali occorre propagare l'etichetta, mentre W contiene i nodi già etichettati.

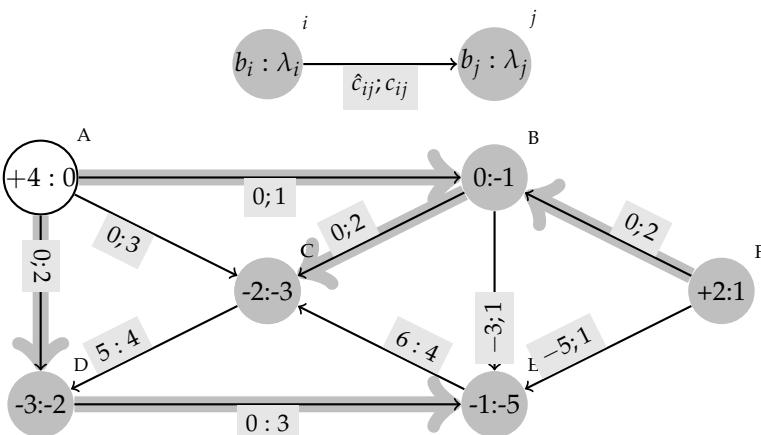
12.4 Calcolo dei coefficienti di costo ridotto

Il calcolo dei coefficienti di costo ridotto per gli archi non appartenenti alla base si riduce semplicemente alla valutazione di

$$\hat{c}_{ij} = c_{ij} - \lambda_B^T A_{ij} = c_{ij} - \lambda_i + \lambda_j$$

Sul grafo:

12.5. Determinazione dell'arco uscente di base



Poiché è presente più di un arco con coefficiente di costo ridotto negativo, si pone il problema della scelta della variabile destinata ad entrare in base. Come nel metodo (generico) del simplex, esistono numerose regole di scelta: si va dalla scelta del coefficiente di costo ridotto minimo (che richiede la valutazione di *tutti* i coefficienti di costo ridotto), alla scelta del primo coefficiente negativo, al minimo coefficiente in una lista limitata di archi.

12.5 Determinazione dell'arco uscente di base

Scelto un arco al quale è associato un coefficiente di costo ridotto negativo, il metodo del simplex prevede la determinazione di una variabile (arco) uscente di base, attraverso il calcolo dei coefficienti della variabile entrante in base nel dizionario corrente: indicata con d la quantità $-A_B^{-1}A_j$, occorre risolvere il sistema $A_B d = -A_j$. Questo sistema è perfettamente analogo a quello già visto per la determinazione del flusso associato alla base A_B . Per valutare i coefficienti d è sufficiente quindi considerare un problema di determinazione del flusso ammissibile lungo gli archi dell'albero di

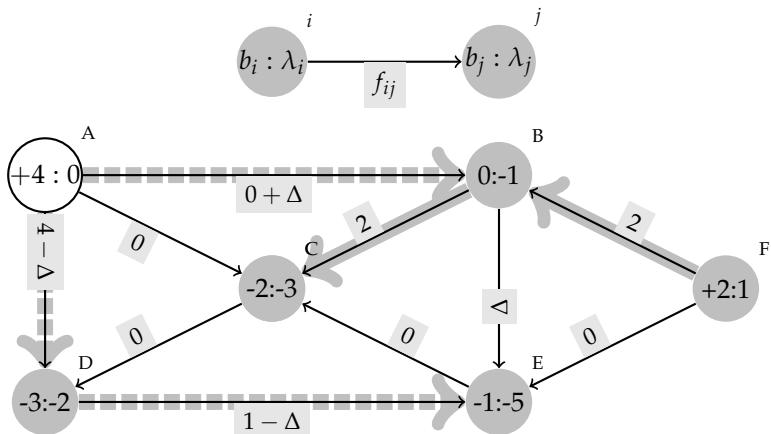
12. L'ALGORITMO DEL SIMPLEX SU RETI

supporto corrente, immaginando un bilancio pari a $+1$ nel nodo i e -1 nel nodo j , se l'arco (i, j) è scelto come arco entrante in base.

Si può anche vedere facilmente come determinare l'arco uscente di base per via grafica. Infatti l'aggiunta di un arco ad un albero, come noto dalle proprietà degli alberi, genera uno ed un solo ciclo. Nell'esempio, aggiungendo l'arco (B, E) , il ciclo è formato dalla seguente sequenza di nodi e archi:

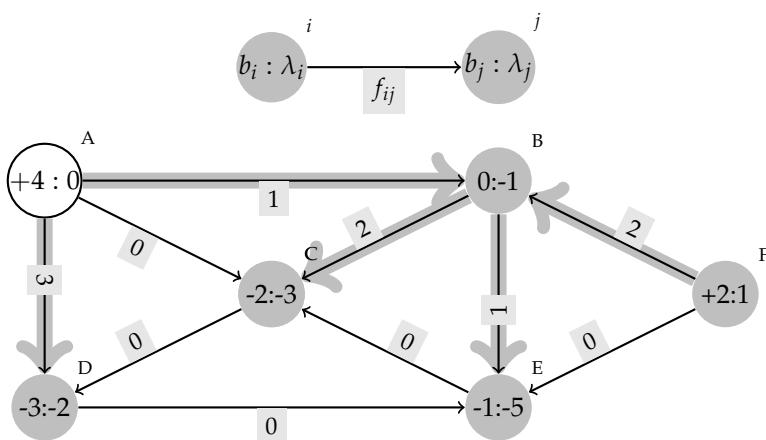
$$B, (B, E), E, (D, E), D, (A, D), A, (A, B), B$$

Nel metodo del simplex si procede aumentando il valore della variabile associata all'arco entrante in base, $f_{(B,E)}$. Immaginando di attribuire un flusso pari a $\Delta \geq 0$ a tale arco, occorre modificare di conseguenza i flussi su tutti gli archi del ciclo in modo tale da mantenere valida la legge di conservazione. In pratica, procedendo lungo la direzione indicata dall'arco entrante, gli archi del ciclo concordi con il verso di percorrenza subiranno un aumento pari a Δ , mentre quelli percorsi in senso inverso avranno una diminuzione di flusso di pari entità:



12.5. Determinazione dell'arco uscente di base

Si vede facilmente che il massimo incremento possibile è pari a $\Delta = 1$ e che, in corrispondenza di tale aumento, l'arco (D, E) vede azzerarsi il flusso. Tale arco viene quindi eliminato dalla base corrente ed il metodo riprende con la seguente situazione iniziale:



Come si vede, non è stato necessario ricalcolare i flussi associati all'albero della nuova base, ma è stato sufficiente aggiornare i flussi relativi al ciclo formatosi con l'aggiunta dell'arco entrante in base. La nuova soluzione di base corrisponde ad un flusso ammissibile di valore pari a 16 unità; questo valore è calcolabile direttamente dal grafo oppure può essere dedotto ricordando che la variabile entrante in base, a livello 1, aveva costo ridotto -3 e, pertanto, era attesa una riduzione di 3 unità della funzione obiettivo.

Riassumendo, la procedura per determinare una nuova soluzione di base a partire da una soluzione di base ammissibile è la

seguente:

```
Dati il grafo  $G = \langle V, E \rangle$ , un albero  $\langle V, T \rangle$ , un nodo radice  
 $r \in V$ ;  
Dati i costi sugli archi  $c_{ij}$  ed una soluzione di base  
ammissibile  $f$ ;  
forall  $(i, j) \in E \setminus T$  do  
| Sia  $\hat{c}_{ij} := c_{ij} - (\lambda_i - \lambda_j)$ ;  
end  
if  $\nexists (v, w) : \hat{c}_{vw} < 0$  then  
| Stop: la soluzione corrente è ottimale  
else  
| Sia  $(v, w) \in E \setminus T : \hat{c}_{vw} < 0$ ;  
| Sia  $C$  l'unico ciclo generato aggiungendo  $(v, w)$  a  $T$ ;  
| Siano  $C^+$  gli archi di  $C$  di verso concorde con  $(v, w)$ ;  
| Siano  $C^-$  gli archi di  $C$  di verso discorde da quello di  
 $(v, w)$ ;  
| if  $C^- = \emptyset$  then  
| | Stop: il problema è illimitato inferiormente;  
| end  
| Sia  $\Delta^* = \min\{f_{ij} : (i, j) \in C^-\}$ ;  
| Sia  $(k, \ell) \in T : f_{k\ell} = \Delta^*$ ;  
| forall  $(i, j) \in C^+$  do  
| |  $f_{ij} := f_{ij} + \Delta^*$ ;  
| end  
| forall  $(i, j) \in C^-$  do  
| |  $f_{ij} := f_{ij} - \Delta^*$ ;  
| end  
| Sia  $T := T \setminus \{(k, \ell)\} \cup \{(v, w)\}$ ;  
end
```

Algoritmo 4: Operazione di pivot

Come si vede dall'algoritmo e come è noto dalla teoria del metodo del simplex, se a nessun arco è associato un coefficiente di

12.5. Determinazione dell'arco uscente di base

costo ridotto negativo, allora la soluzione corrente è ottimale e l'algoritmo termina. Analogamente l'algoritmo termina nel caso in cui non esista una variabile candidata uscente di base: nel grafo questa situazione corrisponde ad aver determinato un ciclo il cui costo totale è negativo, ma lungo il quale il flusso può essere incrementato a piacere senza violare il vincolo di non negatività del flusso. In questo caso il problema è illimitato inferiormente. In tutti gli altri casi, all'aumentare del flusso sull'arco entrante in base esisterà almeno un arco il cui flusso diminuisce fino ad annullarsi: il primo arco (o uno tra i primi archi) il cui flusso si azzera, cioè l'arco con il minimo flusso in direzione discorde rispetto a quella dell'arco entrante in base, diviene l'arco destinato ad uscire. L'operazione di pivot corrisponde ad aggiungere un arco con coefficiente di costo ridotto negativo e ad eliminare l'arco (o uno degli archi) il cui flusso si è annullato. Per come è organizzata l'operazione, si vede immediatamente che la nuova soluzione sarà ancora una soluzione di base, cioè un albero. Infatti si giunge a tale soluzione aggiungendo un arco ad un albero con conseguente creazione di uno ed un solo ciclo ed eliminando un arco da tale ciclo: il grafo ottenuto è dunque ancora un albero.

A questo punto l'algoritmo del simplex può proseguire con un'altra iterazione; non è necessario calcolare il flusso ammissibile, essendo stato ottenuto nel corso della procedura di pivot. Occorre invece ricalcolare la soluzione del duale.

Naturalmente, è anche importante tenere sotto controllo le situazioni degeneri: può certamente accadere, nel corso di un'operazione di pivot, che $\Delta^* = 0$: in questo caso il flusso non varia, ma l'operazione di pivot ed il cambio di albero vengono effettuati ugualmente. Occorre adottare in questi casi procedure anti-ciclo, quali, ad esempio, la regola di Bland, per evitare cicli infiniti e la non terminazione dell'algoritmo. Anche se in queste note si desidera semplicemente illustrare il metodo del simplex su reti, senza volerci addentrare nei numerosissimi aspetti computazionali ed im-

plementativi, vale la pena ricordare come, per il caso delle soluzioni degeneri, esistano procedure che garantiscono la terminazione del metodo, anche differenti da quelle associate alla regola di Bland. Ad esempio, un albero T corrispondente ad una soluzione ammmissibile si dice *fortemente ammissibile* se esiste una radice r tale che ogni arco (i, j) con flusso nullo è *orientato in modo uscente* rispetto alla radice (con questo si intende dire che, identificato l'unico percorso che dal nodo r raggiunge i due nodi dell'arco, tale percorso incontra prima i e poi j). Si può dimostrare che se le soluzioni generate dal metodo del simplex sono tutte fortemente ammissibili, allora esse saranno tutte associate ad alberi distinti fra loro e, data la finitezza del numero di alberi del grafo, questo è sufficiente a garantire la terminazione finita dell'algoritmo. È possibile implementare l'operazione di pivot in modo tale da garantire che, partendo da una soluzione fortemente ammissibile, anche tutte le successive godano della stessa proprietà. La tecnica, piuttosto semplice, per ottenere questo risultato esula dal contenuto di questo volume.

12.6 Ricalcolo della soluzione del duale

Anche per quanto riguarda la soluzione duale complementare a quella rappresentata dal flusso corrente, non è necessario procedere al ricalcolo a partire dalla matrice associata all'albero, ma è sufficiente modificare solo alcuni potenziali. Infatti, l'eliminazione di un arco dalla base corrente genera la sconnessione dell'albero in due componenti (due alberi), una delle quali contenente il nodo radice. Poiché, come si è visto, il calcolo della soluzione del duale si effettua a partire dal nodo radice, per tutti i nodi della componente connessa contenente la radice il valore della soluzione duale non viene modificato. Per quanto riguarda l'altra componente, si può osservare che, con l'eccezione del nodo appartenente all'arco entrato in base, per tutti gli altri nodi la situazione rimane invariata. Cioè, per tali nodi dovrà rimanere invariata la differenza di

12.6. Ricalcolo della soluzione del duale

potenziale agli estremi degli archi che compongono l'albero di supporto corrente. Quindi l'unica variazione ammessa ai potenziali dei nodi di tale componente connessa sarà un aumento o una diminuzione costante, in modo da mantenere inalterata la differenza di potenziale.

Nell'esempio, l'eliminazione dell'arco (D, E) dall'albero genera due componenti connesse, una delle quali contenente i nodi A, B, C, D, F , l'altra ridotta al solo nodo E . Quindi la soluzione duale associata ai primi nodi resta invariata. Per l'attribuzione della soluzione duale alla seconda componente connessa, si distinguono due casi. Poiché il nuovo arco entrante in base è tale da ricreare un albero, cioè un grafo connesso (oltre che aciclico), sicuramente i suoi nodi estremi apparterranno alle due diverse componenti connesse. Sia (v, w) tale arco ed indichiamo con V_1 e V_2 l'insieme dei nodi della prima componente connessa (quella contenente la radice) e della seconda. Sono possibili due casi:

1. $v \in V_1, w \in V_2$. In questo caso, poiché si deve imporre $\lambda_v - \lambda_w = c_{vw}$, essendo λ_v già fissata, sarà sufficiente modificare i valori associati aggiungendo ai nodi di V_2 un valore pari a

$$\lambda_v - \lambda_w - c_{vw}$$

dove con λ si sono indicati i valori attuali della soluzione duale. La soluzione associata al nodo w infatti in questo caso diviene

$$\lambda_w + \lambda_v - \lambda_w - c_{vw} = \lambda_v - c_{vw}$$

come richiesto.

2. $v \in V_2, w \in V_1$: in questo caso per poter attribuire il valore corretto al nodo v occorrerà aumentare tutte le componenti della soluzione duale associate ai nodi di V_2 di una quantità pari a

$$\lambda_w - \lambda_v + c_{vw}$$

così da portare il potenziale associato a v al valore

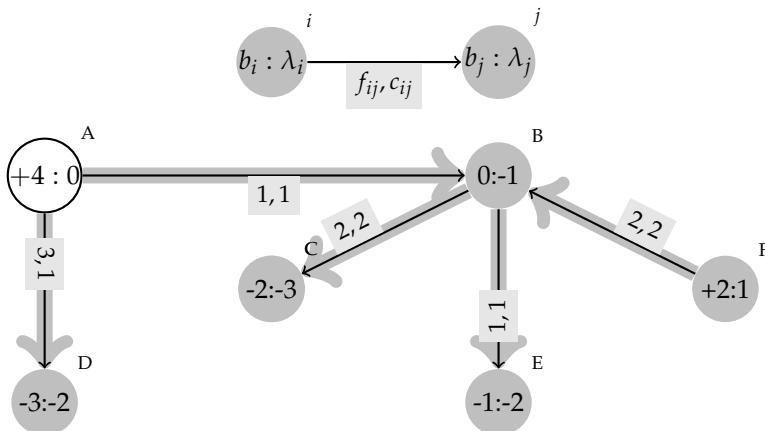
$$\lambda_v + \lambda_w - \lambda_v + c_{vw} = \lambda_w + c_{vw}$$

Un algoritmo per il ricalcolo della soluzione ammissibile del duale può quindi essere definito come segue:

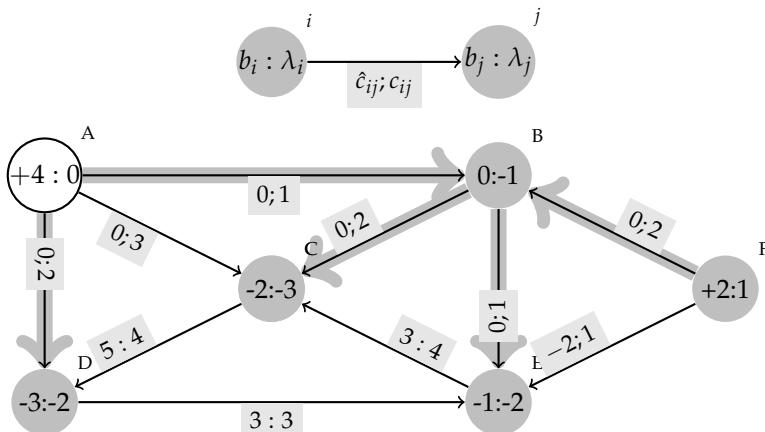
```
Dati il grafo  $G = \langle V, E \rangle$ , un albero  $\langle V, T \rangle$ , un nodo radice  
 $r \in V$ ;  
Dati i costi sugli archi  $c_{ij}$ , la soluzione corrente del duale  $\lambda$ ,  
un arco  $(v, w)$  entrante in base ed uno  $(k, \ell)$  uscente;  
Si elimini  $(k, \ell)$  da  $\langle V, T \rangle$  e siano  $V_1, V_2$  le componenti  
connesse risultanti, con  $r \in V_1$ ;  
if  $v \in V_1, w \in V_2$  then  
  | forall  $i \in V_2$  do  
  |   |  $\lambda_i := \lambda_i + (\lambda_v - \lambda_w - c_{vw})$   
  | end  
else  
  | forall  $i \in V_2$  do  
  |   |  $\lambda_i := \lambda_i + (\lambda_w - \lambda_v + c_{vw})$   
  | end  
end
```

Nell'esempio, l'arco entrante in base è (B, E) e, pertanto, vale il
primo dei due casi appena descritti. Si avrà pertanto $\lambda_E = -5 +$
 $(-1 + 5 - 1) = -2$. La situazione diviene quindi la seguente:

12.6. Ricalcolo della soluzione del duale



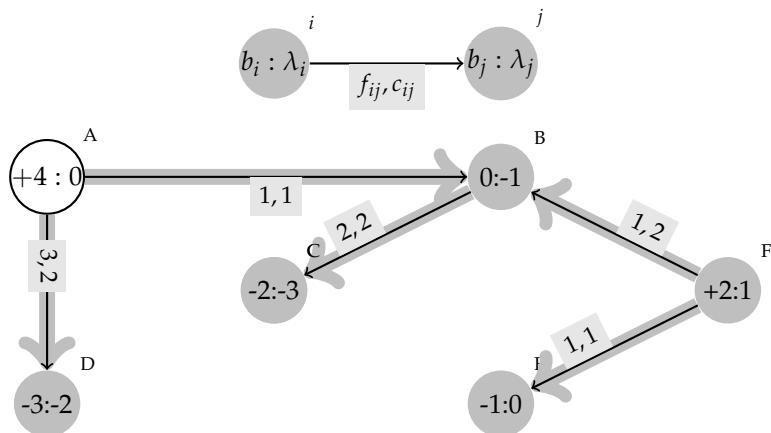
Proseguendo col metodo del simplex, si calcolano i nuovi coefficienti di costo ridotto:



Entrerà dunque in base l'arco (F, E) . Tale arco genera un ciclo $F, (F, E), E, (B, E), B, (F, B), F$. Incrementando di Δ il flusso, attualmente nullo, sull'arco (F, E) , si avrà una diminuzione di Δ unità

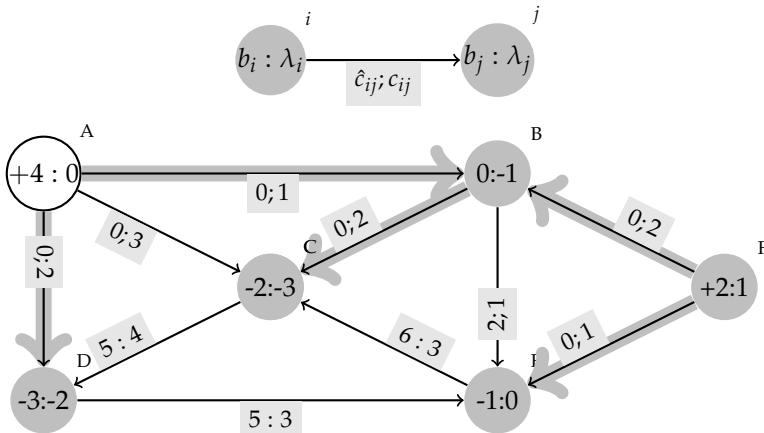
12. L'ALGORITMO DEL SIMPLEX SU RETI

sia sull'arco (B, E) che sull'arco (F, B) ; tale diminuzione dovrà limitarsi ad una unità a causa del vincolo di non negatività imposto sul flusso lungo l'arco (B, E) che, quindi, uscirà di base. Anche in questo caso, l'eliminazione dell'arco (B, E) divide l'albero associato alla base in due componenti connesse, di cui quella non contenente il nodo radice è ridotta al solo nodo E . Eseguendo i calcoli necessari si giunge alla seguente situazione:



La soluzione corrente ha costo pari a 14. I nuovi coefficienti di costo ridotto si ottengono facilmente come

12.7. Inizializzazione del metodo del simplex su reti



e, come si nota, sono tutti non negativi. La soluzione di base corrente è, pertanto, una soluzione ottimale del problema di flusso di costo minimo.

12.7 Inizializzazione del metodo del simplex su reti

Come per il metodo del simplex tradizionale, anche per il simplex su reti esistono diverse strategie per determinare, se esiste, una base ammissibile iniziale. In questa sezione si vedrà come interpretare il metodo delle due fasi su un problema di flusso su reti a costo minimo. Come si è visto, il metodo delle due fasi prevede di risolvere un problema artificiale, o problema prima fase, nel quale, dopo aver ricondotto il termine noto del problema ad essere non negativo, si aggiunge una variabile artificiale ad ogni vincolo e si risolve il problema corrispondente alla minimizzazione della somma delle variabili artificiali. In realtà, l'operazione di riportare il termine noto ad essere non negativo non è strettamente necessaria: si potrebbe equivalentemente lasciare inalterato il termine noto

ed introdurre una variabile artificiale con coefficiente +1 in tutti i vincoli con termine noto non negativo e coefficiente -1 nei vincoli il cui termine noto è invece negativo. Pensando ad un problema di flusso su reti, questa aggiunta di variabili artificiali genera nella matrice dei coefficienti un numero di colonne, pari al numero di nodi del grafo, in ciascuna delle quali solo un elemento è non nullo. Dato infatti un problema di flusso su reti a costo minimo:

$$\begin{aligned} \min & c^T f \\ \text{s.t. } & Af = b \\ & f \geq 0 \end{aligned}$$

il corrispondente problema prima fase può essere scritto come

$$\begin{aligned} \min & 1^T y \\ \text{s.t. } & Af + Dy = b \\ & f, y \geq 0 \end{aligned}$$

dove la matrice D è diagonale e gli elementi sulla sua diagonale principali sono dati da

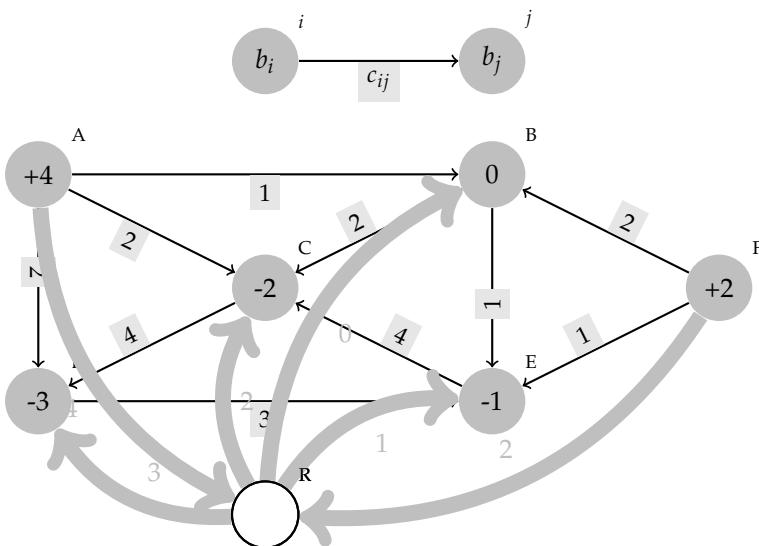
$$\begin{aligned} D_{ii} &= +1 && \text{se } b_i > 0 \\ D_{ii} &= -1 && \text{altrimenti} \end{aligned}$$

Si può pensare che le colonne dei coefficienti delle variabili artificiali y corrispondano ad archi (artificiali) connessi ad un nodo radice, anch'esso artificiale, e che la matrice dei coefficienti del problema prima fase corrisponda ad una matrice di incidenza di un grafo in cui la riga corrispondente al nodo artificiale radice sia stata eliminata. In particolare, gli archi artificiali andranno dal nodo radice ai nodi del grafo con bilancio negativo o nullo; dai nodi con bilancio positivo invece gli archi saranno orientati verso la radice. Si noti che così facendo l'albero costituito dai nodi adiacenti il nodo

12.7. Inizializzazione del metodo del simplex su reti

radice è associato ad una soluzione di base fortemente ammissibile, essendo gli archi a flusso nullo diretti in uscita dal nodo radice.

Nell'esempio utilizzato in questo capitolo, il grafo del problema prima fase sarà il seguente (dove gli archi artificiali, ai quali corrisponde costo pari ad uno, sono rappresentati come curvilinei):



La soluzione del duale associata a questa soluzione iniziale “tutta artificiale” si trova immediatamente associando valore +1 ai nodi con bilancio positivo, -1 a quelli con bilancio negativo, 0 ai nodi di transito. Poiché nel problema prima fase tutte le variabili artificiali hanno costo uno, mentre le variabili originali hanno costo 0, i coefficienti di costo ridotto di queste ultime variabili sono pari a

$$\hat{c}_{ij} = 0 - \lambda_i + \lambda_j$$

mentre i coefficienti di costo ridotto delle variabili artificiali possono anche non essere calcolati, poiché ciascuna variabile artificiale, una

12. L'ALGORITMO DEL SIMPLEXO SU RETI

volta uscita di base, può essere eliminata dal problema. Una volta definito il problema prima fase sul grafo, il metodo del simplex su rete porta, in un numero finito di iterazioni, ad una soluzione ottimale: se tale soluzione ottimale ha valore non nullo, il problema originale è privo di soluzioni ammissibili; in caso contrario, il metodo fornisce un albero ottimale ammissibile con il quale è possibile iniziare il metodo del simplex (su reti) applicato al problema originale.

Il problema del cammino di costo minimo

Il problema dell'individuazione di un *percorso di costo minimo* in un grafo orientato o meno, con un nodo *origine* o *sorgente* ed un nodo *destinazione* o *terminale*, può, in molti casi, essere facilmente formulato come un problema di flusso su reti di costo minimo. Si può infatti pensare che un percorso orientato da un'origine ad una destinazione possa essere individuato immaginando di immettere una unità di flusso dal nodo sorgente e prelevando tale unità dal nodo terminale. Se valgono alcune semplici ipotesi (in particolare se si assume che nel grafo non esistano circuiti chiusi di costo totale negativo), allora si può facilmente dimostrare che, se esistono soluzioni ottimali della formulazione del problema in termini di programmazione lineare ne esiste sempre almeno una binaria; gli archi corrispondenti alle componenti della soluzione pari ad uno formano un percorso di costo minimo sorgente/destinazione. Si potrebbe quindi pensare di utilizzare il metodo del simplex per risolvere tale problema. In effetti, il metodo del simplex su reti visto nel capitolo 12 può essere utilizzato per risolvere questo problema. La formulazione lineare del problema del cammino di costo

13. IL PROBLEMA DEL CAMMINO DI COSTO MINIMO

minimo è la seguente:

$$\begin{aligned} & \min c^T f \\ & Af = e_s - e_t \\ & f \geq 0 \end{aligned}$$

dove A è la matrice di incidenza di un grafo $G = \langle V, E \rangle$, $s \in V$ è la sorgente, $t \in V$ la destinazione, e_s, e_t sono vettori di dimensione $|V|$ le cui componenti sono tutte nulle con ad eccezione di quelle associate ad s ed a t rispettivamente, pari ad 1 e c è il vettore dei costi associati agli archi. Come si è visto in precedenza se questo problema ammette soluzione ottimale allora esiste una soluzione associata ad un albero di supporto del grafo (ammesso che il grafo sia连通的). Dal punto di vista della programmazione lineare si distinguono quindi i seguenti casi

1. il problema non è ammissibile; in questo caso non esiste la possibilità di inviare una unità di flusso dal nodo s al nodo t ; il grafo è quindi sconnesso e, in particolare, non esistono percorsi orientati che connettono sorgente e terminale.
2. il problema è illimitato inferiormente. In questo caso sicuramente esisterà un ciclo orientato nel grafo: in caso contrario, gli archi associati ai flussi non nulli formerebbero un grafo连通的 e aciclico e, cioè, un albero – su tale struttura sarebbe impossibile associare un flusso illimitato. Non solo, ma il ciclo individuato avrebbe sicuramente costo totale negativo.
3. il problema ammette soluzione ottimale. Se il grafo è连通的, allora il teorema 45 ed i teoremi collegati garantiscono che esiste una soluzione ottimale associata ad un albero di supporto. Una soluzione ammissibile, ed ottimale, associata ad un albero di supporto avrà flusso non nullo, pari ad 1, sull'unico percorso che collega s a t . Se invece il grafo non fosse

13.1. Algoritmo di Dijkstra

connesso, allora sicuramente s e t appartengono alla stessa componente connessa, altrimenti non si sarebbe potuta trovare una soluzione. Restringendo l'analisi a tale componente, nuovamente si può individuare un cammino ottimale a partire dalla soluzione del problema di PL .

13.1 Algoritmo di Dijkstra

In questo paragrafo verrà presentato un algoritmo specializzato per il problema del cammino di costo minimo in modo del tutto indipendente dall'utilizzo della programmazione lineare. In realtà l'algoritmo qui presentato non è altro se non un'implementazione efficiente del metodo del simplesso su reti.

L'*algoritmo di Dijkstra*(Dijkstra, 1959; Ahuja et al., 1993) procede attraverso un'*etichettatura* dei nodi del grafo (cioè ad una attribuzione di valori numerici ai nodi); tali etichette vengono variate nel corso dell'algoritmo, ma, da una certa iterazione in poi, divengono "*definitive*" e non subiscono ulteriori modifiche: l'algoritmo rientra nella categoria dei metodi cosiddetti "*label-setting*", nei quali cioè le etichette vengono gradualmente fissate a valori definitivi e procedono fino al momento in cui tutte le etichette non vengono fissate; esiste un'altra categoria di metodi, detti "*label-correcting*", nei quali le etichette subiscono modificazioni ad ogni iterazione, che terminano non appena nel corso di due iterazioni successive nessuna etichetta viene modificata.

L'algoritmo di Dijkstra può essere applicato esclusivamente a grafi con archi di costo non negativo. La sua applicazione a grafi con archi di costo negativo tipicamente porta alla determinazione di soluzioni sub-ottimali.

Lo schema dell'algoritmo di Dijkstra è riportato nello schema 5.

Nella fase di aggiornamento delle etichette viene adottata la convenzione di porre $c_{ij} = +\infty$ se $(i, j) \notin E$. In alternativa, si

13. IL PROBLEMA DEL CAMMINO DI COSTO MINIMO

```

/* Inizializzazione */;
 $R = \emptyset, \rho(s) = 0, \rho(v) = +\infty \quad \forall v \in V \setminus \{s\};$ 
while  $t \notin R$  do
    if  $\min \{\rho(v) : v \notin R\} = +\infty$  then
        return Il grafo è sconnesso;
    else
        /* etichetta minima */;
         $v_{\min} \in \arg \min \{\rho(v) : v \notin R\};$ 
        /* aggiornamento etichette definitive */;
         $R = R \cup \{v_{\min}\};$ 
        /* aggiornamento etichette temporanee */;
        for  $v \notin R$  do
             $\rho(v) = \min \{\rho(v), \rho(v_{\min}) + c_{v_{\min}, v}\}$ 
        end
    end
end

```

Algoritmo 5: Schema dell'algoritmo di Dijkstra

può pensare che l'aggiornamento delle etichette temporanee venga effettuato solo per i nodi v direttamente connessi a v_{\min} .

Se l'algoritmo venisse eseguito fino all'etichettatura definitiva di ogni nodo di V , cioè se il ciclo **while** fosse eseguito finché $R \neq V$, il risultato sarebbe un albero di percorsi di minimo costo dall'origine s ad ogni altro nodo del grafo. Naturalmente, se il grafo G non fosse connesso, ed, in particolare, non esistesse alcun cammino orientato da s ad un certo nodo v , l'etichetta di v resterebbe pari a $+\infty$.

L'algoritmo di Dijkstra, se eseguito su un grafo con costi non negativi, determina, in un numero finito di passi, una soluzione ottimale al problema del cammino di costo minimo.

Definizione 24. Dato un cammino v_1, v_2, \dots, v_k in un grafo G , si dicono *nodi intermedi* di tale cammino tutti i nodi con esclusione dell'ultimo, e cioè $\{v_1, v_2, \dots, v_{k-1}\}$. Se $k = 1$ i nodi intermedi sono

13.1. Algoritmo di Dijkstra

l'insieme vuoto.

Vale la proprietà seguente:

Teorema 49. (Correttezza dell'algoritmo di Dijkstra). *Dato un grafo $G = \langle V, E \rangle$ pesato con costi c non negativi, a ciascuna iterazione dell'algoritmo di Dijkstra, l'etichetta definitiva assegnata ad ciascun nodo $v \in R$ rappresenta il costo di un cammino minimo da s a tale nodo, mentre l'etichetta temporanea di ogni nodo $v \notin R$ rappresenta il minimo costo di un cammino da s a tale nodo in cui i nodi intermedi sono tutti in R .*

Nel teorema, se un nodo non è raggiungibile da alcun cammino, si sottintende che il costo del cammino minimo per “raggiungere” tale nodo è pari a $+\infty$.

Dimostrazione. La dimostrazione procede per induzione sul numero di iterazioni dell'algoritmo.

Al passo iniziale si ha:

$$\begin{aligned} R &= \emptyset \\ \rho(s) &= 0 \\ \rho(v) &= \infty \quad \forall v \neq s \end{aligned}$$

La prima tesi del teorema è verificata, banalmente per mancanza di nodi etichettati definitivamente. Per quanto riguarda la seconda tesi, essendo R vuoto, non esiste alcun cammino da s ad un generico nodo $v \neq s$ con nodi intermedi appartenenti ad R . Quindi il costo di tale cammino è convenzionalmente fissato a $+\infty$. Per quanto riguarda l'etichetta di s , pari a 0, questa è effettivamente il costo ottimale di un cammino da s ad s privo di nodi intermedi. Questa affermazione è vera, essendo il cammino degenere ridotto al solo nodo s l'unico cammino possibile.

Ipotizziamo ora entrambe le tesi valide ad una certa iterazione del metodo e procediamo a dimostrarne la validità anche all'iterazione successiva. Tra un'iterazione e la successiva, l'insieme dei

13. IL PROBLEMA DEL CAMMINO DI COSTO MINIMO

nodi definitivamente etichettati cambia per l'aggiunta del solo nodo v_{\min} . Si dimostra ora che la sua etichetta rappresenta il costo di un cammino ottimale.

Infatti, poiché, per ipotesi di induzione, $\rho(v_{\min})$ rappresenta il costo di un cammino particolare, per dimostrare che questo è il costo minimo sarà sufficiente mostrare che ogni altro cammino da s al nodo v_{\min} ha costo non inferiore a $\rho(v_{\min})$. I cammini uscenti da s e terminanti in v_{\min} possono essere di due tipi:

1. Cammini con tutti i nodi intermedi in R . Cammini di questo tipo, per ipotesi di induzione, non possono avere costo inferiore a $\rho(v_{\min})$.
2. Cammini con almeno un nodo intermedio esterno ad R . Ciascuno di questi cammini contiene almeno un nodo intermedio non etichettato definitivamente. Scelto comunque un cammino di questo tipo, sia u il primo nodo non appartenente ad R lungo tale cammino. Il costo di questo cammino sarà dato dalla somma del costo sottocammino da s ad u , indicato con C_1 e del costo del sottocammino da u a v_{\min} , indicato con C_2 . Sicuramente $C_1 \geq \rho(u)$ per ipotesi di induzione, in quanto l'etichetta di u rappresenta il costo del miglior cammino da s ad u con nodi intermedi tutti in R e C_1 è il costo di un cammino particolare da s ad u con tutti i nodi intermedi definitivamente etichettati. Il costo C_2 è non negativo, grazie all'ipotesi fatta sul costo associato agli archi. Pertanto

$$\begin{aligned}C_1 + C_2 &\geq \rho(u) + C_2 \\&\geq \rho(u) \\&\geq \rho(v_{\min})\end{aligned}$$

dove l'ultima minorazione deriva dall'ipotesi che v_{\min} sia un nodo con etichetta minima.

13.1. Algoritmo di Dijkstra

Si è quindi dimostrato che nessun cammino può avere un costo inferiore a $\rho(v_{\min})$ e, pertanto, tale etichetta rappresenta il costo del cammino ottimo per raggiungere il nodo v_{\min} .

Per dimostrare la seconda parte del teorema, si consideri la regola di aggiornamento delle etichette:

$$\rho(v) = \min\{\rho(v), \rho(v_{\min}) + c_{v_{\min}, v}\} \quad \forall v \notin R.$$

Per ipotesi di induzione, prima dell'introduzione del nodo v_{\min} in R , $\rho(v)$ rappresentava il costo di un cammino di costo minimo con nodi intermedi tutti contenuti in R . Dopo l'inserimento di v_{\min} in R la situazione potrebbe cambiare. Si consideri quindi il nuovo cammino di costo minimo da s a v con nodi intermedi tutti in R . Sia u l'ultimo nodo di R di tale cammino minimo; il costo del cammino sarà quindi dato da

$$\rho(u) + c_{u,v}$$

dato che le etichette definitive, come si è visto, rappresentano il costo di un cammino ottimale. Il costo minimo dei cammini vincolati ad avere tutti i nodi intermedi in R sarà pari a

$$\begin{aligned} \min_{u \in R} \rho(u) + c_{u,v} &= \min\left\{\min_{u \in R \setminus \{v_{\min}\}} \rho(u) + c_{u,v}, \rho(v_{\min}) + c_{v_{\min}, v}\right\} \\ &= \min\{\rho(v), \rho(v_{\min}) + c_{v_{\min}, v}\} \end{aligned}$$

che è esattamente la regola di aggiornamento delle etichette previste dall'algoritmo. \square

Lo schema dell'algoritmo di Dijkstra permette di conoscere il costo dei cammini minimi da s ad uno o a tutti i nodi di un grafo. Per ricostruire la sequenza di nodi che costituiscono tali cammini

13. IL PROBLEMA DEL CAMMINO DI COSTO MINIMO

è sufficiente arricchire la struttura dati e modificare l'algoritmo nel modo seguente:

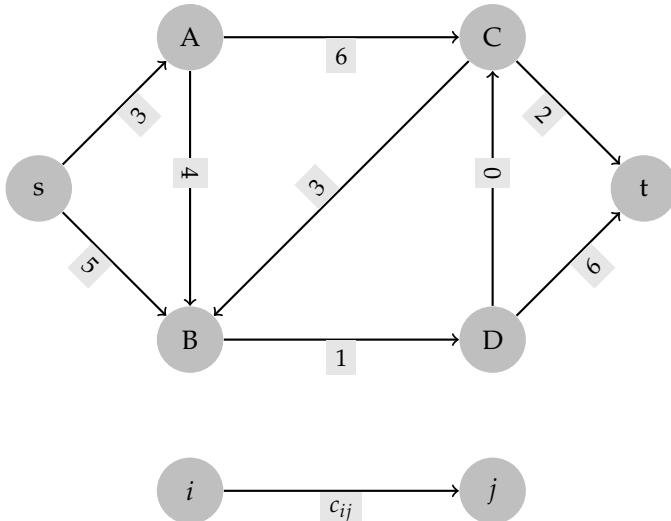
```
/* Inizializzazione */;
 $R = \emptyset, \rho(s) = 0, \rho(v) = +\infty \quad \forall v \in V \setminus \{s\};$ 
pred( $s$ ) =  $s$ ;
pred( $v$ ) =  $\emptyset \quad \forall v \neq s$ ;
while  $t \notin R$  do
    ...
    /* aggiornamento etichette temporanee */;
    for  $v \notin R$  do
        if  $\rho(v) > \rho(v_{\min}) + c_{v_{\min},v}$  then
            pred( $v$ ) =  $v_{\min}$ ;
             $\rho(v) = \rho(v_{\min}) + c_{v_{\min},v}$ 
        end
    end
end
```

Algoritmo 6: Memorizzazione del cammino ottimale nello schema dell'algoritmo di Dijkstra

La struttura $\text{pred}(\cdot)$ contiene, per ciascun nodo, il nodo del grafo che ha fornito il valore attuale dell'etichetta. Per ricostruire il cammino ottimo da s ad un nodo t basterà, al termine dell'algoritmo, leggere a ritroso la sequenza $t, \text{pred}(t), \text{pred}(\text{pred}(t)), \dots$

Esempio 36. In questo esempio si mostreranno le iterazioni compiute dall'algoritmo di Dijkstra su un semplice grafo orientato.

13.1. Algoritmo di Dijkstra

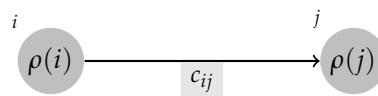
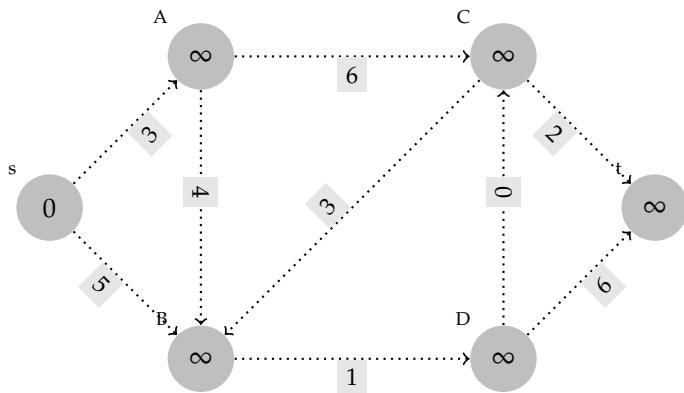


Iniziando l'algoritmo, si fissano le etichette al loro valore di default:

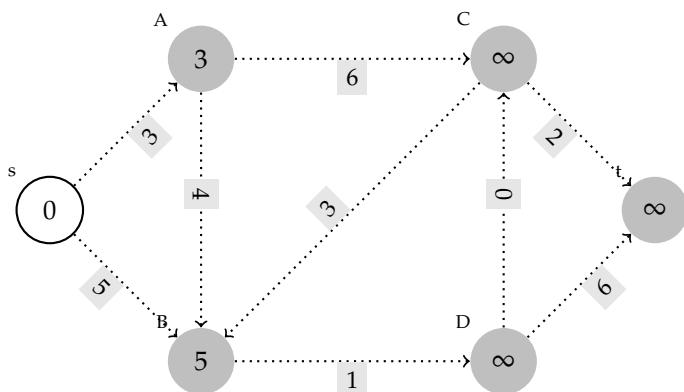
$$\begin{array}{ccccccccc} \rho(s) & \rho(A) & \rho(B) & \rho(C) & \rho(D) & \rho(t) & R & v_{\min} \\ 0 & \infty & \infty & \infty & \infty & \infty & \emptyset & s \end{array}$$

Sul grafo rappresentiamo i singoli passaggi dell'algoritmo indicando esplicitamente l'etichetta associata ai nodi ed evidenziando gli archi che connettono un nodo al suo predecessore.

13. IL PROBLEMA DEL CAMMINO DI COSTO MINIMO

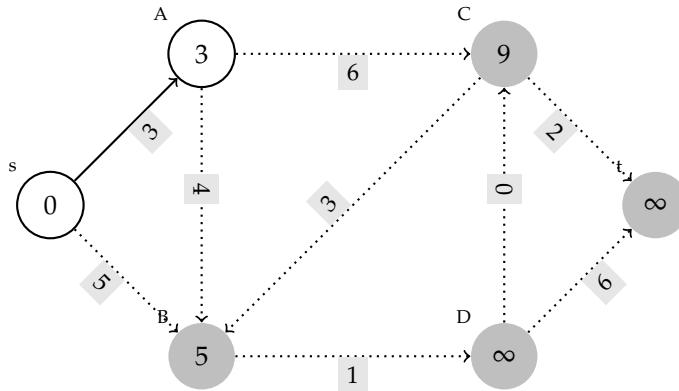


$$\begin{array}{ccccccccc} \rho(s) & \rho(A) & \rho(B) & \rho(C) & \rho(D) & \rho(t) & R & v_{\min} \\ 0 & 3 & 5 & \infty & \infty & \infty & s & A \end{array}$$

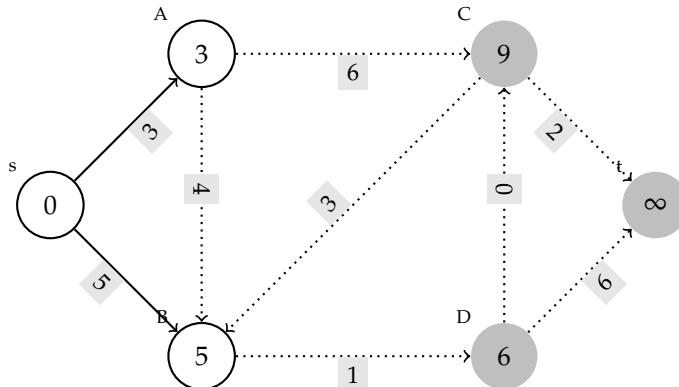


13.1. Algoritmo di Dijkstra

$\rho(s)$	$\rho(A)$	$\rho(B)$	$\rho(C)$	$\rho(D)$	$\rho(t)$	R	v_{\min}
0	3	5	9	∞	∞	s, A, B	

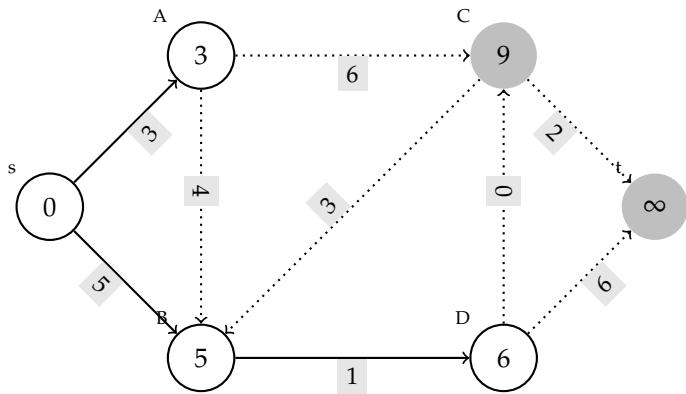


$\rho(s)$	$\rho(A)$	$\rho(B)$	$\rho(C)$	$\rho(D)$	$\rho(t)$	R	v_{\min}
0	3	5	9	6	∞	s, A, B	D

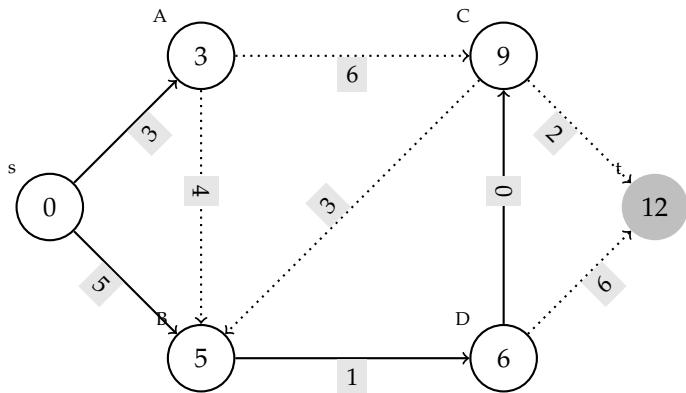


13. IL PROBLEMA DEL CAMMINO DI COSTO MINIMO

$$\begin{array}{ccccccccc} \rho(s) & \rho(A) & \rho(B) & \rho(C) & \rho(D) & \rho(t) & R & v_{\min} \\ 0 & 3 & 5 & 6 & 6 & 12 & s, A, B, D & C \end{array}$$

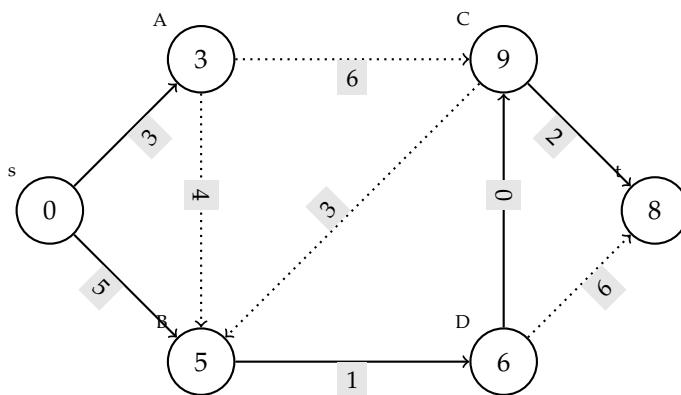


$$\begin{array}{ccccccccc} \rho(s) & \rho(A) & \rho(B) & \rho(C) & \rho(D) & \rho(t) & R & v_{\min} \\ 0 & 3 & 5 & 6 & 6 & 8 & s, A, B, C, D & t \end{array}$$



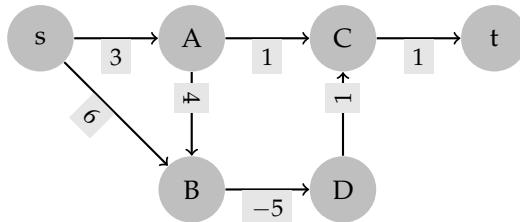
13.1. Algoritmo di Dijkstra

$$\begin{array}{ccccccccc} \rho(s) & \rho(A) & \rho(B) & \rho(C) & \rho(D) & \rho(t) & R & v_{\min} \\ 0 & 3 & 5 & 6 & 6 & 8 & s, A, B, C, D, t & \end{array}$$



Si può anche notare che l'ipotesi di non negatività dei costi associati agli archi è necessaria per il funzionamento dell'algoritmo di Dijkstra.

Esempio 37. Si consideri il grafo rappresentato in figura:



13. IL PROBLEMA DEL CAMMINO DI COSTO MINIMO

Se si esegue l'algoritmo di Dijkstra, si ottengono le seguenti etichette:

$\rho(s)$	$\rho(A)$	$\rho(B)$	$\rho(C)$	$\rho(D)$	$\rho(t)$	R	v_{\min}
0	∞	∞	∞	∞	∞	\emptyset	s
3	6	∞	∞	∞	∞	s	A
	6	4	∞	∞	∞	s, A	C
	6		∞	5	s, A, C		t
	6		∞		s, A, C, t		

e, come si vede facilmente dalla figura, il cammino trovato ha un costo superiore a quello del cammino ottimo (che, in questo caso, si trova facilmente: s, B, D, C, t , di costo totale pari a 3).

Una semplice modifica dell'algoritmo di Dijkstra permette di determinare il cammino di costo minimo anche in presenza di archi con costo negativo. Si tratta di modificare la struttura dell'algoritmo eliminando il concetto di etichetta permanente: tutte le etichette sono modificabili ad ogni iterazione e l'algoritmo termina quando tra due iterazioni successive non si assiste a nessun cambiamento. Esistono anche varianti del metodo di Dijkstra utili per determinare il cammino di costo minimo tra ogni coppia di nodi in un grafo. Tali varianti hanno un costo computazionale complessivo generalmente inferiore a quello ottenuto ripetendo l'esecuzione dell'algoritmo di Dijkstra da ciascun nodo del grafo.

Per quanto riguarda il tempo massimo di esecuzione dell'algoritmo, esso è facilmente calcolabile per la versione base del metodo, così come presentata in questo capitolo. Infatti si ha:

Teorema 50. *L'algoritmo di Dijkstra richiede al più $O(|V|^2)$ operazioni.*

Dimostrazione. In un grafo orientato con $|V|$ nodi, l'algoritmo, nel caso peggiore, richiede l'esecuzione di $|V|$ cicli, ciascuno dei quali comprendente la fase di identificazione del nodo con etichetta minima e la fase di aggiornamento delle etichette. Implementato senza

13.1. Algoritmo di Dijkstra

far uso di strutture dati particolari, l'algoritmo richiede, ad un generico passo, un numero di confronti pari al massimo a $|V \setminus R| - 1$ per determinare v_{\min} ed un numero di aggiornamenti di etichette temporanee pari al più a $|V \setminus R| - 1$; ciascuno di tali aggiornamenti richiede, nel caso peggiore, un confronto ed una somma. Quindi il numero totale di operazioni elementari, sempre nel caso peggiore, è limitato da

$$\sum_{k=1}^{|V|-1} k + 2 \sum_{k=1}^{|V|-1} k = 3 \frac{|V|(|V| - 1)}{2} = O(|V|^2).$$

□

Naturalmente, come sempre accade nella valutazione asintotica di complessità, si è tenuto qui conto del caso peggiore e sono stati trascurati termini che, coll'aumentare della dimensione del grafo, sono trascurabili rispetto a $|V|^2$; inoltre sono state trascurate le costanti moltiplicative. Tutte queste semplificazioni costituiscono una pratica standard nella valutazione di complessità computazionale di un algoritmo. Dalla proprietà sopra dimostrata, si può quindi affermare che il problema della determinazione del cammino ottimo in un grafo con archi di costo non negativo è un problema di complessità *polinomiale*. E' importante osservare però che l'utilizzo di strutture dati efficienti, così come l'uso di algoritmi di tipo differente, possono abbassare la complessità del problema o, a parità di complessità asintotica, possono portare ad implementazioni più efficienti abbassando il coefficiente moltiplicativo. Una struttura particolarmente utile per il problema in esame è il cosiddetto *heap*, una struttura dati che permette di determinare in modo efficiente il minimo tra un insieme di dati numerici, rappresentando i dati stessi mediante una struttura parzialmente ordinata (Fredman & Tarjan, 1987). Si può facilmente intuire che la complessità dell'algoritmo potrà diminuire soltanto nel caso di grafi sparsi, di grafi cioè nei quali siano presenti archi in numero molto inferiore rispetto al ca-

13. IL PROBLEMA DEL CAMMINO DI COSTO MINIMO

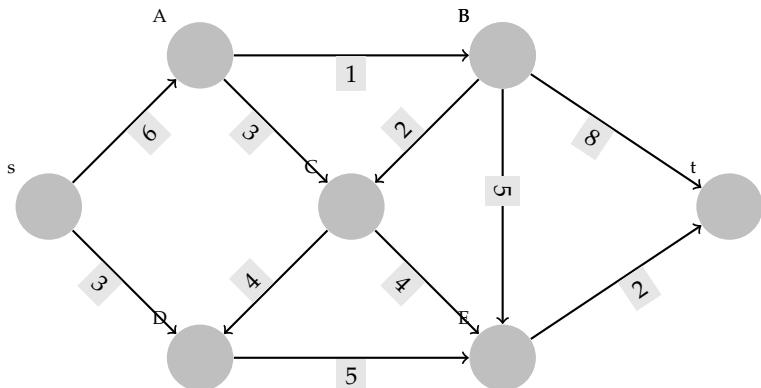
so di un grafo completo. Per i grafi completi infatti la complessità $O(|V|^2)$ è inevitabile, in quanto, essendo presenti $|V|(|V| - 1)$ archi sarà sempre possibile costruire esempi nei quali l'algoritmo di Dijkstra sia obbligato a valutare il costo di ciascuno di essi. Tuttavia, nel caso di grafi sparsi, le operazioni di aggiornamento possono essere svolte in modo più efficiente.

13.2 Applicazione del simplex su reti al problema del percorso di costo minimo

Come caso particolare di applicazione del metodo del simplex su reti si consideri il caso del percorso di costo minimo su un grafo pesato con archi di costo non negativo:

$$\begin{aligned} \min & c^T f \\ Af &= e_s - e_t \\ f &\geq 0 \end{aligned}$$

Si consideri, a titolo di esempio, il grafo seguente:



13.2. Applicazione del simplex su reti al problema del percorso di costo minimo

Il metodo della “grande M ”

E’ noto che un’alternativa al metodo delle due fasi per l’inizializzazione del metodo del simplex è nota in letteratura con il nome di metodo della grande M (si veda il paragrafo 6.3 a pag. 128). Il metodo prevede la ridefinizione del problema di ottimizzazione come

$$\begin{aligned} \min c^T f + M\mathbf{1}^T y \\ Af + Dy = b \\ f, y \geq 0 \end{aligned}$$

dove M è una costante positiva “di valore arbitrariamente grande”. In pratica si può evitare di assegnare un valore numerico ad M , ma considerarla in modo simbolico, immaginando, tutte le volte in cui è necessario effettuare un confronto, che il suo valore sia sempre maggiore di qualunque altro numero positivo. L’idea del metodo è quella di fondere in un’unica fase le due fasi del metodo precedentemente esposto, in modo da tenere conto, nel cercare soluzioni ammissibili, anche del costo del problema originale. Non appena una variabile (arco) artificiale y esce di base, essa può essere eliminata definitivamente dal problema; se al termine dell’esecuzione del metodo del simplex si ottiene una soluzione nella quale alcune variabili artificiali compaiono a livello positivo, si può concludere che il problema originale era privo di soluzioni ammissibili. In caso contrario, si otterrà la soluzione ottimale del problema dato (oppure si dimostrerà che tale problema è illimitato inferiormente). Come nel metodo delle due fasi, l’introduzione delle variabili artificiali y permette la costruzione immediata di una soluzione (albero) ammissibile di base iniziale con cui far partire il metodo del simplex.

Per iniziare le iterazioni del metodo della grande M occorrerebbe introdurre un nodo artificiale r ; occorrerebbe poi introdurre gli archi (r, v) per ciascun nodo $v \neq s$ e l’arco (s, r) . A tutti questi archi

13. IL PROBLEMA DEL CAMMINO DI COSTO MINIMO

viene associato il costo artificiale M . In realtà si vede abbastanza facilmente che, in questo caso, si può evitare l'introduzione di un nodo r artificiale e definire un problema in cui il nodo s è collegato da archi artificiali a costo M verso ogni altro nodo del grafo (si lascia per esercizio la verifica della correttezza di questa impostazione).

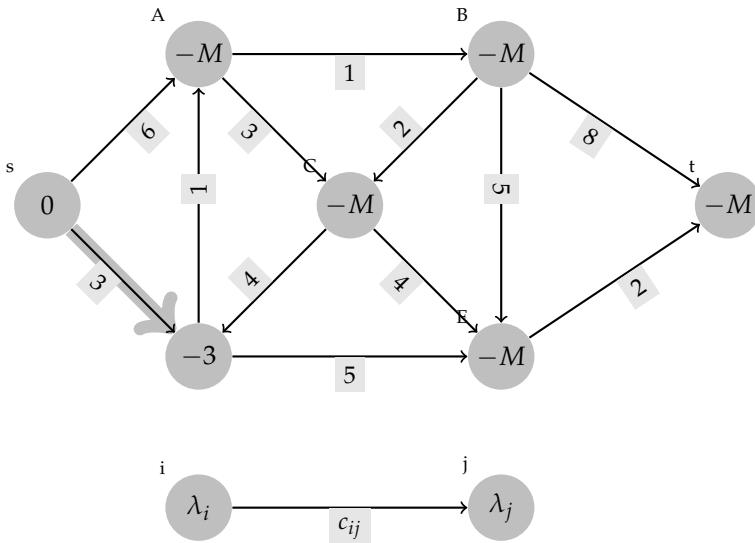
La soluzione di base iniziale, quindi, corrisponde all'albero con radice s ed archi artificiali verso tutti i nodi. La soluzione di base associata avrà flusso 1 sull'arco artificiale (s, t) e zero su tutti gli altri. La soluzione del duale, invece, fissato a 0 il potenziale del nodo s , sarà pari a $-M$ su ogni nodo.

I coefficienti di costo ridotto (solo per gli archi “normali”) risultano pari a

$$\begin{aligned}\hat{c}_{sj} &= c_{sj} - 0 - M & j : (s, j) \in E \\ \hat{c}_{ij} &= c_{ij} - M + M & \forall (i, j) : i \neq s\end{aligned}$$

e, a causa del fatto che M è positivo e “grande”, risulteranno negativi tutti e soli i coefficienti associati ad archi uscenti dal nodo sorgente e in particolare si avrà $\hat{c}_{sA} = 6 - M$, $\hat{c}_{sD} = 3 - M$. Volendo scegliere la regola del coefficiente di costo ridotto minimo, l'arco entrante in base sarà quello uscente dalla sorgente con costo minimo (nel caso dell'esempio: l'arco (s, D)). Aggiungendo tale arco si genera un'unico ciclo, composto dagli archi (s, D) (artificiale) ed (s, D) (naturale). L'operazione di pivot porta all'uscita (definitiva) dell'arco artificiale; l'operazione di pivot in questo caso è degenere (sull'arco non passava alcun flusso) e pertanto la soluzione primale non varia; la soluzione duale invece viene modificata come si vede nel grafo: dove si è evidenziato l'unico arco “normale” attualmente presente in base.

13.2. Applicazione del simplex su reti al problema del percorso di costo minimo



Al passo successivo, i coefficienti di costo ridotto degli archi normali fuori base risultano:

$$\hat{c}_{sA} = 6 - M$$

$$\hat{c}_{CD} = 4 + M - 3$$

$$\hat{c}_{DA} = 1 - M + 3$$

$$\hat{c}_{DE} = 5 - M + 3$$

mentre per tutti gli altri archi essi rimangono invariati. Gli unici archi con coefficiente di costo ridotto negativo risultano essere quindi $(s, A), (D, A), (D, E)$ e, scegliendo il minimo, il prossimo arco entrante in base sarà (D, A) . Di nuovo il flusso non varia; esce definitivamente di base l'arco artificiale (s, A) e la soluzione del duale associata al nodo A diviene pari a -4 . I nuovi coefficienti di costo

13. IL PROBLEMA DEL CAMMINO DI COSTO MINIMO

ridotto sono

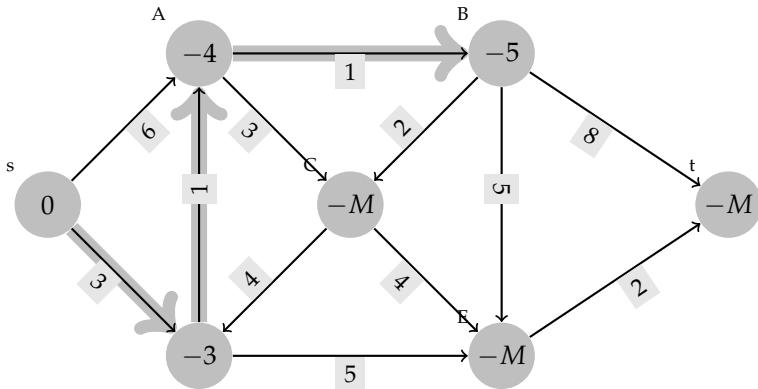
$$\hat{c}_{sA} = 2$$

$$\hat{c}_{AB} = 1 + 4 - M$$

$$\hat{c}_{CD} = 4 + M - 3$$

$$\hat{c}_{DE} = 5 - M + 3$$

Il coefficiente minimo risulta quello associato ad (A, B) che entrerà in base al posto dell'arco artificiale (s, B) . Ricalcolando la soluzione del duale, la situazione risulta essere la seguente:



I coefficienti di costo ridotto sono dati da:

$$\hat{c}_{BC} = 2 + 5 - M$$

$$\hat{c}_{BE} = 5 + 5 - M$$

$$\hat{c}_{Bt} = 8 + 5 - M$$

$$\hat{c}_{AC} = 7 - M$$

$$\hat{c}_{CD} = 7 + M$$

$$\hat{c}_{DE} = 8 - M$$

...

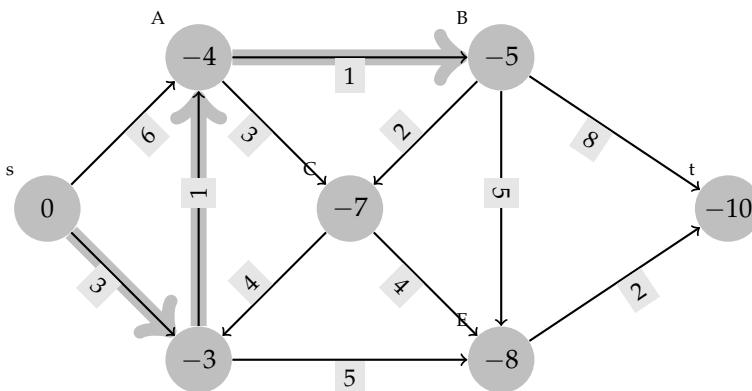
13.2. Applicazione del simplex su reti al problema del percorso di costo minimo

e quindi il coefficiente negativo minimo è pari a $7 - M$, raggiunto sia in corrispondenza di (B, C) che di (A, C) : la scelta è indifferente e porta, in ogni caso, all'uscita dell'arco artificiale (s, C) dalla base. Scegliendo, ad esempio (B, C) , si ottiene una nuova soluzione di base in cui la componente duale associata a C vale -7 . Ricalcolando i coefficienti di costo ridotto, si ottiene $\hat{c}_{CE} = 4 + 7 - M$, mentre $\hat{c}_{AC} = 0$ (come facilmente prevedibile). Il coefficiente minimo risulta quindi associato a (D, E) , il cui ingresso in base porta a -8 il valore duale associato ad E . A questo punto i coefficienti di costo ridotto negativi risultano essere

$$\hat{c}_{Bt} = 13 - M$$

$$\hat{c}_{Et} = 2 + 8 - M$$

e pertanto entra in base (E, t) . Questa non è più un'operazione di pivot degenere, poiché l'arco uscente di base è l'arco artificiale (s, t) che ha flusso unitario. Aggiornando quindi flussi e soluzione del duale si ottiene la seguente configurazione:



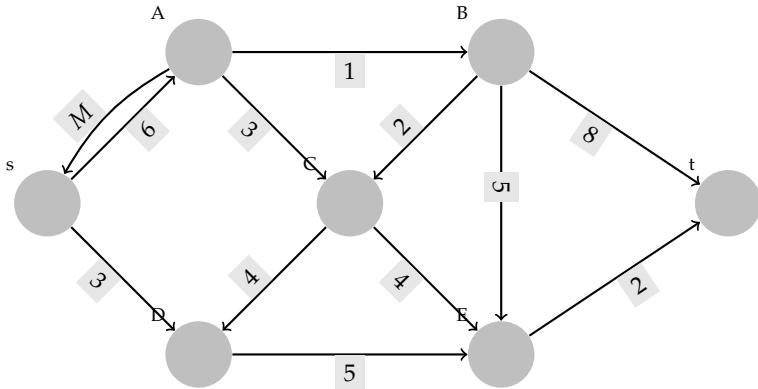
Si vede facilmente che questa soluzione ammissibile è anche ammissibile per il duale e, pertanto, rappresenta un ottimo del problema di cammino di costo minimo.

13. IL PROBLEMA DEL CAMMINO DI COSTO MINIMO

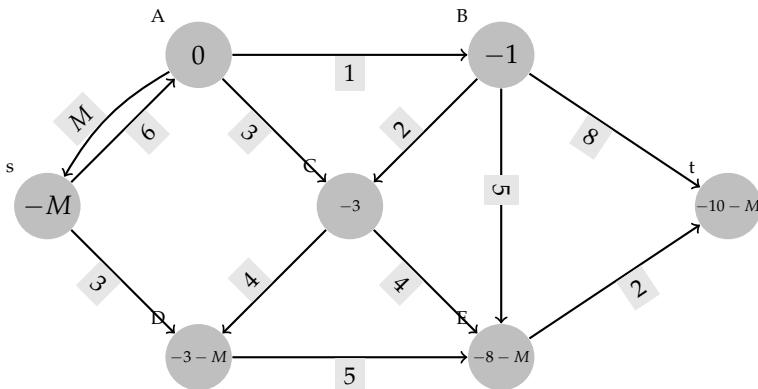
Si può osservare che l'algoritmo del simplesso per il problema in esame segue passi del tutto analoghi a quelli dell'algoritmo di Dijkstra, nel caso di grafi pesati con costi non negativi. In particolare, la soluzione del duale rappresenta l'etichetta dei nodi. La decisione relativa all'arco entrante in base con la regola del coefficiente di costo ridotto minimo è la stessa che prenderebbe il metodo di Dijkstra; ad una qualunque iterazione, se si analizzano, per ogni nodo del grafo, i valori associati ai coefficienti di costo ridotto, questi, se negativi, hanno sempre la forma $C - M$, dove C è il costo di un particolare cammino che da s porta al nodo, non necessariamente ottimale. La minima di queste etichette è facilmente riconoscibile come associata all'etichetta temporanea associata dall'algoritmo di Dijkstra. Si può anzi vedere come l'algoritmo di Dijkstra non sia altro che una riorganizzazione delle operazioni di ricalcolo dei coefficienti di costo ridotto fatta in modo tale da non dover ricalcolare inutilmente tutti i coefficienti ad ogni iterazione.

Rispetto all'algoritmo di Dijkstra, il metodo del simplesso ha alcuni importanti vantaggi. Innanzitutto può essere utilizzato anche nel caso di presenza di archi di costo negativo, praticamente senza alcuna modifica. Inoltre, una volta terminato, l'algoritmo permette con relativo poco sforzo computazionale di calcolare i percorsi ottimali a partire da un nodo differente rispetto ad s . Infatti, se volessimo, ad esempio, calcolare i percorsi ottimali avendo come nodo origine, ad esempio, il nodo A , basterebbe eliminare dall'albero di supporto l'unico arco entrante in A e sostituirlo con un arco artificiale (A, s) , di costo M :

13.2. Applicazione del simplex su reti al problema del percorso di costo minimo



Così facendo si genera un albero di supporto con radice A , dal quale si può iniziare con il metodo del simplex; nel caso in esame, aggiornando la soluzione del duale, si ottiene



Da questa soluzione può iniziare il metodo del simplex.

La dimostrazione formale dell'equivalenza tra questo algoritmo ed il metodo di Dijkstra viene lasciata come esercizio.

Il problema del massimo flusso

14.1 Algoritmo di Ford e Fulkerson

La determinazione della massima quantità di flusso che può circolare su una rete è un problema di notevole interesse applicativo, in particolare nel campo della progettazione di reti di flusso; problemi di massimo flusso si incontrano sia nella progettazione di infrastrutture di trasporto (di liquidi, gas, pacchetti di bit, veicoli), sia in vari altri campi. Anche se a prima vista può essere difficile riconoscerlo, anche il problema della schedulazione di lavori su macchine parallele può, in determinate situazioni, essere rappresentato come un problema di massimo flusso su una rete. In questo capitolo verranno introdotti alcuni concetti fondamentali per la descrizione delle reti di flusso e si arriverà alla definizione di un algoritmo classico per la determinazione del massimo flusso. Esistono algoritmi più recenti di quello di *Ford e Fulkerson*, che risale a (Ford & Fulkerson, 1956), e, in generale più efficienti (si veda, ad esempio, (Goldberg, Tardos, & Tarjan, 1990)); tuttavia pare opportuno qui presentare l'algoritmo originale sia per la sua relativa semplicità ed efficienza, sia perché nel presentare tale algoritmo verranno illustrate alcune proprietà fondamentali delle reti di flusso.

14. IL PROBLEMA DEL MASSIMO FLUSSO

Si consideri una rete orientata

$$N = \langle V, E, s, t, b \rangle$$

dove $\langle V, E \rangle$ è un digrafo, s e t sono rispettivamente un nodo sorgente ed un nodo terminale, b è la capacità massima di ciascun arco. Si ipotizza qui che nella rete siano presenti un'unica sorgente ed un unico nodo terminale, anche se sarebbe semplice estendere la trattazione anche al caso di sorgenti o terminali multipli, ad esempio trasformandoli in nodi di transito collegati ad una sorgente ed un terminale fintizi.

Data una rete N il problema del *massimo flusso* consiste nel determinare, se esiste, un flusso ammissibile f che abbia *valore* massimo. Per flusso ammissibile si intende un flusso che soddisfi alla legge di conservazione del flusso per tutti i nodi, eccetto s e t con bilancio nullo, che abbia flusso uscente da s uguale al flusso entrante in t e che rispetti il vincolo di capacità massima.

Definizione 25. Il *valore* val di un flusso f ammissibile è definito come

$$val = \sum_{v:(s,v) \in E} f_{s,v}$$

Quindi il valore del flusso è definito come il flusso totale erogato dalla sorgente. Questa definizione è valida nel caso in cui dal nodo sorgente esistano solo archi, con flusso non nullo, uscenti. Più in generale, anche se in questa parte non sarà necessario, si potrebbe definire valore del flusso il totale *netto* di flusso uscente dal nodo sorgente, cioè

$$val = \sum_{v:(s,v) \in E} f_{s,v} - \sum_{w:(w,s) \in E} f_{w,s}$$

Il problema della determinazione di un flusso di valore massimo può quindi essere rappresentato come il seguente problema di

14.1. Algoritmo di Ford e Fulkerson

programmazione lineare:

$$\begin{aligned} & \max_{f, val} val \\ & Af = val(e_s - e_t) \\ & f \leq cap \\ & f \geq 0 \end{aligned}$$

dove con A si è indicata la matrice di incidenza del grafo.

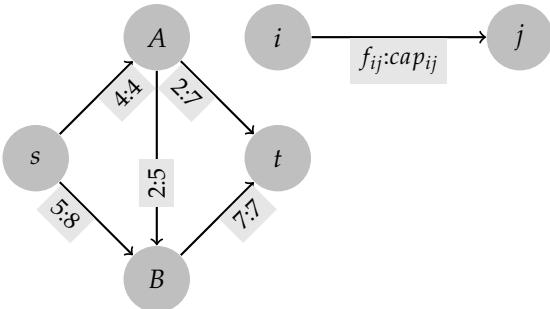
Il problema sopra indicato è un caso particolare dei problemi di flusso su rete a costo minimo. Si vede facilmente che, introducendo un arco fittizio (t, s) con capacità massima illimitata e costo -1 , attribuendo costo nullo a tutti gli altri archi, il problema è equivalente a

$$\begin{aligned} & \min [0 \quad 1]^T \begin{bmatrix} f \\ val \end{bmatrix} \\ & [A \quad -e_s + e_t] \begin{bmatrix} f \\ val \end{bmatrix} = 0 \\ & f \in [0, cap] \\ & val \geq 0 \end{aligned}$$

ed è quindi riconducibile ad un equivalente problema di *circolazione*, cioè ad un problema di flusso su reti in cui su tutti i nodi è imposto un bilancio pari a zero.

Assumendo di conoscere un flusso ammissibile (ad esempio il flusso nullo), per poter individuare una possibilità di incremento del valore complessivo del flusso è utile introdurre il concetto di *rete residua*, un grafo sui cui archi viene indicata la massima variazione possibile di flusso. Si consideri ad esempio la rete seguente:

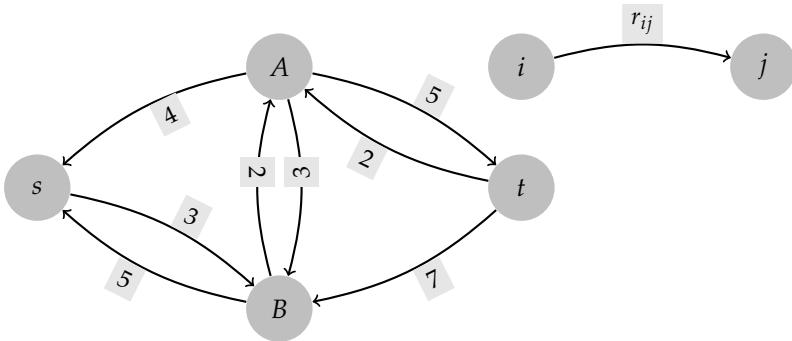
14. IL PROBLEMA DEL MASSIMO FLUSSO



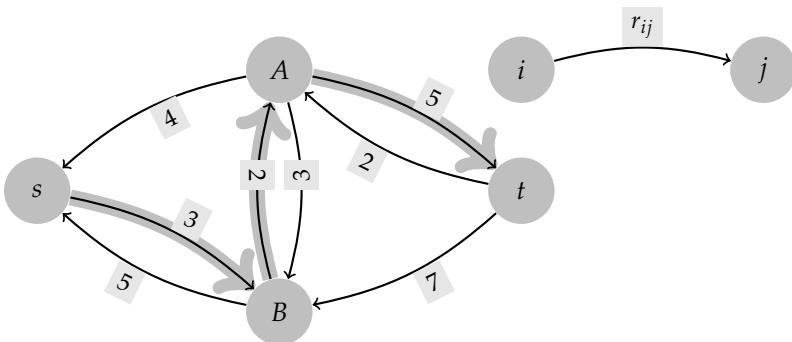
nella quale circolano 9 unità di flusso in totale. E' abbastanza facile vedere in questo esempio che il flusso potrebbe essere aumentato dirottando il flusso dall'arco (A, B) verso l'arco (A, t) : in questo modo si liberano due unità di flusso che possono essere inviate lungo il percorso s, B, t .

Per poter determinare variazioni di flusso di questo tipo in modo sistematico, si utilizza un grafo, detto rete residua, sui cui archi vengono rappresentate le possibili variazioni di flusso. Nel caso in esame, l'arco (s, A) può subire solo una diminuzione di flusso (pari al massimo a 4 unità); lungo l'arco (s, B) si possono avere sia un aumento di flusso, pari al massimo a 3 unità, sia una diminuzione (non superiore a 5). Analogamente su (A, B) il flusso può aumentare fino a 3 unità oppure diminuire al più di due unità. Rappresentando in un grafo orientato tali variazioni massime, con la convenzione che un arco concorde con l'originale rappresenta un possibile aumento di flusso, mentre un arco discorde con l'originale indica una possibile diminuzione, per la rete sopra rappresentata si avrebbe la seguente rete residua:

14.1. Algoritmo di Ford e Fulkerson

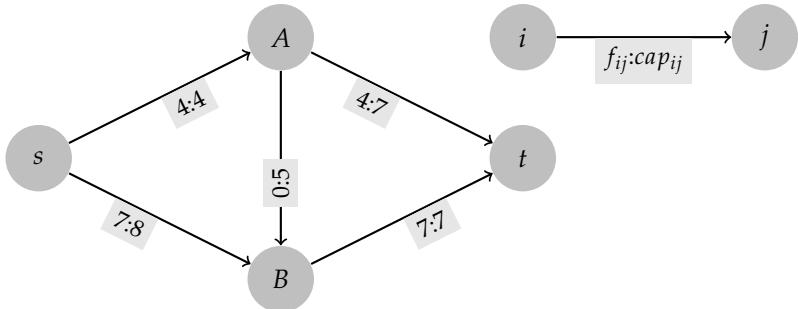


Da questa rete si vede abbastanza facilmente che esiste un percorso orientato da s a t lungo il quale il flusso può essere incrementato di due unità:

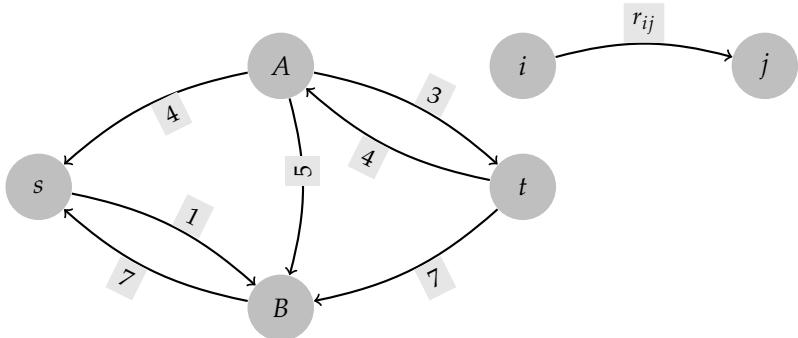


La variazione indicata dal percorso s, b, a, t corrisponde, sul grafo originale, ad un aumento di due unità di flusso sugli archi (s, b) e (a, t) e ad una *diminuzione* della stessa quantità lungo l'arco a, b . Effettuando tale variazione si ottiene il flusso

14. IL PROBLEMA DEL MASSIMO FLUSSO



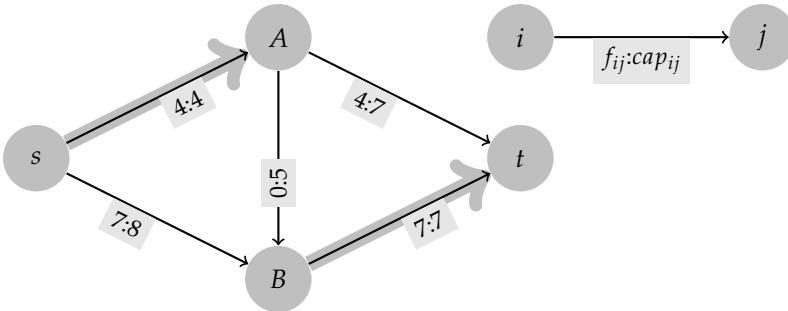
e la rete residua diviene



Si vede facilmente che in questa rete residua non esiste alcun percorso orientato da s a t e, pertanto, non esiste la possibilità di incrementare il flusso. Si è quindi trovata la soluzione al problema del flusso massimo.

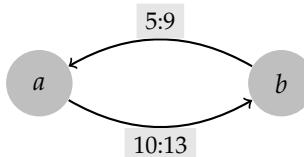
Che il flusso sia il massimo possibile può essere visto anche considerando la seguente rappresentazione:

14.1. Algoritmo di Ford e Fulkerson

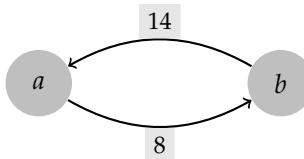


Si vede che tutto il flusso uscente da s deve passare dagli archi evidenziati; essendo tali archi saturi, essi costituiscono un collo di bottiglia che impedisce ad altro flusso di poter raggiungere il nodo terminale.

La rete residua dunque può essere costruita a partire dal grafo originale analizzando le possibilità di incremento (senza violare il vincolo di capacità massima) o di decremento (senza scendere al di sotto della soglia minima). Se nel grafo originale esistessero due archi, con direzione opposta, incidenti sugli stessi nodi



allora, estendendo il ragionamento visto prima, si potrebbe pensare che un eventuale aumento di flusso nella direzione da a verso b potrebbe avvenire sia aumentando fino a tre unità il flusso sull'arco (a, b) , sia diminuendo di 5 unità al massimo il flusso sull'arco (b, a) . Analogamente per quanto riguarda gli incrementi lungo la direzione (b, a) ; la rete residua associata a questa porzione di grafo diverrebbe quindi



In pratica, in presenza di archi doppi come in questo esempio, la massima variazione di flusso lungo una direzione sarà data dalla somma della capacità residua (cioè della differenza fra capacità massima e flusso passante) lungo l'arco con la medesima orientazione, e del valore del flusso lungo l'arco in direzione opposta.

Possiamo quindi introdurre formalmente il concetto come segue:

Definizione 26. Data una rete di flusso $N = \langle V, E, f, b \rangle$ ed un flusso f ammissibile, si definisce *rete residua* un grafo orientato $R = \langle V, E' \rangle$ con gli stessi nodi del grafo originale; gli archi della rete residua sono definiti come:

$$\begin{aligned} E' = & \{(i, j) \in E : f_{ij} < b_{ij}\} \\ & \cup \{(i, j) \in V \times V : (j, i) \in E, f_{ji} > 0\} \end{aligned}$$

Definizione 27. La *capacità residua* di un arco della rete residua R è definita come

$$r_{ij} = b_{ij} - f_{ij} + f_{ji} \quad \forall (i, j) \in E'$$

con la convenzione $f_{ij} = b_{ij} = 0$ se $(i, j) \notin E$.

Per definizione, la capacità residua è sempre strettamente positiva. Partendo dall'analisi della rete residua si può abbastanza facilmente arrivare alla definizione di un algoritmo che, in modo iterativo, proceda identificando percorsi orientati nella rete residua. Ad ogni percorso individuato corrisponde una variazione ammissibile

14.1. Algoritmo di Ford e Fulkerson

di flusso che provoca un aumento del valore del flusso. L'algoritmo si basa sull'applicazione di una semplice procedura di ricerca di un cammino tra il nodo s ed il nodo t nella rete residua; durante la ricerca di tale cammino orientato conviene mantenere una struttura dati che permetta sia la ricostruzione del cammino, sia la determinazione del valore del massimo incremento di flusso possibile. In pratica, sulla rete residua si esegue, ad ogni iterazione, una versione dell'algoritmo di Dijkstra, semplificata per tenere conto del fatto che non si è interessati alla ricerca di un cammino di *costo minimo* ma, semplicemente, di un cammino. Pertanto la procedura di etichettatura assegnerà etichette definitive ai nodi, senza mai necessità di modificare tali etichette. Per la ricostruzione del cammino, come si è visto, è sufficiente memorizzare, nel momento in cui un nodo riceve un'etichetta, il nodo fornitore di tale etichetta. Per quanto concerne invece l'entità della massima variazione possibile, poiché su ogni arco della rete residua è memorizzata la massima variazione applicabile al flusso su tale arco, sarà sufficiente, lungo il cammino, tenere memoria della minima tra le massime variazioni possibili.

Definizione 28. Si dice *cammino incrementabile* o *cammino incrementante* un qualsiasi cammino orientato $s - t$ sul grafo residuo. L'*incremento di flusso* associato a tale cammino è pari alla minima capacità residua degli archi del cammino.

Prima di passare alla descrizione dell'algoritmo per la ricerca del massimo flusso, si può osservare come sia possibile evitare, ad ogni iterazione, di ricostruire la rete residua a partire dal flusso circolante sulla rete. Infatti, una volta individuato un cammino incrementabile, si osservi che l'incremento di flusso associato corrisponde alla massima variazione di flusso possibile, lungo il cammino, sulla rete originale; tale variazione sarà in aumento o in diminuzione rispettivamente nel caso in cui l'arco della rete residua sia o meno orientato come il corrispondente arco sulla rete originale. Pertanto, se ad un percorso incrementabile è associato un incremento pari a

14. IL PROBLEMA DEL MASSIMO FLUSSO

$\Delta > 0$, lungo un arco (i, j) nella rete residua si pone

$$r_{ij} = r_{ij} - \Delta$$

$$r_{ji} = r_{ji} + \Delta$$

con la convenzione che, se a seguito della variazione r_{ij} si annullasse, allora l'arco (i, j) verrebbe eliminato dalla rete residua, mentre nel caso in cui l'arco (j, i) non fosse presente (capacità residua pari a zero), a seguito dell'operazione sopra indicata esso dovrebbe essere incluso nella nuova rete.

La procedura di Ford e Fulkerson può essere descritta in pseudolinguaggio nel modo indicato nell'algoritmo 8.

Nello schema algoritmico sopra riportato sono state indicate due procedure, per l'inizializzazione della rete residua e per la determinazione di un cammino incrementabile, che verranno illustrate tra breve. E' importante però osservare il meccanismo per la ricostruzione del flusso a partire dalla struttura della rete e dai valori associati agli archi della rete residua. Dalla definizione di rete residua discende che

$$r_{ij} = cap_{ij} - f_{ij} + f_{ji}$$

e

$$r_{ji} = cap_{ji} - f_{ji} + f_{ij}$$

dove al solito si sottintende che la capacità di un arco non presente nella rete originale vale 0. Dalle due relazioni sopra indicate deriva

$$f_{ij} - f_{ji} = cap_{ij} - r_{ij} = r_{ji} - cap_{ji}.$$

Un'assegnazione ammissibile di flusso si ottiene immaginando che dei due flussi f_{ij} ed f_{ji} uno soltanto possa essere non nullo. A questo punto discende immediatamente che, se

$$r_{ij} \leq cap_{ij}$$

14.1. Algoritmo di Ford e Fulkerson

Algoritmo di Ford & Fulkerson **Data:** $N = \langle V, E, s, t, cap \rangle, f$:
una rete e un flusso iniziale ammissibile

Result: f : flusso ottimale

$R = \text{CostruisciReteResidua}(N, f);$

while $\text{CamminoIncrementabile}(R, pred, aumento)$ **do**

// Applica l'aumento alla rete residua ;

$j = t;$

while $j \neq s$ **do**

$i = \text{pred}[j];$

$r_{ij} = r_{ij} - \text{aumento};$

$r_{ji} = r_{ji} + \text{aumento};$

$j = i;$

end

end

// Ricostruzione del flusso ottimo;

for $(i, j) \in E$ **do**

if $r_{ij} \leq cap_{ij}$ **then**

$f_{ij} = cap_{ij} - r_{ij};$

$f_{ji} = 0$

else

$f_{ij} = 0;$

$f_{ji} = r_{ij} - cap_{ij}$

end

end

14. IL PROBLEMA DEL MASSIMO FLUSSO

allora si potrà assegnare il flusso come

$$f_{ij} = \text{cap}_{ij} - r_{ij}, \quad f_{ji} = 0$$

mentre se fosse

$$r_{ij} > \text{cap}_{ij}$$

e, pertanto,

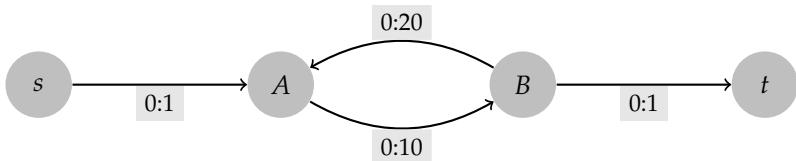
$$r_{ji} > \text{cap}_{ji}$$

si potrebbe porre

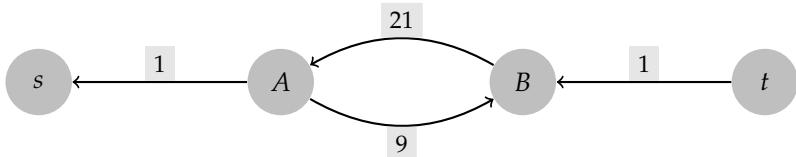
$$f_{ij} = 0, \quad f_{ji} = \text{cap}_{ji} - r_{ji}.$$

Naturalmente questa scelta è obbligata nel caso in cui tra i nodi i e j esista un solo arco. Quando invece esistono nella rete originale sia l'arco (i, j) che l'arco (j, i) , allora esistono soluzioni alternative, come si vede dal seguente semplice esempio.

Esempio 38. Si consideri la rete seguente, nella quale le etichette sugli archi rappresentano il flusso passante e le capacità massime:



La rete residua, al termine dell'algoritmo, risulta la seguente, nella qual le etichette sugli archi rappresentano le capacità residue r_{ij} :

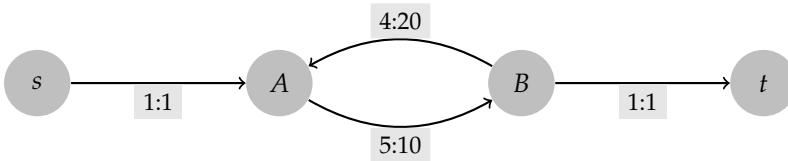


14.1. Algoritmo di Ford e Fulkerson

Per gli archi che connettono i nodi A e B la situazione è la seguente:

$$f_{AB} - f_{BA} = 1$$

Sono quindi ammissibili sia la soluzione $f_{sA} = f_{AB} = f_{Bt} = 1, f_{BA} = 0$, sia infinite altre, come, ad esempio



Per quanto riguarda le altre procedure previste dall'algoritmo, esse possono essere definite come segue.

CostruisciReteResidua **Data:** N, f

Result: R

for $(i, j) \in E$ **do**

$$r_{ij} = cap_{ij} - f_{ij} + f_{ji};$$

$$r_{ji} = cap_{ji} - f_{ji} + f_{ij};$$

if $r_{ij} > 0$ **then**

$$| \quad E_R = E_R \cup \{(i, j)\}$$

end

if $r_{ji} > 0$ **then**

$$| \quad E_R = E_R \cup \{(j, i)\}$$

end

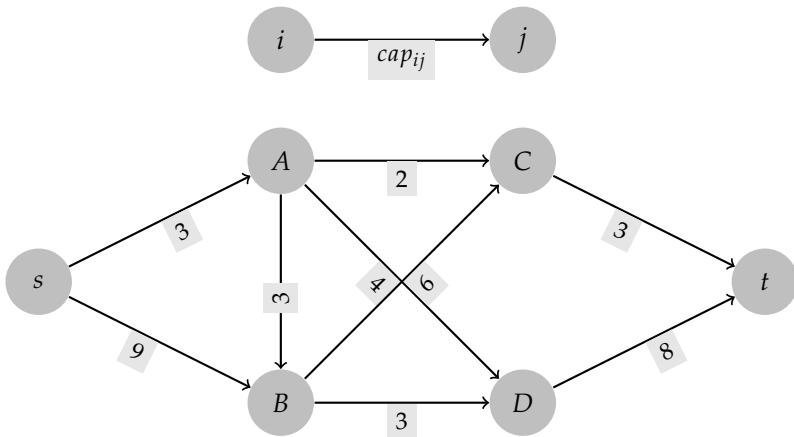
end

La procedura di costruzione di un cammino incrementabile segue abbastanza da vicino lo schema dell'algoritmo di Dijkstra. La differenza principale consiste nel fatto che si tiene traccia esclusivamente dei nodi raggiungibili, nel grafo residuo, dalla sorgente, senza che sia presente un costo da minimizzare e, pertanto, senza

14. IL PROBLEMA DEL MASSIMO FLUSSO

che si renda necessario rivedere le etichette assegnate. Quindi a differenza dall'algoritmo di Dijkstra, i nodi sono divisi in “etichettati” e non; il valore numerico associato all’etichetta di un nodo non rappresenta, come nel metodo di Dijkstra, una stima del costo del cammino ottimo, non essendo presente alcun costo sugli archi. Nel corso della procedura, ad ogni nodo viene assegnato un valore utile per ricostruire il massimo incremento possibile una volta raggiunto il terminale.

Esempio 39. Si consideri il grafo seguente:



Si ipotizzi di iniziare il metodo di Ford & Fulkerson con un flusso nullo (anche se converrebbe iniziare assegnando un flusso ammissibile non nullo); la rete residua iniziale coincide quindi con la rete originale. Utilizzando questa rete residua si può ad esempio ottenere il primo cammino incrementabile procedendo nel modo

14.1. Algoritmo di Ford e Fulkerson

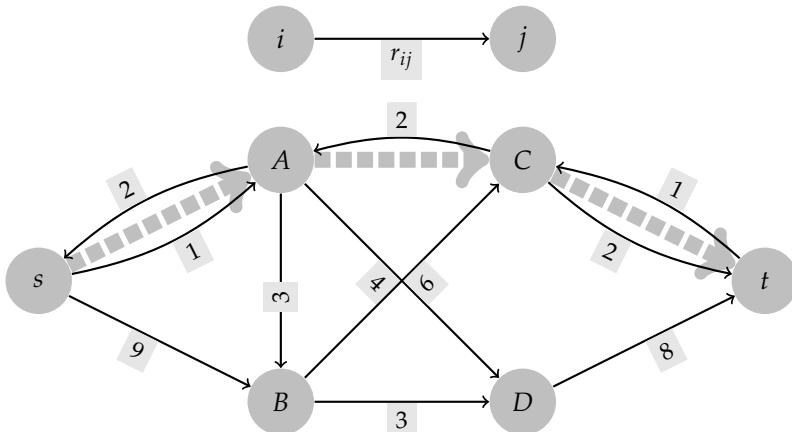
```
CamminoIncrementabile Data:  $R$ 
Result: pred,  $\Delta$ 
//pred: lista dei predecessori lungo il cammino
incrementabile
//  $\Delta$ : flusso incrementabile
Lista L; //Nodi da cui propagare
Set T; // Nodi etichettati Nodo curr;
pred =  $\emptyset$ ;
 $T = \{s\}$ ;
 $L = \{s\}$ ;
aumento[s] =  $+\infty$ ;
while  $t \notin T \&& L \neq \emptyset$  do
    Scegli curr  $\in L$ ;
    for  $j \in V : j \notin T \&& (curr, j) \in E_R$  do
        // Cerca i nodi raggiungibili dal nodo corrente
         $T = T \cup \{j\}$  // Nodo  $j$ : raggiungibile ;
         $L = L \cup \{j\}$  // Occorre propagare la sua etichetta;
        pred[j] = curr;
        aumento[j] = min(aumento[curr],  $r_{curr,j}$ );
    end
     $L = L \setminus \{curr\}$ 
end
if  $t \in T$  then
    | return (true)
else
    | return (false)
end
```

14. IL PROBLEMA DEL MASSIMO FLUSSO

segue:

pred	s	A	B	C	D	t	T	L
\emptyset	∞	—	—	—	—	—	s	s
s		3	9				sAB	AB
A				2	3		$sABCD$	BCD
B							$sABCD$	CD
C					2		$sABCDt$	Dt

Il primo cammino incrementabile è dunque $sACt$, con un incremento di 2 unità di flusso. La rete residua ottenuta inserendo le opportune variazioni dovute al cammino incrementabile appena individuato, evidenziato nel grafico, diviene

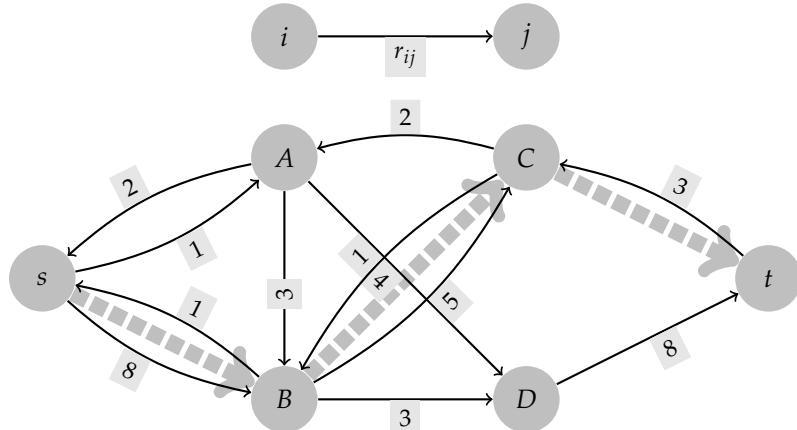


14.1. Algoritmo di Ford e Fulkerson

La seconda iterazione potrebbe procedere nel modo seguente:

pred	s	A	B	C	D	t	T	L
\emptyset	∞	—	—	—	—	—	s	s
s		1	9				sAB	AB
A				1			$sABD$	BD
B				6			$sABCD$	CD
C					1		$sABCDt$	Dt

Il secondo cammino porta ad un incremento di una unità di flusso lungo il percorso $sBCt$ e la situazione diviene:

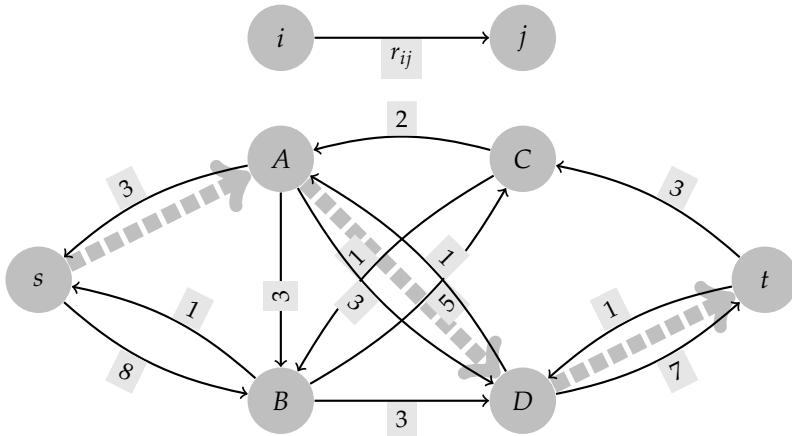


L'iterazione successiva porta alla situazione seguente:

pred	s	A	B	C	D	t	T	L
\emptyset	∞	—	—	—	—	—	s	s
s		1	8				sAB	AB
A				1			$sABD$	BD
B				5			$sABCD$	CD
C							$sABCD$	D
D					1		$sABCDt$	t

14. IL PROBLEMA DEL MASSIMO FLUSSO

(cammino incrementabile: ($sADt$), aumento di 1 unità di flusso)

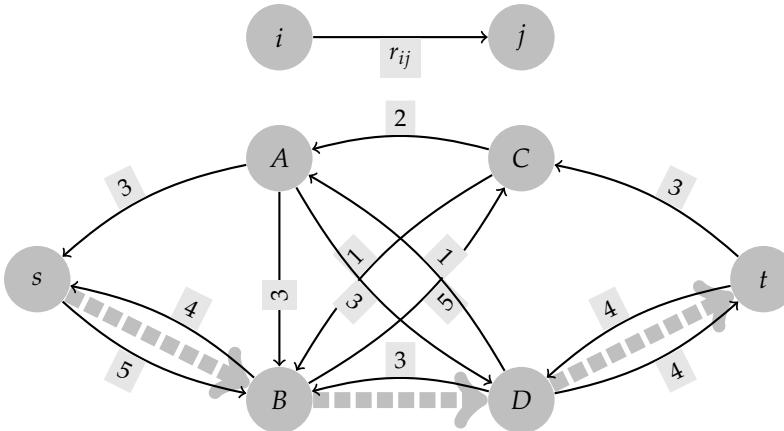


Al passo successivo si ottiene:

pred	s	A	B	C	D	t	T	L
\emptyset	∞	—	—	—	—	—	s	s
s			8				sB	B
B				3	3		$sBCD$	CD
C				2			$sABCD$	AD
A							$sABCD$	D
D					3		$sABCDt$	t

Questo cammino incrementabile, $sBDt$, comporterà un aumento di 3 unità nel flusso totale. La situazione a questo punto è la seguente:

14.1. Algoritmo di Ford e Fulkerson

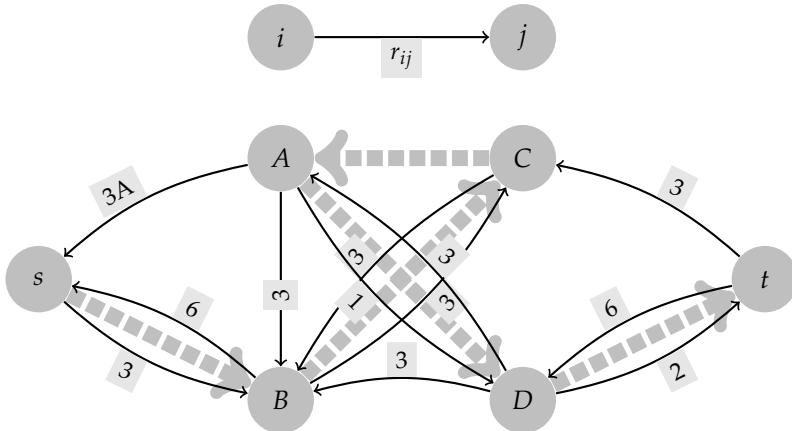


La successiva iterazione procede nel modo seguente:

pred	s	A	B	C	D	t	T	L
\emptyset	∞	—	—	—	—	—	s	s
s			5				sB	B
B				3			sBC	C
C					2		$sABC$	A
A						2	$sABCD$	D
D							$sABCDt$	t

In questa iterazione il cammino incrementabile, con incremento di 2 unità, risulta essere $sBCADt$. La rete residua viene così modificata:

14. IL PROBLEMA DEL MASSIMO FLUSSO

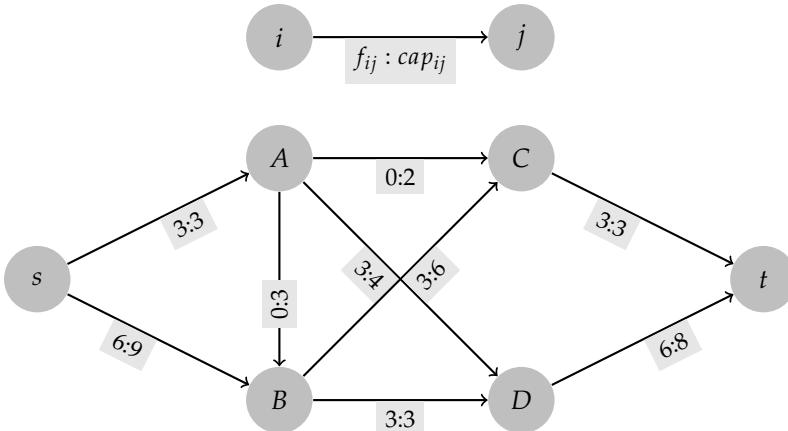


In questo caso, pensando al grafo originale, si può vedere che è stata compiuta una mossa “di ripensamento” che ha portato ad eliminare del flusso dall’arco (A, C) in modo da permettere un miglior sfruttamento delle capacità degli archi. Il flusso che in un primo tempo era stato assegnato a tale arco viene dirottato lungo (A, D) e il risultato complessivo è un aumento globale del flusso in circolazione. Nell’iterazione successiva si ottiene

pred	s	A	B	C	D	t	T	L
\emptyset	∞	—	—	—	—	—	s	s
s			3				sB	B
B				3			sBC	C
C							sBC	\emptyset

e l’algoritmo termina, non essendo più possibile trovare cammini incrementabili. Il flusso massimo è pertanto pari a 9 unità. La ricostruzione del flusso a partire dalla rete residua conduce alla seguente situazione ottimale:

14.1. Algoritmo di Ford e Fulkerson



Finitezza dell'algoritmo di Ford & Fulkerson

L'algoritmo appena presentato procede attraverso una sequenza di costruzioni di cammini nella rete residua. La ricerca di un cammino è sicuramente una procedura finita, ma non è detto che il numero di cammini individuati dall'algoritmo sia anch'esso necessariamente finito. Potrebbe infatti capitare di generare più volte lo stesso cammino incrementabile, a partire da differenti reti residue. Tuttavia, almeno nei casi di interesse pratico, il seguente teorema garantisce la finitezza del metodo.

Teorema 51. *Se il problema del flusso massimo ammette un ottimo finito e le capacità massime sugli archi sono espresse da numeri razionali, l'algoritmo di Ford & Fulkerson, inizializzato con un flusso ammissibile razionale, termina in un numero finito di iterazioni.*

Dimostrazione. Se le capacità degli archi ed i flussi iniziali sono razionali, è sempre possibile ipotizzare che esse siano intere: infatti il problema del flusso massimo è equivalente a quello che si otterrebbe moltiplicando le capacità ed il flusso iniziale per una medesima

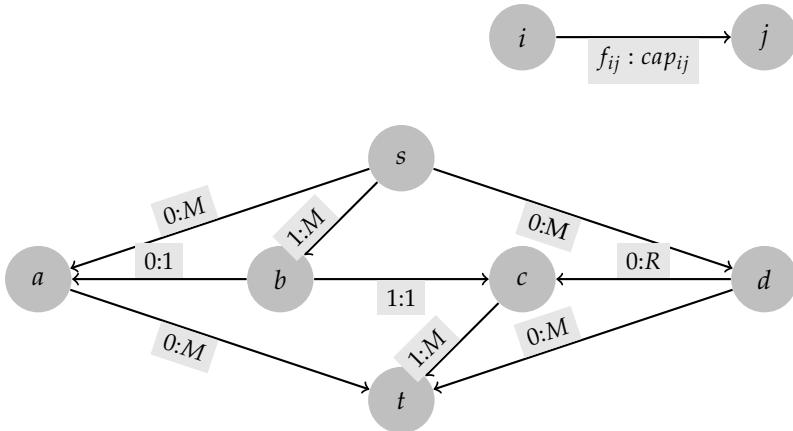
14. IL PROBLEMA DEL MASSIMO FLUSSO

costante. Scegliendo tale costante opportunamente, è possibile trasformare tutte le componenti razionali in interi. Questa operazione corrisponde ad una variazione dell'unità di misura del vettore dei flussi e del valore del flusso totale. In un problema di flusso massimo a capacità e flussi interi, ad ogni iterazione del metodo di Ford & Fulkerson il valore dell'aumento di flusso viene ottenuto dalla scelta, lungo un percorso aumentabile, dell'aumento minimo. Poiché tale valore è il minimo di un numero finito di quantità positive, sarà certamente positivo; inoltre, poiché le capacità massime ed il flusso iniziale sono interi, la capacità residua di ogni arco della rete residua sarà intera. Pertanto l'incremento di flusso, ad ogni iterazione, sarà dato da un numero intero positivo e, pertanto, varrà almeno 1. Se, come ipotizzato, il flusso massimo è finito, l'algoritmo terminerà in un numero di iterazioni pari al più al valore del flusso massimo. \square

Il teorema resta valido anche in alcuni casi più generali. Si possono però costruire esempi di grafi con archi di capacità irrazionale per i quali il metodo di Ford & Fulkerson può generare una successione infinita di iterazioni.

Esempio 40. Si consideri la seguente rete di flusso:

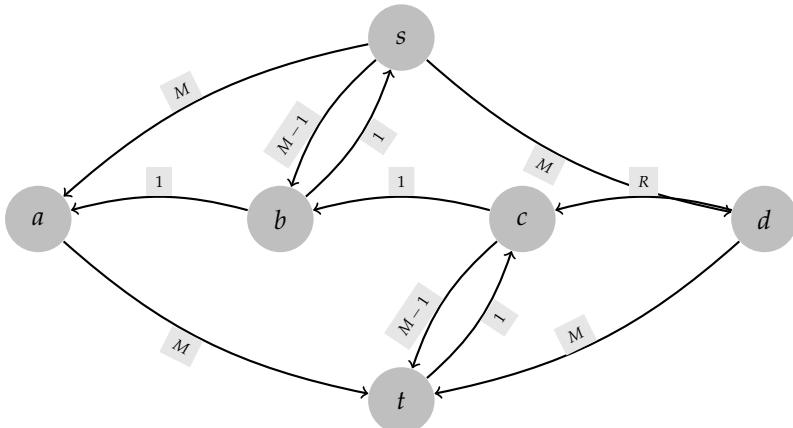
14.1. Algoritmo di Ford e Fulkerson



dove

$$R = \frac{\sqrt{5} - 1}{2}$$

e $M > 2$ è una costante. La rete residua associata a questo flusso è la seguente:

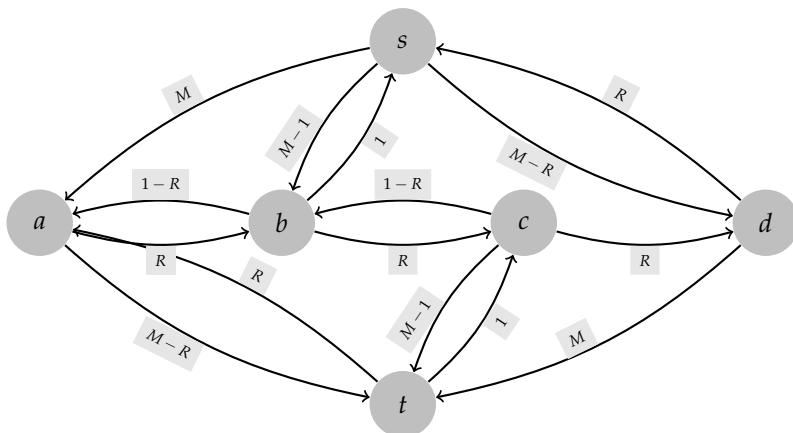


14. IL PROBLEMA DEL MASSIMO FLUSSO

Si consideri il percorso incrementabile

$$P_1 = s - d - c - b - a - t$$

con incremento massimo pari a $R \approx 0.618$. Dopo questo incremento di flusso la rete residua diviene

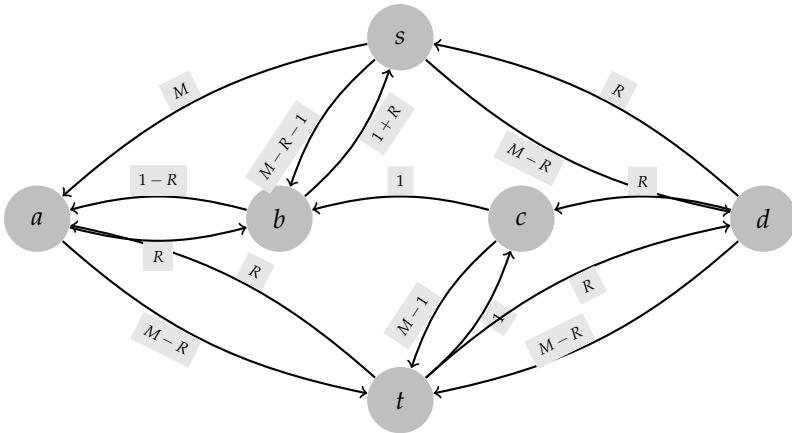


A questo punto si ipotizzi di incrementare il flusso lungo il cammino

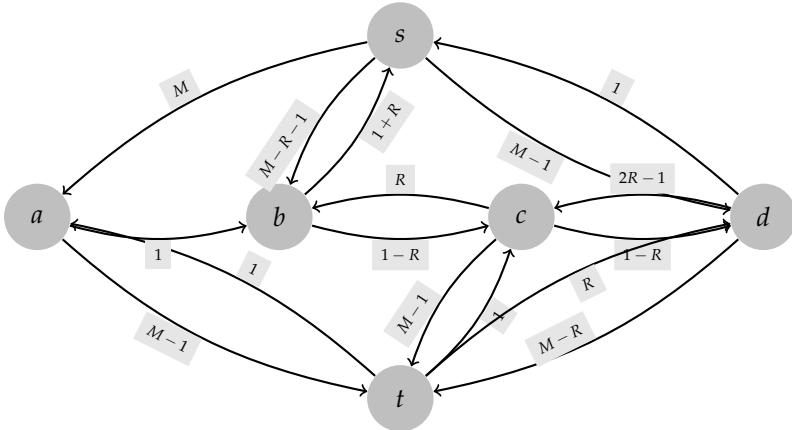
$$P_2 = s - b - c - d - t$$

che ha massimo incremento ammissibile pari a R . La rete residua, dopo tale incremento, assume la forma seguente:

14.1. Algoritmo di Ford e Fulkerson



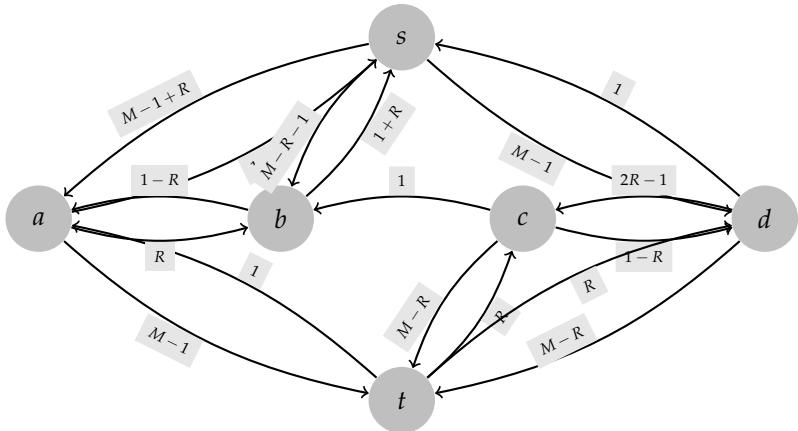
A questo punto si applichi nuovamente il cammino incrementabile P_1 , con incremento pari a $1 - R$:



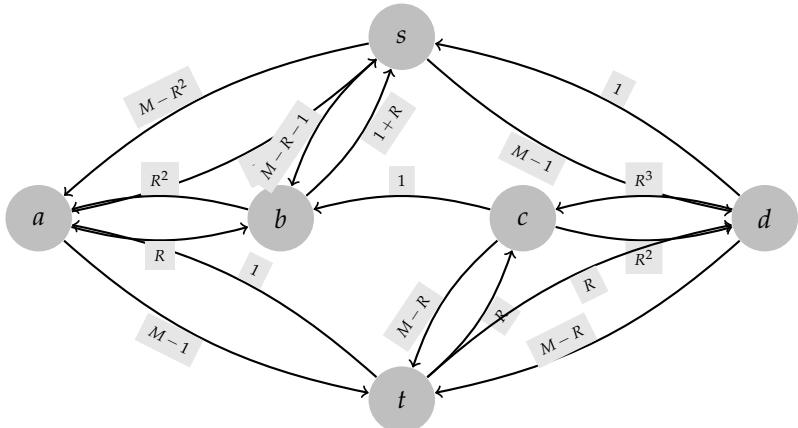
Infine si applichi l'incremento $1 - R$ al cammino

$$P_3 = s - a - b - c - t$$

14. IL PROBLEMA DEL MASSIMO FLUSSO



Sfruttando le caratteristiche della costante R e, in particolare, il fatto che $1 - R = R^2$ e che $R - R^2 = R^3$, la stessa rete residua può anche essere rappresentata come segue:



si può vedere che la situazione è analoga a quella iniziale, dove le capacità residue degli archi $(b,a), (b,c), (d,c)$ passano dai valori

14.1. Algoritmo di Ford e Fulkerson

originali pari rispettivamente a $R^0, 0, R^1$ ai valori $R^2, 0, R^3$; il flusso originale, pari ad 1, è aumentato grazie ai 4 cammini utilizzati, fino a $1 + 2(R + R^2)$.

Si può vedere come la sequenza di cammini incrementabili P_1, P_2, P_1, P_3 possa essere ripetuta all'infinito. Dopo un primo ciclo, il flusso circolante aumenterà di $2(R^3 + R^4)$, e così via. Al limite il flusso tenderà al valore

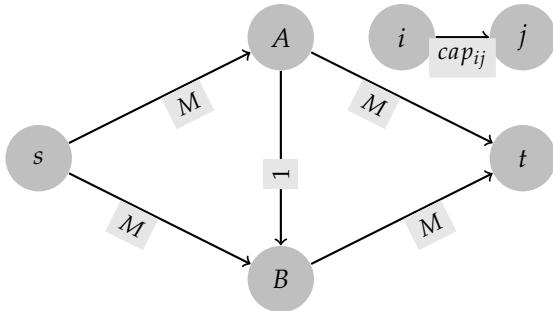
$$\begin{aligned} 1 + 2 \sum_{k=1}^{\infty} R^k &= 1 + 2\left(\frac{1}{1-R} - 1\right) \\ &= 3 + 2R \end{aligned}$$

Da qui si vede che, da un lato, l'algoritmo non possiede terminazione finita. In più, a convergenza, il valore del flusso ottenuto non è quello massimo, che, per la rete in esame, si può vedere facilmente essere pari a $2M + 1$.

Il fatto che la finitezza dell'algoritmo sia garantita solo nel caso di capacità razionali può sembrare sufficiente, nei casi pratici, dato che la rappresentazione comune dei numeri reali sugli elaboratori avviene in parole di memoria di lunghezza finita e, quindi, i numeri utilizzabili dai comuni algoritmi implementati su elaboratore sono sempre numeri razionali. In realtà, il possibile fallimento del metodo nel caso irrazionale deve essere considerato come un segnale d'allarme anche per i casi nei quali la terminazione finita è garantita. In problemi nei quali le capacità associate agli archi hanno valori numerici fortemente differenti fra loro, l'algoritmo può generare sequenze molto lunghe di iterazioni, prima di terminare.

Esempio 41. Si consideri, come esempio di malfunzionamento dell'algoritmo il grafo seguente:

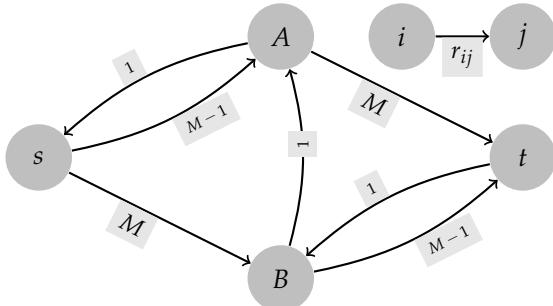
14. IL PROBLEMA DEL MASSIMO FLUSSO



dove M è una costante intera sufficientemente grande (ad esempio, $M = 1000$). E' evidente che il flusso massimo è pari a $2M$ unità. Tuttavia un'implementazione del metodo di Ford & Fulkerson potrebbe procedere come segue: iniziando dal flusso nullo è possibile che un'implementazione dell'algoritmo porti alla scelta del cammino:

$$s \rightarrow A \rightarrow B \rightarrow t.$$

L'incremento massimo lungo tale cammino è pari a 1, a causa della limitata capacità residua sull'arco (A, B) . A questo punto l'arco (A, B) diviene saturo e la rete residua assume la forma

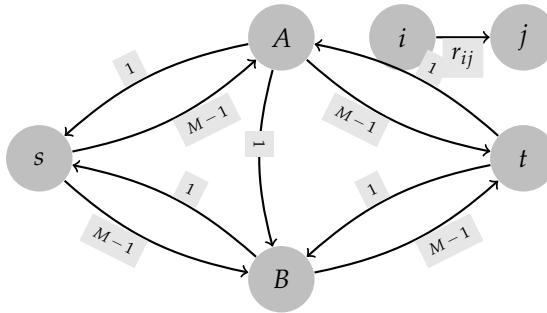


14.1. Algoritmo di Ford e Fulkerson

L'algoritmo potrebbe a questo punto utilizzare il cammino incrementabile

$$s \rightarrow B \rightarrow A \rightarrow t$$

Anche in questo caso l'incremento massimo è pari ad 1; dopo 2 iterazioni quindi il flusso generato è pari a 2 unità, come illustrato nella rete residua seguente:



E' chiaro a questo punto che lo stesso algoritmo potrà procedere ripetendo la stessa sequenza di cammini incrementabili fino a determinare, in $2M$ iterazioni, il flusso ottimo.

Dall'esempio appena presentato si vede che la scelta del cammino incrementabile può influire notevolmente sulla velocità dell'algoritmo: una scelta differente avrebbe portato l'algoritmo a terminare in 2 sole iterazioni. Esistono strategie per la scelta del cammino incrementabile che permettono di evitare situazioni simili a quelle dell'esempio e che garantiscono la terminazione finita. Queste specializzazioni non vengono presentate qui, sia per non appesantire la trattazione, sia perché in realtà l'algoritmo di Ford & Fulkerson è superato in efficienza da diversi altri metodi, per i quali si rimanda alla letteratura scientifica, ad esempio: (Goldberg et al., 1990; Papadimitriou & Steiglitz, 1998; Ahuja et al., 1993; Bertsekas, 1998; Bertsekas & Gallager, 1992).

14.2 Dualità nei problemi di massimo flusso

Il problema del massimo flusso si può, come si è visto, formulare come un problema di *PL*. Vale quindi la teoria della dualità, come per qualsiasi problema di ottimizzazione lineare. Tuttavia, introducendo alcuni importanti concetti relativi alle reti di flusso, si possono dedurre direttamente alcuni teoremi perfettamente analoghi a quelli di dualità debole, forte, complementarità. In realtà questi teoremi possono anche essere ricavati direttamente attraverso un’interpretazione del duale lineare del problema di flusso massimo. Tuttavia, vista l’importanza anche algoritmica del concetto di sezione, si preferisce presentarli qui senza riferimento diretto alla *PL*.

Definizione 29. Data una rete di flusso si definisce *sezione s–t* o *taglio s–t* una qualsiasi bi-partizione dei nodi (W, \bar{W}) tale che $s \in W$, $t \in \bar{W}$.

Nel seguito una sezione *s–t* verrà chiamata semplicemente *sezione*. Ad una sezione corrisponde sempre un insieme di archi “tagliati” dalla sezione, archi cioè che hanno un estremo in un insieme della bipartizione e l’altro nell’insieme complementare.

Definizione 30. Data una sezione (W, \bar{W}) , si definiscono *archi tagliati “in avanti”* dalla sezione, o *archi uscenti dalla sezione*, gli archi

$$(i, j) \in E : i \in W, j \in \bar{W}$$

e, analogamente, si diranno *archi tagliati “all’indietro”*, o *archi entranti nella sezione*, gli archi

$$(i, j) \in E : i \in \bar{W}, j \in W.$$

Definizione 31. Si definisce *capacità di una sezione* la quantità

$$C(W, \bar{W}) = \sum_{(i,j) \in E: i \in W, j \in \bar{W}} b_{ij}$$

14.2. Dualità nei problemi di massimo flusso

pari alla somma delle capacità massime degli archi tagliati in avanti.

Definizione 32. Si definisce *sezione minima* la sezione $s-t$ di capacità minima tra tutte le sezioni di una stessa rete:

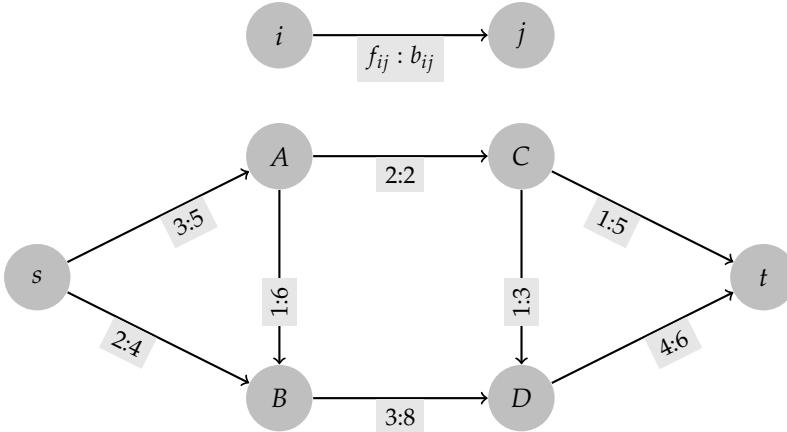
$$(W^*, \bar{W}^*) \in \arg \min_{(W, \bar{W})} C(W, \bar{W})$$

Definizione 33. Si definisce *flusso attraverso una sezione* la quantità

$$F(W, \bar{W}) = \sum_{(i,j) \in E: i \in W, j \in \bar{W}} f_{ij} - \sum_{(j,i) \in E: i \in W, j \in \bar{W}} f_{ji}$$

pari alla differenza fra il flusso circolante sugli archi tagliati in avanti ed il flusso circolante sugli archi tagliati all'indietro.

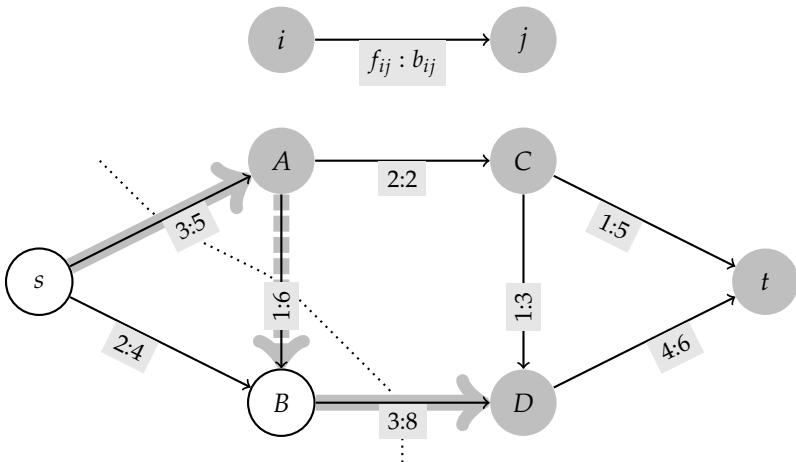
Esempio 42. Data la seguente rete di flusso ed un flusso ammissibile:



la sezione caratterizzata da $W = \{s, A, B\}$ ha capacità pari a 10; il flusso attraverso tale sezione è pari a 5 unità. La sezione associata a $W = \{s, B\}$ ha capacità pari a 13, ed il flusso che la attraversa è pari a $6 - 1 = 5$ unità. Nella figura seguente, i nodi della sezione

14. IL PROBLEMA DEL MASSIMO FLUSSO

hanno colore differente e sono evidenziati gli archi tagliati in avanti ed all'indietro:



Nella figura è anche evidenziata, tramite una linea tratteggiata, la sezione $s - t$.

Vale un'importante proprietà riguardante il flusso attraverso una sezione:

Teorema 52. *Il flusso attraverso una qualsiasi sezione è costante ed è pari al valore del flusso.*

Dimostrazione. Indicando con A la matrice di incidenza del grafo e con v il valore del flusso, è noto che le leggi di conservazione del flusso si possono scrivere, in forma matriciale, nel modo seguente:

$$Af - (e_s - e_t)v = 0$$

dove e_i rappresenta il versore associato al nodo i . Si consideri una generica sezione (W, \bar{W}) e si indichi con $e(W) \in \mathbb{R}^{|V|}$ il vettore

14.2. Dualità nei problemi di massimo flusso

indicatore dell'insieme W , cioè

$$\begin{aligned} e_i(W) &= 1 && \text{se } i \in W \\ e_i(\bar{W}) &= 0 && \text{se } i \in \bar{W} \end{aligned}$$

Si ha quindi

$$\begin{aligned} Af &= (e_s - e_t)v && \text{da cui} \\ e^T(W)Af &= e^T(W)(e_s - e_t)v \\ &= v \end{aligned}$$

dove, nell'ultimo passaggio, si è sfruttato il fatto che $s \in W$ e che $t \in \bar{W}$. Consideriamo, nel membro di sinistra il coefficiente della variabile f associata ad un arco $(i, j) \in E$. Il prodotto tra la riga $e^T(W)$ e la colonna della matrice di incidenza associata all'arco (i, j) è pari a

$$e_i(W) - e_j(W)$$

grazie alla definizione di matrice di incidenza. Tale quantità sarà pari a

$$\begin{aligned} 1 && \text{se } i \in W, j \in \bar{W} \\ -1 && \text{se } i \in \bar{W}, j \in W \\ 0 && \text{altrimenti} \end{aligned}$$

e quindi il prodotto

$$e^T(W)Af$$

sarà pari a

$$\sum_{(i,j) \in E: i \in W, j \in \bar{W}} (+1)f_{ij} + \sum_{(h,k) \in E: h \in \bar{W}, k \in W} (-1)f_{hk}$$

14. IL PROBLEMA DEL MASSIMO FLUSSO

che è la definizione di flusso attraverso la sezione. Si è quindi dimostrato che

$$F(W, \bar{W}) = v$$

□

Il precedente teorema ci autorizza quindi a “misurare” il flusso in una rete osservando il “flusso netto” attraverso una qualsiasi sezione, dove per flusso netto si intende la differenza fra flusso uscente e flusso entrante. Il valore così calcolato è infatti pari a quello ottenuto nella sezione banale $W = \{s\}$, corrispondente alla definizione di valore del flusso.

Il teorema flusso massimo - sezione minima

Il legame tra sezioni $s - t$ e flussi ammissibili è dello stesso tipo di quello esistente tra un problema di *PL* ed il suo duale. Vale infatti il seguente

Teorema 53. (*teorema massimo flusso/minima sezione*). *Si consideri una rete di flusso con capacità finita associata agli archi.*

1. *Il valore di un qualsiasi flusso ammissibile non è mai superiore alla capacità di una qualunque sezione;*
2. *Il flusso massimo ha valore pari a quello della minima capacità di una sezione;*
3. *Un flusso ammissibile ed una sezione (W, \bar{W}) sono ottimali (cioè il flusso è massimo e la sezione è minima) se e solo se il flusso lungo ogni arco uscente dalla sezione è pari alla capacità massima dell'arco ed il flusso lungo ogni arco entrante è pari al minimo.*

14.2. Dualità nei problemi di massimo flusso

Dimostrazione. 1. Si è già visto che il valore v di un flusso ammissibile su una rete è pari al flusso attraverso una qualsiasi sezione, $F(W, \bar{W})$. Si ha quindi che

$$\begin{aligned} v &= F(W, \bar{W}) \\ &= \sum_{i \in W, j \notin W} f_{ij} - \sum_{i \in \bar{W}, j \in W} f_{ij} \\ &\leq \sum_{i \in W, j \notin W} cap_{ij} - \sum_{i \in \bar{W}, j \in W} 0 \\ &= C(W, \bar{W}) \end{aligned}$$

dove nella maggiorazione si è sfruttato il fatto che il flusso è ammissibile e, pertanto, su ogni arco non supera la capacità e non è negativo.

2. Si immagini di conoscere il flusso ottimale f^* di valore v^* . Immaginando di inizializzare l'algoritmo di Ford & Fulkerson con il flusso sugli archi pari al flusso ottimale, sicuramente alla prima iterazione l'algoritmo terminerà senza essere stato in grado di determinare un cammino incrementabile (altrimenti il flusso non sarebbe stato massimo). L'algoritmo di ricerca di un cammino incrementabile, iniziando dal nodo s , genera un insieme T di nodi etichettati che sicuramente non contiene il nodo terminale t . La partizione (T, \bar{T}) è pertanto una sezione $s-t$ (contiene s , ma non t). Se si considerano gli archi attraverso tale sezione, si ha sicuramente $r_{ij} = 0$ per gli archi uscenti (altrimenti, essendo $i \in T$, si sarebbe potuto raggiungere il nodo j). Poiché deve valere la relazione

$$r_{ij} = cap_{ij} - f_{ij} + f_{ji} = 0 \quad \forall i \in T, j \notin T$$

essendo $cap_{ij} - f_{ij}$ e f_{ji} due quantità non negative, la loro somma potrà essere zero se e solo se entrambe sono nulle. Si ha

quindi, per la sezione (T, \bar{T}) ,

$$f_{ij} = cap_{ij}, f_{ji} = 0 \quad \forall i \in T, j \notin T$$

cioè gli archi uscenti sono saturi e quelli entranti sono a flusso nullo. Dalla definizione di capacità della sezione si ha

$$\begin{aligned} C(T, \bar{T}) &= \sum_{i \in T, j \notin T} cap_{ij} \\ &= \sum_{i \in T, j \notin T} f_{ij} - \sum_{i \notin T, j \in T} f_{ij} \\ &= v^* \end{aligned}$$

Si è quindi trovato che la capacità della sezione (T, \bar{T}) è pari al flusso, e la seconda parte del teorema è dimostrata.

3. La condizione sufficiente di ottimalità (se gli archi uscenti sono saturi e quelli entranti nulli allora flusso e sezione sono ottimali) si dimostra seguendo esattamente lo stesso ragionamento visto al punto precedente. Per la condizione necessaria, occorre dimostrare che per qualunque sezione ottimale (e non solo per quella determinata dall'algoritmo di Ford & Fulkerson), vale una proprietà analoga. Sia (W, \bar{W}) una sezione di capacità minima, pari, per quanto visto, al valore del flusso massimo. Deve necessariamente valere

$$\begin{aligned} v^* &= \sum_{i \in W, j \notin W} f_{ij} - \sum_{i \notin W, j \in W} f_{ij} \\ &= \sum_{i \in W, j \notin W} cap_{ij} \end{aligned}$$

cioè deve valere

$$\sum_{i \in W, j \notin W} (cap_{ij} - f_{ij}) + \sum_{i \notin W, j \in W} f_{ij} = 0.$$

14.2. Dualità nei problemi di massimo flusso

Tutti gli addendi nelle sommatorie sopra riportate sono non negativi. Pertanto devono essere tutti nulli, e il teorema è così dimostrato.

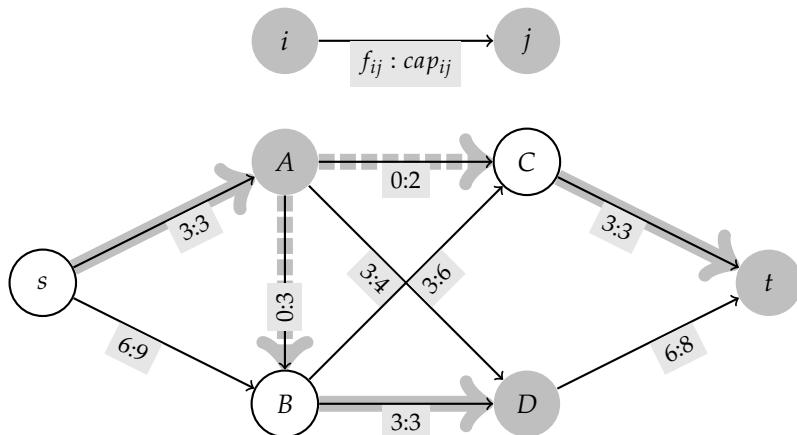
□

Con questa dimostrazione abbiamo anche a disposizione un algoritmo per la determinazione di una sezione di capacità minima: è sufficiente utilizzare l'algoritmo di Ford & Fulkerson: una volta determinato il flusso ottimale, l'ultima etichettatura fornita dalla procedura di ricerca di un cammino incrementabile genera una sezione ottimale.

Corollario 2. *Se l'algoritmo di Ford & Fulkerson termina, i nodi etichettati nel corso dell'ultima iterazione costituiscono una sezione di capacità minima.*

Esempio 39 (cont.)

Nell'esempio di pag. 346, all'ultima iterazione risultano etichettati i nodi s, B, C , che evidentemente costituiscono una sezione.



14. IL PROBLEMA DEL MASSIMO FLUSSO

La capacità di questa sezione è pari a

$$cap_{sA} + cap_{BD} + cap_{Ct} = 3 + 3 + 3 = 9$$

e si nota come gli archi uscenti, cioè $(s, A), (B, D), (C, t)$ siano saturi e quelli entranti, $(A, B), (A, C)$ sono a flusso nullo.

14.3 Reti con capacità minima

In alcuni problemi su alcuni archi è imposto un minimo quantitativo di flusso diverso da zero. Attraverso alcune definizioni e approntando modifiche minime all'algoritmo di Ford & Fulkerson, anche il caso di reti con capacità minima non nulla può essere affrontato in modo simile a quanto visto. Se su ciascun arco di una rete $N = \langle V, E, s, t \rangle$ sono definiti due parametri, C_{min}, C_{max} ed il flusso su ogni arco è vincolato ad assumere valori

$$C_{min_{ij}} \leq f_{ij} \leq C_{max_{ij}}$$

allora si definisce rete residua una rete con gli stessi nodi V , con archi in entrambe le direzioni e con capacità residua generalizzata

$$r_{ij} = (C_{max_{ij}} - f_{ij}) + (f_{ji} - C_{min_{ji}}).$$

dove, nel caso di archi non presenti nel grafo originale, sia la capacità massima che quella minima si intendono pari a zero. Utilizzando questa definizione, l'algoritmo di Ford & Fulkerson può essere direttamente applicato, con l'unica variante da apportare alla fase finale di ricostruzione del flusso. Al termine dell'algoritmo, per determinare un flusso ammissibile ottimale, si può procedere come segue:

$$f_{ij} = \max(C_{min_{ij}}, C_{max_{ij}} - r_{ij}) \quad \forall (i, j) \in E.$$

Definendo la capacità di una sezione come

$$C(W, \bar{W}) = \sum_{i \in W, j \notin W} C_{max_{ij}} - \sum_{j \notin W, i \in W} C_{min_{ji}}$$

14.3. Reti con capacità minima

si può dimostrare il teorema di flusso massimo / sezione minima. Infatti è facile vedere che anche nel caso di reti con capacità minima, il valore del flusso è pari al flusso attraverso una qualunque sezione. Dato un flusso ammissibile ed una qualunque sezione, si ha immediatamente

$$\begin{aligned} v &= \sum_{i \in W, j \notin W} f_{ij} - \sum_{j \notin W, i \in W} f_{ji} \\ &\leq \sum_{i \in W, j \notin W} C_{max_{ij}} - \sum_{j \notin W, i \in W} C_{min_{ji}} \\ &= C(W, \bar{W}) \end{aligned}$$

e, pertanto, il valore del flusso è sempre limitato dalla capacità di una qualunque sezione. Al termine dell'algoritmo di Ford & Fulkerson, i nodi etichettati costituiscono una sezione (T, \bar{T}) . Data la definizione della rete residua, per gli archi uscenti da tale sezione nella rete residua varrà

$$\begin{aligned} r_{ij} &= 0 \quad i \in T, j \notin T \\ &= (C_{max_{ij}} - f_{ij}) + (f_{ji} - C_{min_{ji}}) \end{aligned}$$

Poiché entrambi gli addendi nell'ultima espressione sono non negativi, deve valere

$$f_{ij} = C_{max_{ij}}, \quad f_{ji} = C_{min_{ji}}$$

e pertanto al termine (eventuale) dell'algoritmo di Ford & Fulkerson viene generata una sezione ottimale con archi uscenti saturi ed archi entranti con flusso pari alla minima capacità.

Naturalmente, al contrario di quanto accade nel caso in cui $C_{min} = 0$, la determinazione di un flusso ammissibile iniziale non è banale ed è facile costruire esempi nei quali non esiste alcun flusso ammissibile.

Parte IV

Programmazione Lineare Intera

Introduzione alla Programmazione Lineare Intera

15.1 Introduzione

I problemi di programmazione lineare intera costituiscono un'estensione importantissima del campo di applicazione dei modelli deterministici di programmazione matematica. Attraverso la modellizzazione di variabili a valori discreti è possibile includere in un modello scelte tra diverse alternative e vincoli logici che permettono di formulare modelli di notevole interesse applicativo. Si vedano a questo proposito (Schoen, 2006; Williams, 1999) dove si possono trovare numerosi esempi di problemi lineari con variabili intere.

In questa parte si presenteranno solo alcune semplici ideelegate agli aspetti algoritmici della programmazione lineare intera, senza alcuna pretesa di completezza. Questa cautela, valida per tutto il volume, è sottolineata in particolare in questa parte, dove, per necessità, il discorso relativo alla programmazione intera sarà solo accennato brevemente. La letteratura relativa è molto vasta ed in continua, rapidissima evoluzione – alcuni testi recenti molto interessanti sull'argomento sono (Schrijver, 1998; Papadimitriou & Steiglitz, 1998; Nemhauser & Wolsey, 1988). In questo capitolo saranno solo presentate alcune idee fondamentali sulle quali si basano

15. INTRODUZIONE ALLA PROGRAMMAZIONE LINEARE INTERA

i più attuali metodi di risoluzione.

Prima di analizzare alcuni aspetti teorici, è bene sottolineare che la *programmazione lineare intera* pura, o *PLI*:

$$\begin{aligned} \min c^T x \\ Ax = b \\ x \in \mathbb{Z}^n \\ x \geq 0 \end{aligned}$$

o mista o *MIP*:

$$\begin{aligned} \min c^T x + d^T y \\ Ax + By = b \\ x \geq 0 \\ x \in \mathbb{Z}^n \\ y \geq 0 \end{aligned}$$

è considerevolmente più difficile, dal punto di vista algoritmico, della programmazione lineare, anche se ristretta al caso binario:

$$\begin{aligned} \min c^T \delta \\ A\delta = b \\ \delta \in \{0;1\}^n. \end{aligned}$$

In quest'ultimo problema, ad esempio, è chiaro che una soluzione ottimale può, teoricamente, essere trovata e verificata facilmente: è

sufficiente enumerare i vettori binari δ ad n componenti:

$$\begin{aligned}\delta^1 &= [\ 0 \ 0 \ \dots \ 0 \ 0 \ 0]^T \\ \delta^2 &= [\ 0 \ 0 \ \dots \ 0 \ 0 \ 1]^T \\ \delta^3 &= [\ 0 \ 0 \ \dots \ 0 \ 1 \ 0]^T \\ \delta^4 &= [\ 0 \ 0 \ \dots \ 0 \ 1 \ 1]^T \\ &\vdots \\ \delta^{(2^n)} &= [\ 1 \ 1 \ \dots \ 1 \ 1 \ 1]^T,\end{aligned}$$

per ciascuno dei vettori verificare, tramite sostituzione, l'ammissibilità ($A\delta = b$) e tenere traccia, nel corso della generazione, del vettore ammissibile al quale è associato il costo minimo. Purtroppo questa procedura, teoricamente corretta, non è praticamente realizzabile se non per valori del numero di incognite, n , molto piccoli. Immaginando infatti di avere a disposizione un elaboratore in grado di verificare l'ammissibilità di $2^{20} \approx 10^6$ vettori al secondo, si avrebbero i seguenti tempi di calcolo in funzione di n :

n	$t(s)$	pari a:
20	1	1 secondo
30	2^{10}	17 minuti
50	2^{30}	34 anni
100	2^{80}	38 milioni di miliardi di anni

Se la potenza di calcolo, nei prossimi anni, aumentasse ad esempio di un fattore 1024, portando così la capacità di verifica a $2^{30} \approx 10^9$ vettori al secondo, la situazione diventerebbe la seguente:

n	t	pari a:
20	2^{-10}	1 millisecondo
30	1	1 secondo
50	2^{20}	12 giorni
100	2^{70}	37 000 miliardi di anni

15. INTRODUZIONE ALLA PROGRAMMAZIONE LINEARE INTERA

Invertendo l'analisi, se fissiamo ad esempio in 2^{16} secondi il tempo di calcolo a disposizione (poco più di 18 ore), la seguente tabella mostra la dimensione massima del problema che riusciremmo a risolvere in funzione della velocità della macchina:

velocità (vettori/s)	1024	2048	$\approx 10^6$	$\approx 10^9$	$\approx 10^{12}$
n_{\max}	26	27	36	46	56

Risulta quindi evidente come l'algoritmo proposto, teoricamente funzionante, non potrà mai risolvere problemi realistici. Questa brevissima analisi non ha ovviamente pretese di completezza – si potrebbe benissimo ribattere che, un algoritmo enumerativo del tipo di quello appena visto non potrebbe nemmeno essere pensato, ad esempio, per un problema di *PL* puro (cioè non a variabili intere), problema nel quale le soluzioni ammissibili costituiscono un insieme non numerabile. E tuttavia, come si è visto, esistono algoritmi alternativi alla (impossibile) enumerazione che permettono di risolvere problemi di *PL* di enorme dimensione. Quindi potrebbe sembrare che per la *PL* intera questo algoritmo, seppur finito, non sia il migliore ed altri schemi computazionali possano essere individuati per permettere di risolvere problemi di dimensione mediogrande. In parte questo è vero e, in particolare negli ultimi anni, si è assistito allo sviluppo di metodologie specializzate in grado di determinare soluzioni ottimali per problemi di notevole dimensione. Tuttavia, per la maggior parte dei modelli di *PL* intera, non esistono algoritmi in grado di risolvere *qualunque* esempio con tempi di calcolo che crescano a velocità minore dell'esponenziale col crescere della dimensione. Che possa o meno esistere un algoritmo in grado di risolvere un generico problema di *PL* intero "efficientemente", o, almeno, alcune sottoclassi specializzate, è un tema di ricerca ancora aperto, anche se la comunità scientifica ritiene estremamente improbabile l'esistenza di un algoritmo efficiente di tipo generale per la *PL* intera (si veda, a questo proposito, la letteratura relativa alla *teoria della complessità computazionale*, (Garey & Johnson, 1979)).

15.2. Connessioni tra *PL* e *PL* intera.

Ciò che sembra ormai acquisito è che nel caso dei modelli di ottimizzazione intera, la fase modellistica e quella algoritmica siano fortemente legate. A differenza da quel che accade per i problemi di *PL*, nel caso dell'ottimizzazione discreta una formulazione “matematicamente corretta” di un problema non è in genere sufficiente per poter risolvere il problema stesso. Spesso occorre un lavoro molto accurato di formulazione del problema prima di poter iniziare la sua risoluzione; solo recentemente alcuni risolutori sono stati equipaggiati di fasi di “pre-processing” orientate al miglioramento automatico della formulazione. Cercheremo nel prossimo paragrafo di dare una semplice giustificazione a questa affermazione.

15.2 Connessioni tra *PL* e *PL* intera.

Si consideri un problema di *PL* intera

$$\begin{aligned} z_I^* &= \min c^T x \\ Ax &= b \\ x &\geq 0 \\ x &\in \mathbb{Z}^n \end{aligned} \tag{15.1}$$

Si dice *rilassamento lineare* o *rilassamento continuo* del problema (15.1) il problema di *PL*

$$\begin{aligned} z^* &= \min c^T x \\ Ax &= b \\ x &\geq 0, \end{aligned} \tag{15.2}$$

nel quale è stato eliminato il vincolo di interezza. Questo è effettivamente un rilassamento, secondo la definizione data nel paragrafo 7.2, in quanto la funzione obiettivo nei due problemi (intero e rilassato) è la medesima e, nel rilassamento, l'insieme ammissibile è più

ampio. Siano

$$Q = \{x \in \mathbb{Z}^n : Ax = b, x \geq 0\}$$

$$P = \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$$

gli insiemi ammissibili dei due problemi. Vale naturalmente

$$Q \subseteq P$$

e, di conseguenza,

$$z^* \leq z_I^*.$$

Ricordando il teorema 29 a pag. 150, nel caso del rilassamento continuo, poiché la funzione obiettivo del problema di *PL* intera e del rilassamento è la stessa, vale:

Corollario 3. *Se una soluzione ottimale del problema (15.2) è intera, allora essa è soluzione ottimale del problema intero (15.1).*

Naturalmente vale una proprietà analoga per i problemi interi misti. A volte, quindi, è possibile risolvere un problema di *PL* intera utilizzando gli strumenti della *PL*. Questo non autorizza però a pensare che il legame tra un problema intero ed il suo rilassamento lineare sia in generale così forte. Un tipico esempio di ragionamento errato è quello secondo il quale per risolvere un problema intero è sufficiente risolvere il rilassamento continuo ed arrotondare la soluzione all'intero più vicino. Se questo può essere non lontano dal vero per problemi nei quali le variabili intere assumono valori molto grandi, un ragionamento simile è assolutamente errato nel caso in cui i valori interi siano “piccoli”, come, ad esempio, nel caso estremo della *PL* binaria.

15.2. Connessioni tra PL e PL intera.

Esempio 43. Si consideri il semplicissimo problema

$$\begin{aligned} \min & -x - y \\ \text{s.t. } & -2x + 5y \leq 5 \\ & 2x - 2y \leq 1 \\ & x, y \geq 0 \quad \text{interi.} \end{aligned}$$

rappresentato in Figura 15.1, dove le soluzioni ammissibili sono indicate da punti evidenziati ed in grigio è rappresentato il politopo del rilassamento lineare.

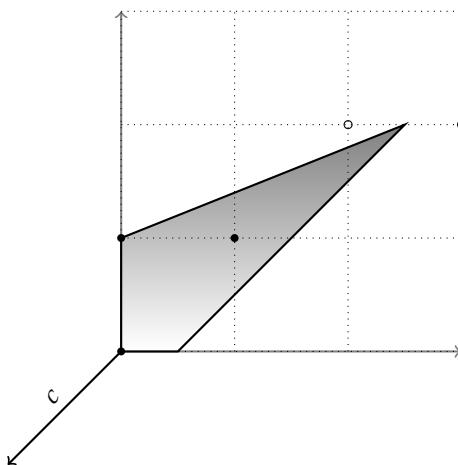


Figura 15.1: Un problema di PLI ed il suo rilassamento lineare

Si vede facilmente che la soluzione

$$x = \frac{5}{2}, y = 2$$

è ottimale per il rilassamento lineare, con valore $-\frac{9}{2}$. Sia per ispezione diretta, che tramite rappresentazione grafica, si vede anche

15. INTRODUZIONE ALLA PROGRAMMAZIONE LINEARE INTERA

facilmente che le uniche soluzioni ammissibili del problema intero sono

$$(0,0), (0,1), (1,1)$$

alle quali sono associati i valori $0, -1, -2$ rispettivamente. Quindi la soluzione ottimale del problema intero è

$$x = y = 1.$$

Arrotondando la soluzione del rilassamento lineare, sia per difetto che per eccesso, non si ottiene nemmeno una soluzione ammissibile (i punti corrispondenti agli arrotondamenti sono rappresentati da piccole circonferenze nella figura).

Si è visto quindi che arrotondare una soluzione non intera non è, in generale, una buona strategia per risolvere i problemi di *PL* intera. Oltretutto è bene ricordare che anche l'arrotondamento è un'operazione computazionalmente molto onerosa: in un problema con n variabili, se, data una soluzione di un rilassamento, si desiderasse controllare l'ammissibilità di un arrotondamento, occorrerebbe effettuare due tentativi per la prima variabile (arrotondamento per eccesso o per difetto), due per la seconda, ..., giungendo così a dover valutare 2^n soluzioni possibili, senza alcuna garanzia di poterne trovare una ammissibile. Questa osservazione dovrebbe essere sufficiente a far comprendere come un utilizzo non ragionato dello strumento dell'arrotondamento non possa portare a risultati apprezzabili. Tuttavia è importante osservare che tra rilassamento e problema intero il legame esiste ed una sua comprensione approfondita costituisce la chiave per la risoluzione dei problemi di ottimizzazione intera di grande dimensione. Si può osservare innanzitutto che uno stesso problema di *PL* intera può essere descritto in modo equivalente in diversi modi.

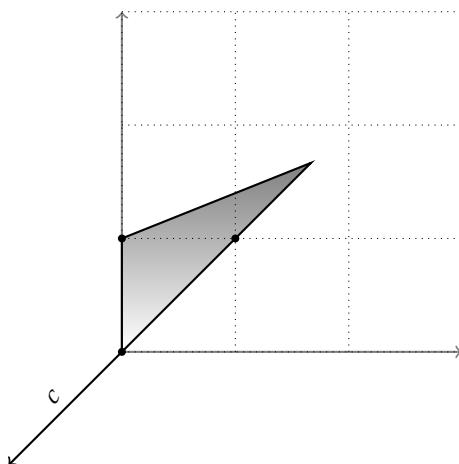
Esempio 43 (cont.)

Riprendendo l'esempio appena fatto, si può vedere facilmente che

15.2. Connessioni tra PL e PL intera.

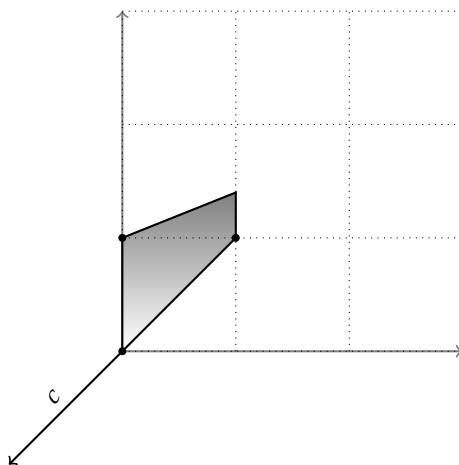
il problema dato è equivalente a ciascuno dei problemi seguenti:

$$\begin{aligned} \min & -x - y \\ -2x + 5y & \leq 5 \\ x - y & \leq 0 \\ x, y & \geq 0 \quad \text{interi.} \end{aligned}$$



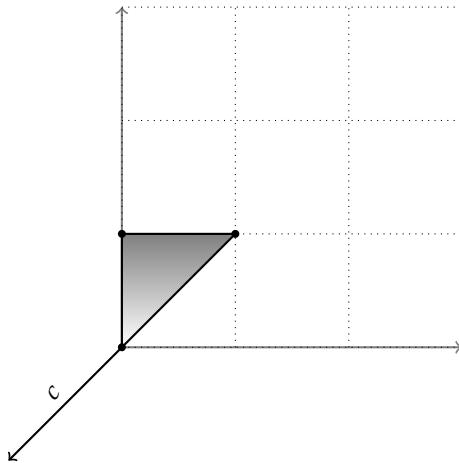
$$\begin{aligned} \min & -x - y \\ -2x + 5y & \leq 5 \\ x - y & \leq 0 \\ x & \leq 1 \\ x, y & \geq 0 \quad \text{interi.} \end{aligned}$$

15. INTRODUZIONE ALLA PROGRAMMAZIONE LINEARE INTERA



$$\begin{aligned} \min & -x - y \\ \text{s.t.} & y \leq 1 \\ & x - y \leq 0 \\ & x, y \geq 0 \quad \text{interi.} \end{aligned}$$

15.2. Connessioni tra PL e PL intera.



Nel primo caso, si è rafforzato il vincolo $2x - 2y \leq 1$ osservando che esso è equivalente al vincolo $x - y \leq 0.5$ e che, essendo per ipotesi x e y intere, il vincolo può essere sostituito da $x - y \leq 0$ senza che alcuna soluzione intera ammissibile venga eliminata. Le altre due formulazioni possono essere ottenute facendo ragionamenti simili o analizzando una rappresentazione grafica del problema. L'ultima rappresentazione è particolarmente importante, perché l'insieme

$$\{(x, y) : y \leq 1, x - y \leq 0, x, y \geq 0\}$$

è il triangolo che ha come vertici $(0, 0)$, $(0, 1)$, $(1, 1)$; tutti i vertici del poliedro sono dunque soluzioni intere e, pertanto, qualunque sia la funzione obiettivo da ottimizzare, il metodo del simplex produrrà sempre soluzioni intere.

Siano date, per un problema di PL intera, due diverse rappre-

sentazioni:

$$\begin{aligned} \min c^T x \\ Ax = b \\ x \geq 0 \quad \text{intero} \end{aligned}$$

e

$$\begin{aligned} \min c^T x \\ A'x = b' \\ x \geq 0 \quad \text{intero} \end{aligned}$$

e siano

$$P = \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$$

e

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

i poliedri corrispondenti ai due rilassamenti lineari. Le due rappresentazioni sono equivalenti (dal punto di vista della *PL* intera) se

$$P \cap \mathbb{Z}^n = P' \cap \mathbb{Z}^n.$$

La nozione di equivalenza corrisponde alla correttezza formale della rappresentazione: entrambe le forme sono modi corretti di rappresentare un dato problema di *PL* intera. Se tra due formulazioni equivalenti vale la relazione

$$P \subset P'$$

allora si dice che la prima formulazione è più *forte* (più stringente) della seconda. Tra le due formulazioni, in generale sarà preferibile quella più forte. Se si indicano con z^* e z'^* i valori ottimali dei due problemi rilassati, vale infatti evidentemente:

$$z'^* \leq z^* \leq z_I^*.$$

15.2. Connessioni tra PL e PL intera.

Quindi un rilassamento più forte approssima meglio il valore dell'ottimo. Esiste sempre una rappresentazione minimale, detta anche *rappresentazione ideale*: tale rappresentazione è caratterizzata dal fatto che il poliedro associato al rilassamento della rappresentazione ideale è contenuto nel poliedro associato al rilassamento di qualunque rappresentazione equivalente del problema stesso. Vale infatti la seguente, fondamentale, proprietà:

Teorema 54. (*Schrijver, 1998*) *Ogni problema di PL intera a coefficienti razionali*

$$\min \left\{ c^T x : x \in P \cap \mathbb{Z}^n \right\}$$

ammette una rappresentazione ideale costituita dal problema

$$\begin{aligned} & \min c^T x \\ & x \in \text{conv}(P \cap \mathbb{Z}^n) \end{aligned}$$

dove con $\text{conv}(X)$ si indica la chiusura convessa dell'insieme X , cioè l'insieme convesso minima contenente X . Inoltre $\text{conv}(P \cap \mathbb{Z}^n)$ è un poliedro i cui vertici sono interi.

Il teorema enunciato afferma quindi che per ogni problema di PL intera (ma una proprietà simile vale per i problemi misti) esiste una rappresentazione ideale, data dalla chiusura convessa delle soluzioni ammissibili intere; se si conoscesse tale rappresentazione ideale, il problema potrebbe essere risolto mediante il metodo del simplesso, dato che il rilassamento avrebbe vertici tutti interi.

Naturalmente in genere la rappresentazione ideale non è nota; più spesso però, succede che anche se la rappresentazione ideale è nota, essa non è utilizzabile per via del numero di vincoli: frequentemente il numero di vincoli della rappresentazione ideale cresce esponenzialmente con la dimensione del problema. Quindi semplicemente valutare esplicitamente se una soluzione soddisfa o meno i vincoli richiederebbe un tempo esponenziale e anche memorizzare la forma di tali vincoli richiederebbe sia un tempo che uno spazio di

memoria esponenziali. Tuttavia esistono algoritmi basati sulla conoscenza di una rappresentazione “implicita” del poliedro ideale, che procedono utilizzando algoritmi specializzati per la verifica di ammissibilità che non richiedono il calcolo esplicito dell'espressione dei vincoli.

Esistono classi importanti di problemi di *PL* intera per i quali è nota la rappresentazione ideale e per i quali tale rappresentazione è “compatta”, cioè è ottenuta tramite un numero non esponenziale di vincoli.

Definizione 34. Una matrice M quadrata, con elementi tutti interi tale che

$$\det(M) = \pm 1$$

si dice *unimodulare*.

Teorema 55. Se una matrice quadrata M ad elementi interi è unimodulare, allora la sua inversa M^{-1} è intera.

Dimostrazione. Data una matrice M invertibile, l'inversa può essere ottenuta mediante l'espressione

$$M^{-1} = \frac{\text{adj}(M)}{\det(M)} \quad (15.3)$$

dove la matrice $\text{adj}(M)$, detta *aggiunta* o *aggiogata*, ha l'elemento di posto ij pari al determinante della matrice ottenuta da M eliminando la i -esima colonna e la j -esima riga, moltiplicato per -1 se la somma $i + j$ è dispari. Se, come ipotizzato, gli elementi di M sono interi, il numeratore nella (15.3) è intero ed essendo il denominatore pari a -1 oppure a $+1$, segue l'interezza dell'inversa. \square

Corollario 4. Se in un problema di *PL* in forma standard tutte le matrici di base sono unimodulari ed il termine noto è intero, qualunque sia il vettore di costi se esiste una soluzione ottimale, esiste una soluzione ottimale di base intera.

15.2. Connessioni tra PL e PL intera.

Definizione 35. Una matrice intera si dice *totalmente unimodulare* se ogni sua sottomatrice quadrata invertibile è unimodulare.

Per un problema di *PL* intera che ha come matrice dei vincoli una matrice totalmente unimodulare, quindi, la rappresentazione è ideale – non vale però il viceversa: in genere una rappresentazione ideale di un problema di *PL* intera non è caratterizzata da una matrice dei vincoli totalmente unimodulare. La definizione di totale unimodularità è di poca utilità pratica, essendo ovviamente molto difficile verificare la proprietà richiesta dal corollario per singoli problemi. Ciò che si può invece fare è identificare classi di problemi per i quali la proprietà di unimodularità totale può essere garantita a priori. Vale una condizione sufficiente:

Teorema 56. (Papadimitriou & Steiglitz, 1998). *Si consideri una matrice A con m righe ed n colonne ed elementi tutti pari a $0, +1, -1$. Se ogni colonna di A possiede al massimo due elementi non nulli, allora le seguenti affermazioni sono equivalenti:*

1. *la matrice A è totalmente unimodulare;*
2. *l'insieme delle righe di A può essere ripartito in due sottoinsiemi in modo tale che:
 - a) se una colonna contiene due elementi uguali a $+1$ oppure due elementi uguali a -1 , questi non appartengono alla stessa partizione;
 - b) se una colonna contiene un elemento pari a $+1$ ed uno pari a -1 tali elementi appartengono alla stessa partizione*

Esempio 44. La matrice

$$A = \begin{bmatrix} +1 & -1 & -1 & & & +1 \\ & +1 & -1 & -1 & -1 & +1 \\ -1 & +1 & & +1 & +1 & -1 & +1 \\ & & & & -1 & & \end{bmatrix}$$

è totalmente unimodulare. Per verificarlo si può usare la partizione delle righe di A in: A_1 = prima, terza e quarta riga; A_2 = seconda riga.

Corollario 5. *La matrice di incidenza (nodi-archi) di un grafo orientato è totalmente unimodulare.*

Dimostrazione. E' sufficiente considerare la partizione banale $A_1 = A, A_2 = \emptyset$. \square

Si può anche dimostrare la seguente proprietà:

Teorema 57. *Se una matrice A è totalmente unimodulare, lo sono anche le matrici*

$$\begin{bmatrix} A & I \end{bmatrix}, \begin{bmatrix} A & -I \end{bmatrix}, \begin{bmatrix} A \\ I \end{bmatrix}, \begin{bmatrix} A \\ -I \end{bmatrix}.$$

Quindi tutti i problemi di flusso su reti, cammino di costo minimo, flusso massimo, flusso di costo minimo, trasporto, assegnamento bipartito, sono caratterizzati da una matrice dei coefficienti totalmente unimodulare. Una proprietà simile, ma solo per il caso di problemi di flusso rappresentabili senza limiti superiori al flusso su ciascun arco, era stata già dimostrata nel paragrafo 11.3 a pag. 267, quando si è dimostrata la corrispondenza tra basi ed alberi nei problemi di flusso.

Si consideri il seguente problema: dati due insiemi di nodi O e D :

$$\begin{aligned} \min \quad & \sum_{i \in O, j \in D} c_{ij} x_{ij} \\ \text{s.t.} \quad & \sum_{i \in O} x_{ij} = 1 \\ & \sum_{j \in D} x_{ij} = 1 \\ & x_{ij} \in \{0, 1\} \end{aligned}$$

15.2. Connessioni tra PL e PL intera.

A prima vista non sembra un problema di flusso, ma se si introduce una sorgente fittizia collegata a tutti i nodi di O , con flusso uscente $|O|$ e capacità massima sugli archi pari a 1, e, analogamente, si introduce un nodo terminale fittizio, con archi provenienti dai nodi di D e capacità massima pari ad 1, il problema diviene un problema di flusso su rete di costo minimo. La matrice ad esso associata è quindi totalmente unimodulare e resta tale anche per il problema originale: se da una matrice totalmente unimodulare si eliminano alcune righe la matrice risultante è ancora totalmente unimodulare.

Quindi se il vincolo $\delta_{ij} \in \{0,1\}$ viene sostituito dal vincolo $0 \leq \delta_{ij} \leq 1$, le soluzioni di base saranno sicuramente binarie. Questo modello prende il nome di *assegnamento bipartito* o *abbinamento*: ogni soluzione di base del problema corrisponde ad una corrispondenza biunivoca tra elementi di O ed elementi di D ; tra questi si sceglie quello di costo totale minimo.

Un modo alternativo per dimostrare che il poliedro dell'assegnamento bipartito è ideale è considerare la struttura della matrice dei coefficienti, che, ad esempio in un problema 3×3 , ha la forma

$$\begin{bmatrix} 1 & 1 & 1 & & & \\ & & & 1 & 1 & 1 & \\ 1 & & & 1 & & 1 & \\ & 1 & & 1 & & 1 & \\ & & 1 & & 1 & & 1 \\ & & & 1 & & & 1 \end{bmatrix}$$

e si vede immediatamente come suddividere le righe in due insiemi che separano i due elementi pari ad 1 che compaiono in ogni colonna.

E' importante osservare però che se in un problema di flusso viene aggiunto anche un solo vincolo che non sia un'equazione di bilancio o un limite superiore o inferiore sul valore del flusso su un arco, la proprietà di interezza delle soluzioni di base generalmente diventa falsa.

Ad esempio, in un problema di cammino di costo minimo, se esistono cicli con costo totale negativo, la soluzione del rilassamento lineare è illimitata inferiormente. Tuttavia in un problema del genere si potrebbe richiedere di determinare un cammino *semplice* di costo minimo, cioè un cammino nel quale nessun arco e nessun nodo viene ripetuto. La non ripetizione del passaggio attraverso un arco si ottiene facilmente imponendo un limite superiore pari ad 1 su ogni arco:

$$\begin{aligned} \min c^T f \\ Af = e_s - e_t \\ f \leq 1 \\ f \geq 0 \end{aligned}$$

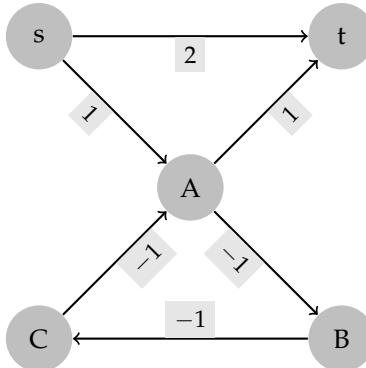
e la rappresentazione resta ideale. Tuttavia se si impone anche la non ripetizione dei nodi, occorre aggiungere vincoli che limitino il flusso totale entrante o uscente da ogni nodo:

$$\begin{aligned} \min c^T f \\ Af = e_s - e_t \\ \sum_{j:(i,j) \in E} f_{ij} \leq 1 \\ f \leq 1 \\ f \geq 0 \end{aligned}$$

Questa aggiunta, pur necessaria, non è sufficiente a garantire che il risultato sia un cammino semplice. Potrebbe infatti essere generata una soluzione sconnessa, composta da un cammino $s \rightarrow t$ e da uno o più cicli sconnessi da tale cammino:

Esempio 45. Si consideri il grafo in figura

15.2. Connessioni tra PL e PL intera.



rappresentante un problema di cammino di costo minimo. E' evidente che il cammino semplice di costo minimo ha costo totale 2 e può essere indifferentemente scelto come $s \rightarrow t$ oppure come $s \rightarrow A \rightarrow t$. Tuttavia la soluzione di costo minimo, anche con l'imposizione del flusso uscente da ogni nodo non superiore a 1, corrisponde ad un flusso unitario sull'arco $s \rightarrow t$ e da un flusso unitario sul ciclo sconnesso $A \rightarrow B \rightarrow C \rightarrow A$. Il costo di questa soluzione è -1.

Per ottenere un cammino semplice occorre quindi inserire altri vincoli. Ad esempio si può utilizzare un vincolo che imponga che, scelto comunque un sottoinsieme di nodi non comprendente s e t , il flusso su ogni arco interno a tale insieme non superi il flusso uscente da tale insieme. In questo modo, se esistesse un ciclo, il vincolo verrebbe violato:

$$f_{ij} \leq \sum_{u \in U, v \notin U} f_{uv} \quad \forall U \subset V : i, j \in U, s, t \notin U.$$

Esempio 45 (cont.)

Per il grafo dell'esempio precedente, i vincoli da inserire sono:

$$\begin{aligned} U = \{A, B\} : f_{AB} &\leq f_{At} + f_{BC} \\ U = \{B, C\} : f_{BC} &\leq f_{CA} \\ U = \{A, C\} : f_{CA} &\leq f_{At} + f_{AB} \\ U = \{A, B, C\} : f_{AB} &\leq f_{At} \\ &f_{BC} \leq f_{At} \\ &f_{CA} \leq f_{At} \end{aligned}$$

che, dopo una breve analisi, possono essere ridotti al solo vincolo

$$f_{BC} \leq f_{At}$$

essendo i restanti vincoli implicati da questo o dalla legge di conservazione del flusso.

Questo vincolo aggiuntivo, pur non introducendo nella matrice dei coefficienti alcun elemento diverso da 0, ±1, distrugge la totale unimodularità.

Risolvendo il rilassamento continuo del problema precedente si ottiene una soluzione di costo pari a $\frac{1}{2}$, nella quale su ogni arco il flusso è $\frac{1}{2}$; in altre parole il flusso unitario si divide in due parti uguali a 0.5, la prima delle quali percorre l'arco $s \rightarrow t$, l'altra il cammino $s \rightarrow A \rightarrow B \rightarrow C \rightarrow A \rightarrow t$.

15.3 Metodi di taglio

I primi algoritmi utilizzati per la programmazione lineare intera si basarono sulla definizione di taglio.

Definizione 36. Dato un insieme S , una disequazione $a(x) \leq 0$ si dice *valida* per S se vale $a(x) \leq 0 \quad \forall x \in S$.

Definizione 37. Dato un problema di ottimizzazione

$$\begin{aligned} & \min f(x) \\ & x \in S \end{aligned}$$

ed un suo rilassamento

$$\begin{aligned} & \min g(x) \\ & x \in Q \end{aligned}$$

una disequazione valida per S , $a(x) \leq 0$ si dice *taglio* se, indicata con

$$\bar{x} \in \arg \min_{x \in Q} g(x)$$

una soluzione ottimale del rilassamento, si ha

$$a(\bar{x}) > 0.$$

Quindi una disequazione valida è un taglio se è violata in corrispondenza dell'ottimo di un rilassamento. Basandosi su questa definizione molti metodi di ottimizzazione lineare intera sono basati sulla costruzione di rilassamenti opportuni, sulla loro risoluzione e sull'aggiunta alla formulazione originale di disequazioni che garantiscano di "tagliare" l'ottimo del rilassamento, senza tagliare alcuna soluzione ammissibile del problema originale. Il problema

$$\begin{aligned} & \min f(x) \\ & x \in S \\ & a(x) \leq 0 \end{aligned}$$

è ancora equivalente all'originale, ma, se il vincolo aggiunto è un taglio, la formulazione è più forte.

La tecnica sulla quale si basano molti metodi di taglio è indicata nel seguente teorema.

Teorema 58. *Sia dato un problema di ottimizzazione in cui tutte le variabili sono vincolate ad essere intere e non negative. Se*

$$a^T x \leq b$$

è una disequazione valida, allora anche

$$\sum_{j=1}^n \lfloor a_j \rfloor x_j \leq \lfloor b \rfloor$$

è valida.

Dimostrazione. Infatti, ricordando che per qualsiasi numero reale α vale

$$\lfloor \alpha \rfloor \leq \alpha$$

grazie alla non negatività delle variabili x anche la seguente disequazione è valida:

$$\sum_{j=1}^n \lfloor a_j \rfloor x_j \leq b$$

Il membro di sinistra di questa disequazione è intero. Poiché se α è un intero e $\beta \in \mathbb{R}$, vale

$$\alpha \leq \beta \text{ sse } \alpha \leq \lfloor \beta \rfloor$$

segue immediatamente che

$$\sum_{j=1}^n \lfloor a_j \rfloor x_j \leq \lfloor b \rfloor$$

□

15.3. Metodi di taglio

Storicamente uno dei primi e tra i più importanti metodi di taglio è il metodo dei tagli di Gomory (Gomory, 2010). La tecnica di Gomory viene riportata qui per fornire un semplice esempio di come si possano costruire disequazioni valide e tagli per problemi lineari interi. Gomory stesso, dopo aver proposto la tecnica, ne indicò l'inefficacia, dovuta soprattutto alla lentezza di convergenza ed all'instabilità numerica. La tecnica restò quindi per molti anni solo un esercizio teorico. Recentemente (Balas, Ceria, & Cornuéjols, 1993) i tagli di Gomory sono stati in un certo riscoperti e adattati ai moderni risolutori di problemi interi. Oggi nessun risolutore di buona qualità fa a meno di questi tagli.

Si consideri un problema del tipo

$$\begin{aligned} \min c^T x \\ Ax = b \\ x \geq 0, \quad \text{intero} \end{aligned}$$

ed il rilassamento continuo

$$\begin{aligned} \min c^T x \\ Ax = b \\ x \geq 0. \end{aligned}$$

Sia \bar{x} una soluzione ottimale del rilassamento continuo. Naturalmente se $\bar{x} \in \mathbb{Z}^n$, allora si è trovata una soluzione ottima per il problema intero. Se viceversa almeno una componente di \bar{x} non è intera, si può definire un taglio. Si consideri il dizionario finale del metodo del simplex applicato al rilassamento lineare:

$$x_B = \bar{x}_B - A_B^{-1} A_N x_N$$

o, equivalentemente, in forma esplicita

$$x_{B_i} + \sum_{j \in N} \bar{a}_{ij} x_j = \bar{x}_{B_i}, \tag{15.4}$$

dove con \bar{A} si è indicata la matrice $A_B^{-1}A_N$. Poiché il rilassamento lineare è un problema standard, si ha in particolare $x \geq 0$; quindi, seguendo la stessa traccia del teorema dimostrato sopra, si ha

$$x_{B_i} + \sum_{j \in N} \lfloor \bar{a}_{ij} \rfloor x_j \leq \bar{x}_{B_i}.$$

Questa è una disequazione valida sia per il problema intero che per il rilassamento lineare. Non può mai essere un taglio quindi. Ma se si sfrutta il fatto che le incognite nel problema intero devono assumer valore intero, allora si ha che tutto il membro di sinistra deve essere intero. Ricordando che se α è un numero intero. Si ottiene quindi

$$x_{B_i} + \sum_{j \in N} \lfloor \bar{a}_{ij} \rfloor x_j \leq \lfloor \bar{x}_{B_i} \rfloor \quad (15.5)$$

Teorema 59. *La disequazione (15.5) è un taglio se e solo se \bar{x}_{B_i} non è intero.*

Dimostrazione. Infatti la soluzione ottimale di base del rilassamento lineare è data da

$$\begin{aligned} x_{B_i} &= \bar{x}_{B_i} \\ x_N &= 0 \end{aligned}$$

sostituendo questa soluzione in (15.5), si ottiene

$$\bar{x}_{B_i} + \sum_{j \in N} \lfloor \bar{a}_{ij} \rfloor 0 \leq \lfloor \bar{x}_{B_i} \rfloor$$

che, se \bar{x}_{B_i} non è intero, non può valere. \square

Spesso al posto di questo taglio se ne introduce uno equivalente, ottenuto sottraendo membro a membro la (15.5) dalla (15.4):

$$\sum_{j \in N} (\bar{a}_{ij} - \lfloor \bar{a}_{ij} \rfloor) x_j \geq (\bar{x}_{B_i} - \lfloor \bar{x}_{B_i} \rfloor)$$

15.3. Metodi di taglio

e introducendo i simboli

$$f_{ij} = (\bar{a}_{ij} - \lfloor \bar{a}_{ij} \rfloor)$$

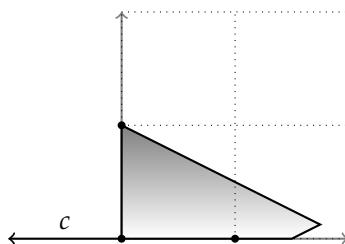
$$f_{x_{B_i}} = (\bar{x}_{B_i} - \lfloor \bar{x}_{B_i} \rfloor)$$

per le *parti frazionarie* dei termini \bar{a}_{ij} e \bar{x}_{B_i} , si ottiene il cosiddetto *taglio di Gomory frazionario*

$$\sum_{j \in N} f_{ij} x_j \geq f_{x_{B_i}}$$

Esempio 46. Si consideri il problema

$$\begin{aligned} & \min -x \\ & x + 2y \leq 2 \\ & 2x - 4y \leq 3 \\ & x, y \geq 0 \quad \text{interi.} \end{aligned}$$



Aggiungendo le variabili di slack che, essendo i coefficienti dei vincoli ed i termini noti interi, saranno anch'esse intere, si ottiene il dizionario

$$\begin{aligned} z &= -x \\ s &= 2 - x - 2y \\ t &= 3 - 2x + 4y \end{aligned}$$

15. INTRODUZIONE ALLA PROGRAMMAZIONE LINEARE INTERA

che, dopo due operazioni di pivot, diviene:

$$z = -\frac{7}{4} + \frac{1}{2}s + \frac{1}{4}t$$

$$y = \frac{1}{8} + \frac{1}{8}t - \frac{1}{4}s$$

$$x = \frac{7}{4} - \frac{1}{2}s - \frac{1}{4}t$$

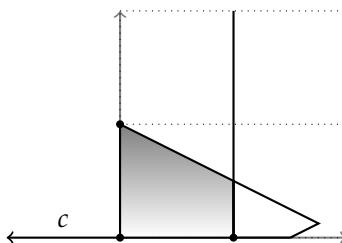
Sia la variabile in base x che y hanno valore non intero. E' quindi possibile effettuare un taglio di Gomory a partire da una qualsiasi delle due equazioni corrispondenti. Utilizzando l'euristica secondo la quale a volte è meglio effettuare un taglio sulla variabile che ha la parte frazionaria maggiore, si può generare un taglio a partire dall'equazione associata ad x :

$$\frac{1}{2}s + \frac{1}{4}t \geq \frac{3}{4}$$

Sfruttando la definizione delle variabili di slack s e t , si può ricostruire il taglio di Gomory nello spazio delle variabili originali del problema, al solo scopo di poterlo visualizzare. Si ottiene in particolare

$$\frac{1}{2}(2 - x - 2y) + \frac{1}{4}(3 - 2x + 4y) \geq \frac{3}{4}$$

da cui con facili passaggi si arriva al vincolo $x \leq 1$:



15.3. Metodi di taglio

Inserito il taglio nel dizionario, si ha:

$$\begin{aligned}z &= -\frac{7}{4} + \frac{1}{2}s + \frac{1}{4}t \\y &= \frac{1}{8} + \frac{1}{8}t - \frac{1}{4}s \\x &= \frac{7}{4} - \frac{1}{2}s - \frac{1}{4}t \\w &= -\frac{3}{4} + \frac{1}{2}s + \frac{1}{4}t\end{aligned}$$

Questo dizionario è pronto per l'esecuzione del metodo del simplexo duale; effettuando un'operazione di pivot duale si ottiene:

$$\begin{aligned}z &= -1 + w \\y &= -\frac{1}{4} - \frac{1}{2}w + \frac{1}{4}t \\x &= 1 - w \\s &= \frac{3}{2} + 2w - \frac{1}{2}t\end{aligned}$$

che ancora non è ottimale; dopo un'altra operazione di pivot si ottiene

$$\begin{aligned}z &= -1 + w \\t &= 1 + 2w + 4y \\x &= 1 - w \\s &= 1 + w - 2y\end{aligned}$$

che è ottimale e intera.

L'esempio sopra riportato non è indicativo della reale efficienza del metodo dei tagli di Gomory che, in genere, è molto scarsa.

L'esempio può essere utilizzato per illustrare come un miglioramento della formulazione di un problema possa portare ad una risoluzione più efficiente.

15. INTRODUZIONE ALLA PROGRAMMAZIONE LINEARE INTERA

Esempio 46 (cont.)

Il vincolo

$$2x - 4y \leq 3$$

presente nel problema precedente può essere migliorato osservando che entrambi i membri possono essere divisi per il massimo comune denominatore dei coefficienti del membro di sinistra:

$$x - 2y \leq \frac{3}{2};$$

questo vincolo è valido sia per il problema intero che per il suo rilassamento lineare. Tuttavia, grazie all'interezza di x e y , il vincolo può essere rafforzato in

$$x - 2y \leq 1.$$

Riscrivendo il dizionario, si ottiene

$$z = -x$$

$$s = 2 - x - 2y$$

$$t = 1 - x + 2y$$

da cui, con due operazioni di pivot, si arriva a

$$\begin{aligned} z &= -\frac{3}{2} + \frac{1}{2}t + \frac{1}{2}s \\ x &= \frac{3}{2} - \frac{1}{2}t - \frac{1}{2}s \\ y &= \frac{1}{4} - \frac{1}{4}s + \frac{1}{4}t \end{aligned}$$

Un taglio di Gomory sull'equazione associata ad x ha la forma

$$\frac{1}{2}t + \frac{1}{2}s \geq \frac{1}{2}$$

15.3. Metodi di taglio

mentre quello costruito sull'equazione associata ad y è

$$\frac{1}{4}s + \frac{3}{4}t \geq \frac{1}{4}.$$

Si vede che il secondo taglio è più debole del primo. Aggiungendo il primo taglio alla formulazione si ottiene il dizionario

$$\begin{aligned} z &= -\frac{3}{2} + \frac{1}{2}t + \frac{1}{2}s \\ x &= \frac{3}{2} - \frac{1}{2}t - \frac{1}{2}s \\ y &= \frac{1}{4} - \frac{1}{4}s + \frac{1}{4}t \\ w &= -\frac{1}{2} + \frac{1}{2}s + \frac{1}{2}t \end{aligned}$$

e, dopo un'operazione di pivot, si arriva al dizionario ottimo:

$$\begin{aligned} z &= -1 + w \\ x &= 1 - w \\ y &= -\frac{1}{2}w + \frac{1}{2}t \\ s &= 1 + 2w - t \end{aligned}$$

Si è visto che, anche in un semplice esempio come questo, una formulazione migliore del problema iniziale può portare ad un aumento della velocità anche per un metodo notoriamente lento quale quello basato sui tagli di Gomory. L'esempio precedente e la tecnica dei tagli di Gomory illustrano le possibilità di utilizzo di strumenti di tipo generale per la riformulazione di problemi di *PL* intera. Si è visto in pratica come per generare migliori formulazioni sia utile inserire disequazioni valide generate dalla combinazione di vincoli validi e dall'utilizzo dell'ipotesi di interezza; anche la tecnica dei

15. INTRODUZIONE ALLA PROGRAMMAZIONE LINEARE INTERA

tagli di Gomory può essere vista come un metodo per il miglioramento della formulazione iniziale. Un metodo per la generazione di disequazioni valide è il seguente:

1. Generazione di un vincolo ridondante: ad esempio, da un sistema di disequazioni del tipo

$$Ax \leq b$$

si generano vincoli ridondanti mediante combinazioni lineari a coefficienti non-negativi (combinazioni coniche):

$$\lambda^T A x \leq \lambda^T b, \quad \lambda \geq 0$$

2. Utilizzando l'ipotesi di non negatività di x si genera il vincolo ridondante

$$\left[\begin{array}{c} \lambda^T A \\ \lambda \end{array} \right] x \leq \left[\begin{array}{c} \lambda^T b \\ \lambda \end{array} \right], \quad \lambda \geq 0$$

3. Sfruttando l'ipotesi di interezza il vincolo può essere rafforzato:

$$\left[\begin{array}{c} \lambda^T A \\ \lambda \end{array} \right] x \leq \left[\begin{array}{c} \lambda^T b \\ \lambda \end{array} \right].$$

Esempio 46 (cont.)

Sommendo i vincoli validi

$$x + 2y \leq 2$$

$$x - 2y \leq 1$$

15.4. Il metodo Branch & Bound

si ottiene il vincolo ridondante $2x \leq 3$, equivalente a $x \leq 3/2$, da cui, imponendo l'interezza, si ottiene $x \leq 1$; il dizionario risulta ora

$$\begin{aligned} z &= -x \\ s &= 2 - x - 2y \\ t &= 1 - x + 2y \\ u &= 1 - x \end{aligned}$$

e, con un'unica operazione di pivot, si giunge all'ottimo intero, senza bisogno di effettuare tagli di Gomory:

$$\begin{aligned} z &= -1 + u \\ s &= 1 + u - 2y \\ x &= 1 - u \\ t &= u + 2y \end{aligned}$$

Esistono molte varianti del metodo dei tagli di Gomory, così come varie proprietà teoriche che garantiscono la finitezza, la possibilità di eliminare alcuni tagli dopo alcune iterazioni, la possibilità di inserire tagli equivalenti con coefficienti interi, ecc.

In questo volume non si desidera approfondire ulteriormente l'argomento; è sufficiente ricordare che i metodi basati sulla generazione di tagli sono attualmente considerati gli strumenti più efficaci nella risoluzione di problemi interi a grande dimensione.

15.4 Il metodo Branch & Bound

L'algoritmo più comunemente implementato nel software commerciale per la *PL* intera si basa sull'idea di un'enumerazione implicita delle soluzioni ammissibili del problema. Lo schema su cui si basano i metodi di tipo Branch & Bound è il seguente: si consideri un

15. INTRODUZIONE ALLA PROGRAMMAZIONE LINEARE INTERA

problema di *PL* intera:

$$\begin{aligned} & \min c^T x \\ & Ax = b \\ & x \geq 0 \quad \text{intero.} \end{aligned}$$

Sia $P = \{x : Ax = b, x \geq 0\}$ il poliedro associato al rilassamento lineare

1. Viene risolto il rilassamento lineare del problema originale determinando $x^* \in \arg \min \{c^T x : x \in P\}$
2. se x^* è intera, allora essa è anche la soluzione ottimale del problema;
3. altrimenti, sia x_j^* una componente di x^* non intera. Il problema originale è equivalente a

$$\begin{aligned} & \min c^T x \\ & x \in P \cap \mathbb{Z}^n \cap \left(\left\{ x : x_j \leq \lfloor x_j^* \rfloor \right\} \cup \left\{ x : x_j \geq \lceil x_j^* \rceil \right\} \right) \end{aligned}$$

4. Dal problema originale vengono originati due sottoproblemi:

$$\begin{aligned} & \min c^T x \\ & x \in P \\ & x_j \leq \lfloor x_j^* \rfloor \\ & x \quad \text{intero} \end{aligned}$$

15.4. Il metodo Branch & Bound

e

$$\begin{aligned} \min c^T x \\ x \in P \\ x_j \geq \lceil x_j^* \rceil \\ x \quad \text{intero} \end{aligned}$$

ai quali viene applicato, in modo recursivo, il medesimo algoritmo; delle due soluzioni ottimali eventualmente trovate, quella di valore inferiore è la soluzione ottimale del problema.

Lo schema del metodo prevede dunque la generazione di un albero binario di problemi di PL intera: ogni nodo corrisponde ad un problema di PL intera. Le foglie dell'albero di Branch & Bound, cioè quei nodi che non è più necessario suddividere, sono caratterizzati dal verificarsi di una delle seguenti condizioni:

1. Il poliedro associato al rilassamento lineare è vuoto;
2. la soluzione del rilassamento lineare è intera.

Esiste tuttavia un terzo caso che permette di troncare l'esplo-razione dell'albero, e che corrisponde all'operazione di “*boun-ding*”:

3. Sia x_I^* una soluzione ammissibile intera del problema, il cui valore è $z_I = c^T x_I^*$; se il valore della soluzione del rilassa-mento lineare è maggiore o uguale a z_I , il nodo corrente vie-ne tagliato – non occorre cioè proseguire nell'operazione di suddivisione.

Il motivo per il quale l'operazione di bounding è valida è sem-plicemente legato al fatto che nel passaggio dal problema associato ad un nodo ai due problemi generati dal nodo stesso, l'insieme

ammissibile del rilassamento lineare si riduce. Poiché il valore dell'ottimo di un rilassamento costituisce, nei problemi di minimo, un limite inferiore al valore dell'ottimo e nel passaggio da un problema ai due sottoproblemi tale valore non può diminuire, ne segue che se ad una data iterazione il valore del rilassamento è peggiore del valore di una soluzione ammissibile (che costituisce un limite superiore all'ottimo), nessuna eventuale soluzione intera di valore strettamente minore può essere trovata nel sottoalbero che ha come radice il nodo corrente.

Più precisamente, sia z_I^* il valore ottimale del problema, z_I il valore di una soluzione ammissibile del problema, \bar{z} il valore ottimale di un sottoproblema generato dall'algoritmo Branch & Bound e \bar{z}_R il valore del rilassamento di tale sottoproblema. Vale:

$$\begin{array}{ll} z_I \geq z_I^* & \text{per definizione di ottimalità} \\ \bar{z} \geq z_I^* & \text{per definizione di sottoproblema} \\ \bar{z}_R \leq \bar{z} & \text{per definizione di rilassamento} \end{array}$$

Si ha quindi che se $\bar{z}_R \geq z_I$, segue

$$\begin{aligned} \bar{z} &\geq \bar{z}_R \\ &\geq z_I \\ &\geq z^* \end{aligned}$$

e pertanto la soluzione ottimale o è già stata trovata, oppure certamente non si troverà risolvendo esattamente il sottoproblema corrente.

Se il problema ha una funzione obiettivo a coefficienti interi, allora anche il valore ottimale della soluzione sarà intero. Quindi, nelle operazioni di bounding, si potrà a volte rafforzare il criterio come segue: se per un sottoproblema vale che

$$\lceil \bar{z}_R \rceil \geq z_I$$

15.4. Il metodo Branch & Bound

allora il sottoproblema può essere considerato come implicitamente visitato, in quanto nessuna soluzione strettamente migliore di quella di valore z_I potrà essere trovata.

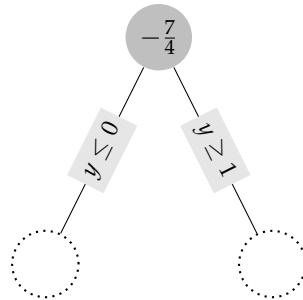
Esempio 47. Si consideri il problema

$$\begin{aligned} \min -x \\ x + 2y &\leq 2 \\ 2x - 4y &\leq 3 \\ x, y &\geq 0 \quad \text{interi.} \end{aligned}$$

già risolto con il metodo dei tagli di Gomory; naturalmente, come in precedenza, anche qui valgono tutte le considerazioni fatte a proposito della buona formulazione iniziale del problema, che potrebbe essere facilmente sostituito con uno equivalente per il quale il rilassamento lineare fornisce già la soluzione ottimale. Tuttavia, per esigenze didattiche, vediamo come il problema può essere risolto mediante la tecnica di Branch & Bound. Inizialmente viene risolto il rilassamento lineare, che porta al dizionario seguente:

$$\begin{aligned} z &= -\frac{7}{4} + \frac{1}{2}s + \frac{1}{4}t \\ y &= \frac{1}{8} + \frac{1}{8}t - \frac{1}{4}s \\ x &= \frac{7}{4} - \frac{1}{2}s - \frac{1}{4}t \end{aligned}$$

Entrambe le variabili in base sono frazionarie; occorre dunque sceglierne una sulla quale effettuare l'operazione di branching. Scegliendo ad esempio y , si generano due sottoproblemi, caratterizzati dall'aggiunta del vincolo $y \leq 0$ e $y \geq 1$ rispettivamente



Scegliendo di proseguire con il nodo di sinistra dell'albero di Branch & Bound, si ottiene il dizionario

$$\begin{aligned} z &= -\frac{7}{4} + \frac{1}{2}s + \frac{1}{4}t \\ y &= \frac{1}{8} + \frac{1}{8}t - \frac{1}{4}s \\ x &= \frac{7}{4} - \frac{1}{2}s - \frac{1}{4}t \\ y &\leq 0 \end{aligned}$$

che, grazie all'introduzione di una variabile di slack w , diviene

$$\begin{aligned} z &= -\frac{7}{4} + \frac{1}{2}s + \frac{1}{4}t \\ y &= \frac{1}{8} + \frac{1}{8}t - \frac{1}{4}s \\ x &= \frac{7}{4} - \frac{1}{2}s - \frac{1}{4}t \\ w &= -\frac{1}{8} - \frac{1}{8}t + \frac{1}{4}s. \end{aligned}$$

Questo dizionario è pronto per l'esecuzione del metodo del simplesso duale (sarebbe stato anche possibile sfruttare il fatto che il vincolo $y \leq 0$ corrisponde a fissare a 0 il valore dell'incognita y).

15.4. Il metodo Branch & Bound

Operando un passo di pivot si ottiene

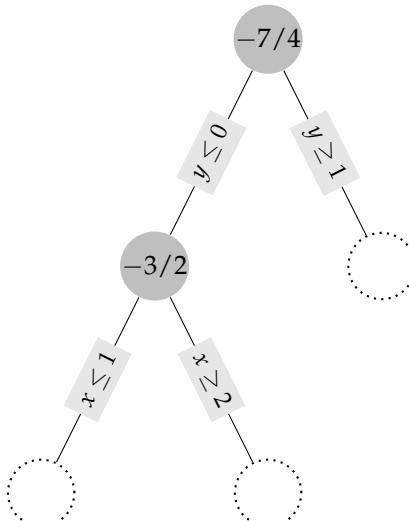
$$z = -\frac{3}{2} + 2w + \frac{1}{2}t$$

$$y = -w$$

$$x = \frac{3}{2} - 2w - \frac{1}{2}t$$

$$s = \frac{1}{2} + 4w + \frac{1}{2}t$$

che è ancora frazionario. Si può quindi ulteriormente suddividere questo ramo, ad esempio, utilizzando il valore frazionario di x , con la generazione di due problemi caratterizzati dai vincoli $x \leq 1$ e $x \geq 2$, rispettivamente:



Ancora si pone il problema della scelta dell'ordine con cui visitare i nodi dell'albero di Branch & Bound. Seguendo una strategia spesso impiegata in pratica, si decida ad esempio di proseguire con

15. INTRODUZIONE ALLA PROGRAMMAZIONE LINEARE INTERA

la visita in profondità (strategia LIFO o Last-In First-Out) di analisi di uno dei due nodi generati dall'ultima operazione di suddivisione. Si immagini ad esempio di proseguire lungo il ramo associato al vincolo $x \leq 1$:

$$\begin{aligned} z &= -\frac{3}{2} + 2w + \frac{1}{2}t \\ y &= -w \\ x &= \frac{3}{2} - 2w - \frac{1}{2}t \\ s &= \frac{1}{2} + 4w + \frac{1}{2}t \\ x &\leq 1 \end{aligned}$$

che equivale a

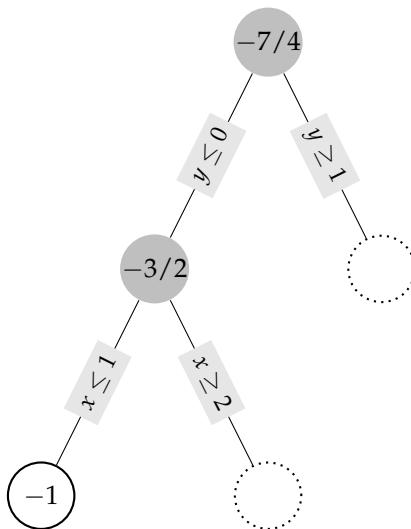
$$\begin{aligned} z &= -\frac{3}{2} + 2w + \frac{1}{2}t \\ y &= -w \\ x &= \frac{3}{2} - 2w - \frac{1}{2}t \\ s &= \frac{1}{2} + 4w + \frac{1}{2}t \\ u &= -\frac{1}{2} + 2w + \frac{1}{2}t \end{aligned}$$

e, con un passo di pivot, si ottiene

$$\begin{aligned} z &= -1 + u \\ y &= -w \\ x &= 1 - u \\ s &= 1 + 2w + u \\ t &= 1 + 2u - 4w \end{aligned}$$

15.4. Il metodo Branch & Bound

che è intera. Il valore -1 costituisce quindi un limite superiore valido per il problema originale:



A questo punto il nodo è completamente esplorato e può essere utilizzato per le operazioni di bounding. Anzi, utilizzando il criterio speciale per il caso in cui la funzione di costo ha coefficienti interi, si potrebbe già dedurre che questa è la soluzione ottimale del problema, in quanto il rilassamento del nodo radice ha valore $-7/4$ e $\lceil -7/4 \rceil = -1$.

Volendo invece proseguire con la strategia LIFO, senza usare questa opportunità di bounding, esclusivamente a scopo illustrativo, si può effettuare una risalita nell'esplorazione dell'albero (*backward search*).

track), passando ad esplorare il ramo destro:

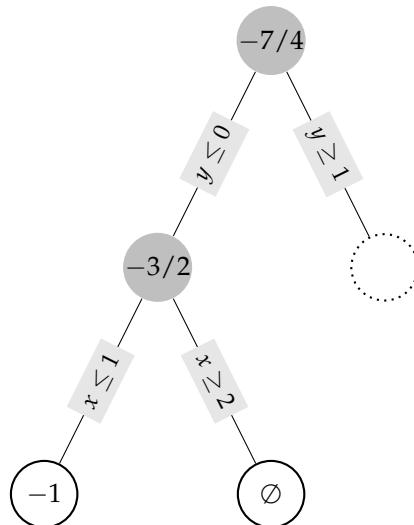
$$\begin{aligned}z &= -\frac{3}{2} + 2w + \frac{1}{2}t \\y &= -w \\x &= \frac{3}{2} - 2w - \frac{1}{2}t \\s &= \frac{1}{2} + 4w + \frac{1}{2}t \\x &\geq 2\end{aligned}$$

che si trasforma immediatamente in

$$\begin{aligned}z &= -\frac{3}{2} + 2w + \frac{1}{2}t \\y &= -w \\x &= \frac{3}{2} - 2w - \frac{1}{2}t \\s &= \frac{1}{2} + 4w + \frac{1}{2}t \\u &= -\frac{1}{2} - 2w - \frac{1}{2}t\end{aligned}$$

Questo problema non ammette soluzioni ammissibili, come si può vedere dall'ultima equazione del dizionario. Il nodo è quindi esplorato e si può ulteriormente risalire.

15.4. Il metodo Branch & Bound



Nella lista dei nodi da esplorare resta a questo punto il discendente “destro” della radice dell’albero:

$$z = -\frac{7}{4} + \frac{1}{2}s + \frac{1}{4}t$$

$$y = \frac{1}{8} + \frac{1}{8}t - \frac{1}{4}s$$

$$x = \frac{7}{4} - \frac{1}{2}s - \frac{1}{4}t$$

$$y \geq 1$$

15. INTRODUZIONE ALLA PROGRAMMAZIONE LINEARE INTERA

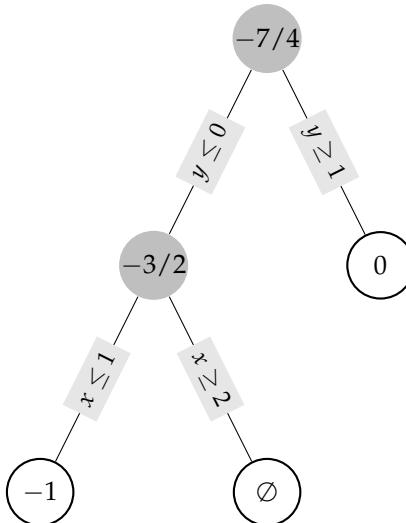
Eseguendo il simplex duale, si ottiene

$$\begin{aligned}z &= -\frac{7}{4} + \frac{1}{2}s + \frac{1}{4}t \\y &= \frac{1}{8} + \frac{1}{8}t - \frac{1}{4}s \\x &= \frac{7}{4} - \frac{1}{2}s - \frac{1}{4}t \\w &= -\frac{7}{8} + \frac{1}{8}t - \frac{1}{4}s\end{aligned}$$

e, con un'operazione di pivot, si giunge a

$$\begin{aligned}z &= s + 2w \\y &= 1 + w \\x &= -s - 2w \\t &= 7 + 8w + 2s\end{aligned}$$

Questa soluzione è intera e, pertanto, termina l'esplorazione di questo ramo e anche dell'intero albero di Branch & Bound.



Metodo Branch & Bound per problemi di ottimizzazione combinatoria

Per meglio illustrare il meccanismo del metodo Branch & Bound, si utilizzerà un esempio di *problema di zaino binario*. Il problema riguarda la scelta ottimale tra n oggetti, o attività o progetti da eseguire, per ciascuno dei quali è noto un “valore” c ed un “peso” a . Data la capacità massima b di un “contenitore”, si desidera scegliere la combinazione di oggetti il cui peso totale non supera la capacità ed il cui valore complessivo è massimo. Il problema si può formulare come *PLI* come segue:

$$\begin{aligned} \max c^T \delta \\ \sum_{j=1}^n a_j \delta_j \leq b \\ \delta \in \{0, 1\}^n \end{aligned}$$

15. INTRODUZIONE ALLA PROGRAMMAZIONE LINEARE INTERA

con $a_j > 0, c_j > 0 \quad \forall j$ e $b > 0$.

Vale la seguente importante proprietà:

Teorema 60. *Si consideri un problema di zaino binario.*

Se gli indici sono stati permutati in modo tale che

$$\frac{c_1}{a_1} \geq \frac{c_2}{a_2} \geq \cdots \geq \frac{c_n}{a_n}$$

allora una soluzione ottimale del rilassamento lineare si ottiene ponendo

$$\delta_j = 1$$

per $j = 1, 2, \dots, j^ - 1$, e*

$$\delta_{j^*} = \left(b - \sum_{j=1}^{j^*-1} a_j \right) / a_{j^*}$$

dove j^ è tale che*

$$\sum_{j=1}^{j^*-1} a_j \leq b$$

$$\sum_{j=1}^{j^*} a_j > b$$

La dimostrazione è lasciata come esercizio – può essere ottenuta, ad esempio, applicando la teoria della dualità per i problemi di *PL*.

Quindi per risolvere il rilassamento continuo di un problema di zaino 0 – 1 è sufficiente ordinare gli oggetti in modo decrescente rispetto al rapporto valore/peso, ed inserire nello zaino tutti gli oggetti che possono essere inseriti; del primo oggetto che non è possibile inserire nello zaino, si sceglie quella frazione che porta alla saturazione della capacità.

Conoscere la strategia per la risoluzione di questi rilassamenti lineari permetterà di risolvere il problema mediante l'algoritmo

15.4. Il metodo Branch & Bound

Branch & Bound, ma senza la necessità di utilizzare il metodo del simplesso.

Si noti che il problema è formulato come un problema di massimo – il criterio per la potatura dell'albero di Branch & Bound dovrà quindi essere ribaltato.

Esempio 48. Si consideri un problema di zaino di capacità massima pari a 17, nel quale i valori ed i pesi siano i seguenti:

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>tot</i>
<i>c</i>	2	3	2	1	3	5	4	
<i>a</i>	1	2	2	1	4	10	9	≤ 17
$\frac{c}{a}$	2	$\frac{3}{2}$	1	1	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{4}{9}$	

La soluzione del problema rilassato è

$$A = B = C = D = E = 1, F = \frac{7}{10}$$

con valore $2 + 3 + 2 + 1 + 3 + 5 \times \frac{7}{10} = 14.5$. Dal problema iniziale, si generano quindi due sottoproblemi, caratterizzati rispettivamente dall'aggiunta del vincolo $F \leq 0$ e $F \geq 1$. Si noti che in un problema con variabili binarie, l'operazione di branching equivale a fissare il valore di una variabile ad uno dei suoi due valori possibili. I due problemi generati sono dunque ancora problemi di zaino, con i seguenti dati:

$F = 0$	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>G</i>	<i>tot.</i>	
<i>c</i>	2	3	2	1	3	4		
<i>a</i>	1	2	2	1	4	9	≤ 17	

e

$F = 1$	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>G</i>	<i>tot.</i>	
<i>c</i>	2	3	2	1	3	4	(+5)	
<i>a</i>	1	2	2	1	4	9	≤ 7	

15. INTRODUZIONE ALLA PROGRAMMAZIONE LINEARE INTERA

Si immagini di adottare la strategia di risolvere per primo il problema in cui la variabile è assegnata al valore 1 e di esplorare l'albero “in profondità”, cioè scendendo fino ad una foglia prima di risalire (“backtrack”). La soluzione del rilassamento del problema (15.7) è

$$A = B = C = D = 1, E = \frac{1}{4}$$

con valore pari a $5 + 2 + 3 + 2 + 1 + 3 \times \frac{1}{4} = 13.75$. A questo punto vengono generati due sottoproblemi, corrispondenti ai seguenti zaini:

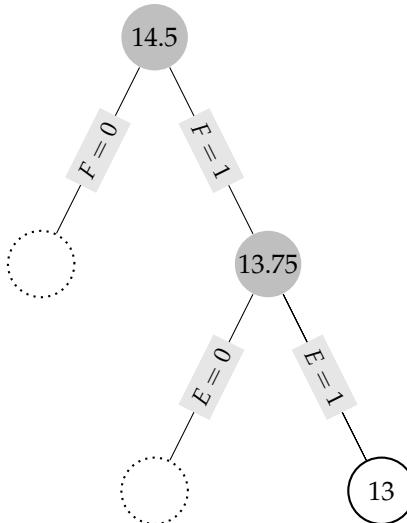
$$\begin{array}{l} E = 0, F = 1 \\ c \quad \quad \quad 2 \quad 3 \quad 2 \quad 1 \quad 4 \quad (+5) \\ a \quad \quad \quad 1 \quad 2 \quad 2 \quad 1 \quad 9 \quad \leq 7 \end{array} \quad (15.8)$$

$$\begin{array}{l} E = 1, F = 1 \\ c \quad \quad \quad 2 \quad 3 \quad 2 \quad 1 \quad 4 \quad (+8) \\ a \quad \quad \quad 1 \quad 2 \quad 2 \quad 1 \quad 9 \quad \leq 3 \end{array} \quad (15.9)$$

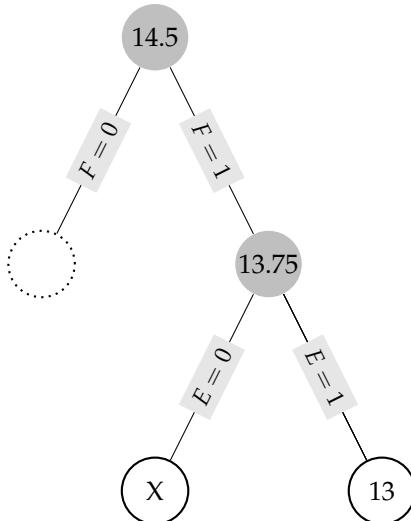
Proseguendo con la strategia di scendere in profondità sul ramo destro ($E = 1$), occorre risolvere il problema (15.9). Prima di farlo però si può notare che la variabile G in questo sottoproblema deve essere fissata a 0, avendo peso superiore alla capacità residua. In realtà il rilassamento lineare potrebbe essere risolto anche senza fissare il valore di G , ma ovviamente c’è convenienza dal punto di vista numerico a sfruttare queste conoscenze sul problema. Eliminando dunque la variabile G si ottiene la soluzione intera

$$A = B = 1$$

con valore $8 + 2 + 3 = 13$. Si è quindi giunti ad una foglia dell’albero di Branch & Bound, ed occorre risalire per completare l’esplorazione di tutto l’albero. Graficamente si può rappresentare la situazione come segue:



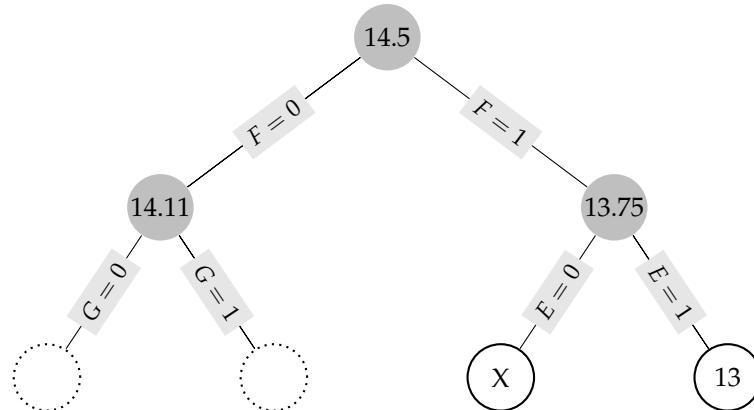
L'operazione di backtrack dovrebbe ora portare al nodo corrispondente al problema (15.8); tuttavia si può osservare che il rilassamento del nodo “genitore” di questo problema ha un valore pari a 13.75. Questo è un valore migliore di 13, il record corrente; tuttavia, essendo i valori associati agli oggetti dati da numeri interi, anche il valore dell'ottimo sarà intero. Quindi, tra i discendenti del nodo nessuno potrà avere valore migliore di 13. E' quindi inutile esplorare questo ramo dell'albero perché, nell'ipotesi migliore, potrebbe portare ad una soluzione alternativa, ma sempre di valore 13. Il nodo viene quindi considerato come implicitamente visitato:



Si può quindi risalire alla radice dell'albero; qui il valore, 14.5, non permette alcuna conclusione: potrebbe esistere una soluzione di valore 14 nel ramo sinistro dell'albero. In ogni caso si può già affermare che l'ottimo del problema ha valore pari a 13 oppure a 14 – a volte questa incertezza è ritenuta tollerabile e l'algoritmo può essere arrestato.

Iniziando con il rilassamento del problema (15.6) si ottiene la soluzione $A = B = C = D = E = 1, G = \frac{7}{9}$, di valore $2 + 3 + 2 + 1 + 3 + \frac{7}{9} \times 4 = 14.\bar{1}$.

15.4. Il metodo Branch & Bound



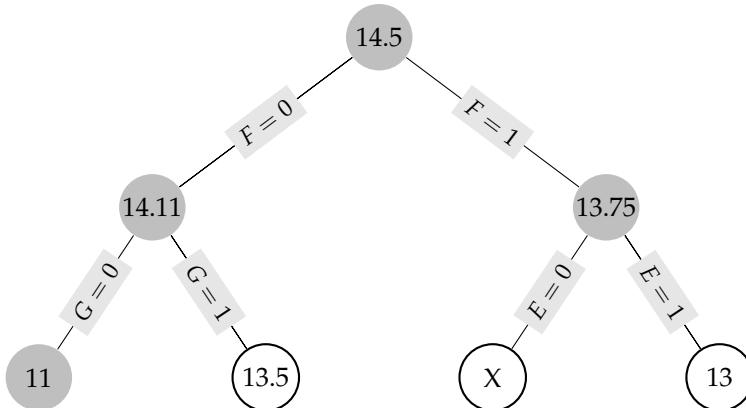
A questo punto è necessario effettuare una ramificazione su G : il problema corrispondente a $G = 1$ è il seguente:

$F = 0, G = 1$	A	B	C	D	E	
c	2	3	2	1	3	(+4)
a	1	2	2	1	4	≤ 8

il cui rilassamento ha ottimo $A = B = C = D = 1, E = \frac{1}{2}$, con valore $4 + 2 + 3 + 2 + 1 + 3 \times \frac{1}{2} = 13.5$. Di nuovo, questo ramo dell'albero può essere tagliato in quanto non sarà possibile produrre soluzioni con valore migliore di 13. Scendendo dall'altro ramo, il problema diviene

$F = 0, G = 0$	A	B	C	D	E	
c	2	3	2	1	3	
a	1	2	2	1	4	≤ 17

la cui soluzione è intera: $A = B = C = D = E = 1$, con valore $2 + 3 + 2 + 1 + 3 = 11$.



Pertanto si è dimostrata l'ottimalità della soluzione

$$A = B = E = F = 1.$$

Gli esempi appena presentati dovrebbero essere sufficienti per comprendere l'idea di fondo sulla quale si basa il metodo Branch & Bound. E' bene sottolineare come l'implementazione del metodo si presti a moltissime varianti che possono influire, anche in modo drastico, sul suo funzionamento. In particolare una buona implementazione del metodo deve prevedere:

1. Strategie di visita dell'albero: ad ogni iterazione del metodo, occorre scegliere e risolvere un problema associato all'insieme delle foglie "attive" dell'albero di Branch & Bound, cioè a quei nodi cui corrispondono rilassamenti né vuoti, né a soluzione intera, né con valore peggiore del valore della miglior soluzione intera. Si elencano di seguito alcune delle più comuni strategie di visita:
 - a) Visita in profondità, basata su una lista di priorità di tipo LIFO, nella quale viene risolto il problema associato al

nodo attivo di più recente inserzione nella lista. Questa strategia ha il vantaggio che, a volte, conduce relativamente in fretta a determinare una prima soluzione ammissibile intera; inoltre la gestione della memoria, per quanto riguarda i nodi ancora da esplorare, è più semplice ed efficiente che con altre strategie.

- b) Visita in larghezza, o priorità FIFO: ad ogni iterazione viene risolto il rilassamento lineare del nodo meno recente
 - c) Visita adattativa, basata sul valore del rilassamento: ad ogni iterazione viene scelto un nodo al quale è associato il problema rilassato con valore migliore (minimo limite inferiore): l'idea di questa strategia è quella di esplorare per primi i nodi che hanno maggiore probabilità di portarci alla soluzione ottima; in alternativa può invece essere scelto il nodo che ha il valore peggiore: in questo caso la strategia corrisponde al desiderio di arrivare velocemente a tagliare parti consistenti dell'albero.
2. Strategia di suddivisione: in presenza di più variabili sulle quali effettuare l'operazione di suddivisione di un problema, occorre introdurre un criterio di scelta.

Inoltre è bene osservare che lo schema può essere reso più generale, sostituendo alla suddivisione operata intorno al valore frazionario di una variabile una suddivisione differente, operata mediante separazione del poliedro corrispondente ad un rilassamento nell'unione di due (o più) poliedri.

Anche per quanto riguarda il rilassamento è bene osservare che non è necessario utilizzare sempre il rilassamento lineare: l'algoritmo richiede semplicemente la determinazione di un limite inferiore al valore dell'ottimo, che può essere ottenuta mediante rilassamenti di tipo più generale.

Bibliografia

- Ahuja, R. K., Magnanti, T. L., & Orlin, J. B. (1993). *Network flows: theory, algorithms and applications*. New Jersey: Prentice Hall.
- Balas, E., Ceria, S., & Cornuéjols, G. (1993). A lift-and-project cutting plane algorithm for mixed 0-1 programs. *Mathematical Programming*, 58, 295-324.
- Bazaraa, M. S., Jarvis, J. J., & Sherali, H. D. (1990). *Linear programming and network flows*. Toronto: John Wiley & Sons.
- Bertsekas, D. P. (1998). *Network optimization: continuous and discrete models*. Belmont, Massachussets: Athena Scientific.
- Bertsekas, D. P., & Gallager, R. G. (1992). *Data networks*. Prentice-Hall.
- Bertsimas, D., & Tsitsiklis, J. N. (1997). *Introduction to linear optimization*. Belmont, Mass. (USA): Athena Scientific.
- Borgwardt, K. H. (1982). The average number of pivot steps required by the simplex-method is polynomial. *Mathematical Methods of Operations Research*, 26, 157-177.
- Chvátal, V. (1983). *Linear programming*. New York: W.H. Freeman & Co.

BIBLIOGRAFIA

- Dantzig, G. B. (1951). Activity analysis of production and allocation. In T. Koopmans (Ed.), (pp. 339–347). Wiley & Chapman–Hall.
- Dantzig, G. B. (1990). The diet problem. *Interfaces*, 20(4), 43–47.
- Dantzig, G. B., & Thapa, M. N. (1997). *Linear programming 1: Introduction*. New York: Springer-Verlag.
- Dantzig, G. B., & Thapa, M. N. (2003). *Linear programming 2: Theory and extensions*. New York: Springer-Verlag.
- Dax, A. (1997). An elementary proof of Farkas' lemma. *SIAM Review*, 39(3), 503–507.
- Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, 269–271.
- Farkas, J. (1902). Über die theorie der einfachen ungleichungen. *Reine Angew. Math.*, 124, 1–24.
- Ford, L. R., & Fulkerson, D. R. (1956). Maximal flow through a network. *Canadian Journal of Mathematics*, 8, 399–404.
- Fourer, R., Gay, D. M., & Kernighan, B. W. (2002). *AMPL: A modeling language for mathematical programming* (Second ed.). Brooks/Cole Publishing Company / Cengage Learning.
- Fredman, M. L., & Tarjan, R. E. (1987). Fibonacci heaps and their uses in improved network optimization algorithms. *Journal of the Association for Computing Machinery*, 596–615.
- Garey, M. R., & Johnson, D. S. (1979). *Computers and intractability: A guide to the theory of NP-completeness*. Freeman.
- Garille, S. G., & Gass, S. I. (2001). Stigler's diet problem revisited. *Operations Research*, 49(1), 1–13.
- Goldberg, A. W., Tardos, E., & Tarjan, R. L. (1990). Network flow algorithms. In B. Korte, L. Lovász, H. Prömel, & A. Schrijver (Eds.), *Algorithms and combinatorics* (pp. 101–164). Springer-Verlag.

Bibliografia

- Gomory, R. E. (2010). Outline of an algorithm for integer solutions to linear programs and an algorithm for the mixed integer problem. In T. Jünger et al. (Eds.), *50 years of integer programming 1958–2008* (pp. 77–103). Springer-Verlag.
- Kantorovich, L. V. (1940). A new method of solving some classes of extremal problems. *Doklady Akad Sci USSR*, 211–214.
- Karmarkar, N. (1984). A new polynomial time algorithm for linear programming. *Combinatorica*, 4, 373–395.
- Kim, E., & Santos, F. (2010). An update on the hirsch conjecture. *Jahresbericht der deutschen Mathematiker-Vereinigung*, 112, 73–98.
- Klee, V., & Minty, G. J. (1972). How good is the simplex algorithm? In A. Press (Ed.), *Inequalities III* (pp. 159–175).
- Luenberger, D. G. (2003). *Linear and nonlinear programming* (Second ed.). Kluwer Academic Publishers.
- Morgenstern, O., & Neumann, J. von. (1944). *Theory of games and economic behavior*. Princeton: Princeton University Press.
- Nemhauser, G. L., & Wolsey, L. A. (1988). *Integer and combinatorial optimization*. New York: J. Wiley & Sons.
- Padberg, M. (1999). *Linear optimization and extensions*. Berlino: Springer Verlag.
- Papadimitriou, C. H., & Steiglitz, K. (1998). *Combinatorial optimization: Algorithms and complexity*. Courier Dover Publications.
- Schoen, F. (2006). *Modelli di ottimizzazione per le decisioni*. Bologna: Società Editrice Esculapio.
- Schrijver, A. (1998). *Theory of linear and integer programming*. John Wiley & Sons.
- Stigler, G. (1945). The cost of subsistence. *J. Farm Econom.*, 25, 303–314.

BIBLIOGRAFIA

- Vanderbei, R. J. (2001). *Linear programming: Foundations and extensions* (2nd ed.).
- Williams, H. (1999). *Model building in mathematical programming*. J. Wiley & Sons.
- Ye, Y. (1997). *Interior point algorithms: Theory and analysis*. J. Wiley & Sons.

Indice analitico

- abbinamento, 391
- aggiogata, 388
- aggiunta , 388
- albero, 260
 - albero di supporto, 261
 - albero ricoprente, 261
- algoritmo di Dijkstra, 311
 - correttezza, 313
- analisi di sensitività, 215
- analisi di post–ottimalità, 215
- arco, 251
 - entrante nella sezione, 362
 - incidente, 255
 - multiplo, 253
 - tagliato “all’indietro”, 362
 - tagliato “in avanti”, 362
 - uscente dalla sezione, 362
- arg, 22
- assegnamento bipartito, 391
- backtrack, 414
- base, 49, 52
- base degenere, 53
- bilancio, 266
- Bland, regola di, 103
- bounding, 407
- cammino, 255
 - semplice, 256
- cammino incrementabile, 341
- cammino incrementante, 341
- capacità massima, 266
- capacità di una sezione, 362
- capacità residua, 340
- cappio, 253
- cardine, 94
- catena, 256
- catena semplice, 256
- ciclo, 256
- circolazione, 335
- circuito, 256
- coefficienti di costo ridotto, 86
- complementarità stretta, 177
- componente connessa, 260

INDICE ANALITICO

- congettura di Hirsch, 142
connesso, 257
convesso, 68
correttezza dell'algoritmo di Dijkstra, 313
curve di livello, 77
destinazione, 309
diametro di un poliedro, 143
digrafo, 253
Dijkstra, algoritmo di, 311
dimensione, 66
dimensione di un insieme, 67
dimensione di un sottospazio affine, 67
direzione ammissibile in , 60
diseguazione valida, 69, 394
diseguazioni lineari, 24
distanza fra due vertici, 143
dizionario, 86
duale del problema della dieta, 184
duale del problema di miscelezione, 188
duale lagrangiano, 153
duale standard, 153
dualità
interpretazione economica, 184
dualità debole, teorema di, 165
dualità forte, teorema di, 167
due fasi, metodo, 121
eccedenza, 31
equazioni lineari, 24
equivalenti, 37, 39, 40
equivaleanza fra problemi di ottimizzazione, 37
estremo, 71
etichetta definitiva, 311
etichettatura, 311
faccetta, 71
faccia, 70
faccia massimale, 71
flusso attraverso una sezione, 363
Ford e Fulkerson, 333
forma standard, 26, 27
formulazione forte, 386
fortemente ammissibile, 300
fortemente connesso, 257
funzione affine, 23
funzione di costo, 24
funzione lineare, 23
funzione obiettivo, 21, 24
gap di dualità, 166
grado, 255
grado entrante, 255
grado uscente, 255
grafo, 251
connesso, 257
grafo orientato, 253
grafo semplice, 253
illimitato, 96
problema, 22

- incremento di flusso, 341
inammissibile, 21
incognita libera, 33
incognite, 24
insieme ammissibile, 21
interpretazione economica della dualità, 184
intervallo di dualità, 166
iperpiano, 67
iperpiano di supporto, 70

Kirchoff, legge di, 266

lagrangiano, rilassamento, 149
lato, 251
legge di conservazione del flusso, 266
legge di Kirchoff, 266
Lemma di Farkas, 172
lista di adiacenza, 258

massimo flusso, 334
matrice di adiacenza, 258
matrice di incidenza, 259
matrice unimodulare, 388
metodo della M grande, 128
metodo due fasi, 121
Minkowski-Weil, teorema di, 132
MIP, 376

nodi adiacenti, 255
nodi intermedi, 312
nodo, 251

normale, 67
origine, 309
ottimizzazione
 problema di, 21
ottimo finito, 22
ottimo globale, 22

parti frazionarie, 399
payoff, 197
percorso di costo minimo, 309
pivot, 94
pl
 soluzione di base, 53
 base, 52
 soluzione, 49
 soluzione ammissibile, 50
 soluzione di base degenera, 53
 soluzione ottimale, 50
PLI, 376
poliedro, 68
politopo, 68
potenziali, 280
primale, 159
problema del flusso di costo minimo, 266
problema di PL illimitato, 96
problema di ottimizzazione, 21
problema di PL
 forma standard, 26
problema di zaino binario, 417
problema non ammissibile, 21
problema prima fase, 121

INDICE ANALITICO

- problemi equivalenti, 37
problemi equivalenti di massimizzazione, 39
problemi minimax, 46
programmazione lineare, 23
programmazione lineare intera, 376
programmazione mista intera, 376
punto interno, metodi, 148
radice, 284
regola di Bland, 103
rete di flusso, 265
rete residua, 335, 340
rilassamento, 149
rilassamento continuo, 379
rilassamento lagrangiano, 149, 151
rilassamento lineare, 379
scarti complementari, teorema degli, 173
scarto, 31
semispazi affini, 67
semplici, 256
sensitività, analisi di, 215
sezione, 362
sezione $s-t$, 362
sezione minima, 363
simplesso, 83
simplesso su reti, 277
slack, 31
soluzione, 49
soluzione ammissibile, 22, 50
soluzione di base, 49, 53
soluzione di base degenere, 53
soluzione duale complementare, 178
soluzione ottimale, 50
sorgente, 309
sottogradiente, 219
sottografo, 260
sottospazio, 65
sottospazio affine, 66
spazio, 65
spigolo, 71
standard, 26
strategia, 197
strategia deterministica, 198
strategie pure, 199
surplus, 31
taglio, 237, 395
taglio $s-t$, 362
taglio di Gomory frazionario, 399
tensione, 280
teorema degli scarti complementari, 173
teorema di dualità debole, 165
teorema di dualità forte, 167
teorema fondamentale della programmazione lineare, 53
teorema massimo flusso/minima sezione, 366

teoria dei giochi, 197
teoria della complessità computazionale, 378
terminale, 309
totalmente unimodulare, 389

unimodulare, 388

valida, 394
valore, 334
variabili artificiali, 121
variabili di base, 53
variabili di decisione, 24
vertice, 71, 251
vincoli di diseguaglianza, 24
vincoli di eguaglianza, 24
vincoli di non negatività, 24
vincoli di segno, 24
vincolo attivo, 72

Cover design: © 2012 Carlotta Schoen
Typesetting: L^AT_EX using the *memoir* class

Questo è il primo di una serie di volumi didattici
sulla teoria, gli algoritmi, i modelli e le applicazioni della
Ricerca Operativa.

Il testo è adatto per un corso al triennio nelle Facoltà di
Ingegneria, Scienze, Economia e contiene un'introduzione
alla programmazione lineare, alla programmazione intera
ed agli algoritmi elementari per l'ottimizzazione su grafi.
Si basa su appunti in uso da anni presso i Corsi di Laurea
in Ingegneria Informatica ed Ingegneria Gestionale.

Fondamenti di Ricerca Operativa

Fabio Schoen

fabio.schoen@unifi.it

gol.dsi.unifi.it

Facoltà di Ingegneria
Università degli Studi di Firenze

ID:12502182
www.lulu.com

ISBN 978-1-105-49496-3



90000

9 781105 494963