Francesco Mason

METODI MATEMATICI PER LA GESTIONE DELLE AZIENDE

Parte I Programmazione lineare e grafi

Questo testo è nato dal proposito di sviluppare alcuni temi classici di Ricerca Operativa, riorganizzandoli sulla falsariga di alcuni problemi concreti che si pongono ad aziende private e pubbliche che devono gestire operazioni di trasporto, di cose o persone, in situazioni di scarsità di risorse. Questo filo conduttore è stato sperimentato in un corso universitario di primo livello ed i problemi presi come riferimento sono quelli del trasporto di persone (aziende municipalizzate di trasporto urbano ed extraurbano), della gestione di servizi di igiene urbana (tipicamente, aziende che effettuano l'asporto dei rifiuti) e di aziende di produzione che, nell'ambito della funzione di logistica distributiva, devono prendere decisioni quotidiane su come e quando consegnare le merci o in ogni modo visitare la clientela.

I temi oggetto di analisi derivano dall'esperienza maturata, sia a livello di docenti sia di laureandi, in occasione della preparazione di dissertazioni di laurea e in progetti di ricerca a livello nazionale, nell'ambito del Dipartimento di Matematica Applicata dell'Università di Venezia e del Dipartimento di Elettronica, Elettrotecnica e Informatica dell'Università di Trieste. I docenti coinvolti sono titolari di corsi di Ricerca Operativa e ricercatori universitari del corrispondente settore disciplinare; i laureandi provengono dalle Università di Venezia e Trieste, in larga parte, ma si è avuta anche qualche presenza di studenti delle Università di Padova e di Udine.

L'obiettivo del testo, piuttosto che essere quello di effettuare una rassegna, più o meno completa, dei temi di Ricerca Operativa, è quello di illustrare un modo di lavorare, o, se si vuole, una filosofia che è importante cogliere nella disciplina.

In realtà, il condurre fino in fondo un'operazione del genere (e in altre parole, ancorare ad alcuni problemi specifici grappoli di modelli), oltre che presentare delle difficoltà tecniche, non è del tutto giustificato: alcune nozioni di base, in Ricerca Operativa come in qualsiasi altra disciplina, vanno necessariamente premesse, senza doversi preoccupare più di tanto di collegarle alla risoluzione di un problema specifico.

Pertanto si può suddividere il contenuto del libro in due parti: una preliminare (riguardante elementi di Programmazione Matematica, in particolare di Programmazione Lineare, di Teoria dei Grafi e di Complessità Computazionale) ed una seconda, che si potrebbe definire monografica, dove i temi di Ottimizzazione Combinatoria (Problemi di Cammino, Flusso, Localizzazione, Assegnazione ecc.) trovano una sistemazione non del tutto usuale, ma con una maggiore finalizzazione rispetto a quello cui si è abituati nei manuali più noti.

Per consentire un utilizzo anche in corsi di durata molto limitata, alcune sezioni, non essenziali per un discorso organico, sono sviluppate in carattere più piccolo: la loro eliminazione non toglie coerenza ai temi restanti.

Il materiale può essere utilizzato in più di un corso e con un differente livello di approfondimento: sta al docente fare le sue scelte!

1. ELEMENTI INTRODUTTIVI.

1.1 La filosofia della Ricerca Operativa.

Tre termini ricorrenti in Ricerca Operativa sono: **problema**, **modello**, **algoritmo**. Darne una definizione univoca non è facile, anche perché in qualche misura 'sfumano' uno nell'altro, ma a volte sono usati in maniera certamente impropria, scambiandoli tra loro, mentre non è bene confonderli. Si può pensare ad un **problema** come ad una circostanza nella quale un decisore, ad es. un responsabile di un settore aziendale, percepisce di dover fare delle scelte: come agire in un determinato contesto, o, più a monte, quale impostazione dare ad una certa struttura. Egli intuisce che vi possono essere varie possibilità, alcune delle quali magari non ancora evidenziate completamente. Un problema può essere espresso con quesiti del tipo: come meglio organizzarsi? Come condurre, al meglio, una certa operazione? Come si comporta un certo sistema in determinate circostanze?

Un problema sorge sia quando si tratta di **decidere o costruire una politica** da seguire, sia quando si cerca (solamente) di **capire il comportamento** di un sistema o di più sistemi alternativi.

Per esemplificare, può trattarsi di individuare come va effettuata al minimo costo (e con la massima soddisfazione della clientela) una spedizione di merce in una fissata rete di comunicazioni (problema a livello tattico o operativo), ma anche di stabilire dove è il caso di localizzare un nuovo impianto e se è il caso di servirsi di magazzini intermedi (problemi che, in genere, hanno valenza strategica).

Può trattarsi solo (ma spesso è proprio la prima cosa da fare!) di approfondire lo studio del funzionamento di un settore di un'azienda, (ad es., un reparto produttivo) reale o ipotetico, per avere un'idea poi di come potrebbero andare le cose se si adottassero certe scelte o si imponessero altre condizioni.

Quando si comincia a delineare la natura del problema e, allo scopo di capire chi sono gli elementi da considerare, si individuano le variabili in gioco, i legami tra loro, quale è - se vi è - l'obiettivo (o gli obiettivi) da raggiungere e quali sono i vincoli da rispettare, si va verso la costruzione di un **modello**.

Tipicamente, nel campo della Ricerca Operativa, si costruisce un modello matematico, che, nella sua veste più completa, con uno o più obiettivi specificati, acquista l'aspetto di un **modello di programmazione**. In altri casi, limitandosi alla comprensione di un sistema, si hanno dei modelli che potremmo definire di **simulazione**.

Ogni modello è una fotografia della realtà, ma non coincide con la realtà stessa: in genere la semplifica. La semplificazione deve essere un compromesso ragionevole tra l'aderenza al caso concreto (più si vuole essere vicini alla realtà,

meno si può semplificare) e l'uso di strumenti governabili (per evitare grossi ostacoli di natura matematica, occorre sfrondare il modello da variabili di secondaria importanza, riducendo la complessità del modello stesso).

Il modello va 'risolto': se è un modello di programmazione, si deve trovare una **soluzione** o una **politica ottima** (oppure un insieme di soluzioni alternative, possibilmente in numero contenuto, tra cui scegliere), mentre per un modello di simulazione si deve ottenere un insieme di valori per le variabili che descrivono il comportamento del sistema. La costruzione di una soluzione è opera di un **algoritmo**, cioè di una procedura che, utilizzando i dati in input e le relazioni espresse nel modello, costruisce un output, sia esso una decisione o una raffigurazione del sistema.

Nella fase di costruzione del modello è utile (e può diventare anche il vero scopo di una ricerca) analizzare e rendere riproducibili le procedure decisionali già usate in azienda: questo si può fare tutte le volte che si cerca di migliorare una performance, ma si lavora nell'ambito di un problema ben noto e quindi già affrontato e in qualche modo risolto, seppure approssimativamente.

La pura e semplice ricostruzione di una tecnica risolutiva, con lo scopo di giungere agli stessi risultati aziendali, è molto interessante per due motivi: fa prendere coscienza dei vincoli specifici della azienda, e quindi mette in condizioni di costruire un modello più aderente alla realtà, e consente generalmente di ridurre i tempi aziendali di costruzione delle soluzioni (che sovente sono ottenute con procedimenti manuali) meccanizzando la procedura.

Ad esempio, la ricostruzione della tecnica di formazione dei turni uomo in un'azienda di trasporto, così come la conduce l'operatore, permette in genere di ridurre i tempi di compilazione degli orari del personale, da qualche settimana a meno di un'ora: questo dà anche la possibilità concreta di effettuare in tempi accettabili delle simulazioni che permettono di capire le conseguenze che si potrebbero avere se cambiano alcune delle regole di formazione dei turni stessi (spesso, queste sono oggetto di contrattazione sindacale e non è facile capire per l'azienda il peso reale delle innovazioni, proprio perché rifare i turni richiede troppo tempo). Naturalmente occorre una stretta collaborazione tra lo studioso di Ricerca Operativa e la persona che, all'atto della costruzione della soluzione dell'azienda, si renda disponibile per spiegare i motivi e la logica del suo procedere.

Si diceva all'inizio che alcuni termini 'sfumano' uno nell'altro: in effetti, a volte non c'è un confine preciso tra problema e modello perché formulando il problema si precisano già tipologie di relazioni, mentre a sua volta si formula un modello in un certo modo già pensando alla tecnica in grado di fornire una soluzione, per cui un certo modello è già, in qualche misura, anche un algoritmo. Si coglie meglio la differenza tra i vari concetti, almeno come saranno intesi in questa esposizione, quando uno stesso problema conduce a modelli diversi e a sua volta uno stesso modello può essere affrontato (risolto) con differenti algoritmi.

1.2 Modelli normativi e modelli descrittivi.

I modelli della Ricerca Operativa sono distinguibili in due categorie: modelli **normativi** e modelli **descrittivi**. Nei primi si individua una soluzione che può essere, a seconda delle circostanze, la migliore in assoluto o la migliore che si

riesce a costruire in tempi accettabili; nei modelli descrittivi si cerca invece di avere una riproduzione con formule del sistema che si studia per capirne meglio il comportamento, ma senza pretendere di individuare nell'ambito del modello stesso soluzioni che migliorino la prestazione.

Alla luce di queste definizioni, per esemplificare, la ricostruzione della tecnica di formazione dei turni di cui si è detto sopra conduce ad un modello descrittivo, mentre quando si cerca - con un altro modello - di ridurre al minimo il personale necessario per la copertura delle corse, si entra in una logica di tipo normativo.

Beninteso, 'normativo' è un aggettivo di comodo da usare con cautela: esso non implica che la soluzione individuata sia quella da usare a tutti i costi, poiché si tratta pur sempre di una soluzione nata da un modello che semplifica la realtà. Tale soluzione va sottoposta agli esperti e ai decisori dell'azienda interessata e devono essere questi a esprimerne il grado di rispetto dei vincoli. Occorre ricordare infatti che spesso i vincoli sono individuati in interviste con l'azienda un po' alla volta, perché molti di essi non sono percepiti esplicitamente, all'interno della azienda stessa, come tali.

In sintesi, un modello normativo presenta due elementi caratteristici:

- una (o almeno una) **funzione oggetto**, da rendere massima o minima (si osservi che è molto frequente, se non la regola, il caso in cui un'azienda persegue contemporaneamente più obiettivi, anche contrastanti tra di loro, per cui la presenza di una sola funzione oggetto è spesso, ancora una volta, un elemento di semplificazione);
- un **insieme di vincoli**, cioè condizioni che devono essere rispettate da qualsiasi soluzione si voglia mettere in atto.

Un modello di questo tipo è un modello di programmazione matematica.

1.3 Modelli di programmazione matematica.

In un modello di programmazione matematica, figurano, come si è appena detto, (almeno) una funzione obiettivo, o funzione oggetto, ed un insieme di vincoli, cioè relazioni che devono essere soddisfatte tra le variabili.

La **funzione oggetto** (assumendo, almeno per ora, che sia unica) consiste in un'espressione il cui significato può essere interpretato come quello di un **costo** - nel qual caso si cercherà di ridurne il valore quanto più possibile - oppure di un **guadagno -** caso in cui viceversa si cercherà di ottenere il valore massimo.

In un modello di Programmazione Matematica la funzione oggetto è data premettendo il tipo di obiettivo, di massimo o di minimo, come segue:

$$\max \quad 3x + 2y - xy,$$

$$\min \quad xyw + 5x - \tfrac{1}{2} w.$$

Il valore assunto dalla funzione oggetto sarà indicato con z.

I **vincoli** possono essere sotto forma di **equazioni o disequazioni**. In ogni caso, le variabili che compaiono, sia nella funzione oggetto sia nei vincoli, sono

definite **variabili decisionali** e possono essere pensate come le componenti di un vettore. Questo vettore, se soddisfa tutti i vincoli del problema, è detto **soluzione ammissibile**. L'insieme delle soluzioni ammissibili costituisce la **regione ammissibile**, **RA**.

Un vincolo può assumere l'aspetto:

$$x + 6y - w \geq 15,$$

oppure

$$3x - 6y + \log w \leq 1,$$

oppure ancora

$$2x + 3y - xyw = 10.$$

Si preferisce evitare la presenza di vincoli di diseguaglianza in senso stretto del tipo '>' oppure '<', anziché '≥' o '≤', perché è preferibile che la regione ammissibile sia un insieme chiuso, come si dirà tra poco.

Tra i vincoli figurano spesso **condizioni di non negatività** per le variabili decisionali (o per le componenti di un vettore soluzione). Queste condizioni di solito sono trattate a parte. Un vettore che soddisfi tutti i vincoli tranne quelli di non negatività è detto semplicemente **soluzione**.

Tra le soluzioni ammissibili occorre individuare la migliore, e cioè quella per cui il costo è minimo o il guadagno è massimo, soluzione che sarà detta **soluzione ottima**: la sua esistenza non è sempre garantita, ma è certamente assicurata se la funzione oggetto è continua e la regione ammissibile è un insieme chiuso e limitato, da cui la preferenza per i vincoli espressi in senso debole**.**

I problemi di programmazione matematica sono classificati in genere sulla base delle particolari forme che assumono funzione oggetto e vincoli. Queste particolarità, da un lato sono di natura matematica, ma al tempo stesso caratterizzano le interrelazioni tra le variabili decisionali.

Il caso più studiato - e sotto certi aspetti il più semplice - è quello della **programmazione lineare** (brevemente **PL**), nella quale sia funzione oggetto che vincoli sono lineari, cioè contengono solo polinomi di grado 1 nelle variabili decisionali. Le variabili sono poi di tipo continuo e quindi il loro campo di variazione è dato da intervalli chiusi (eventualmente semirette se non addirittura rette).

Come affermato in precedenza, va ribadito che la linearità presenta vari risvolti, alcuni di natura matematica, altri di natura logica.

Dal punto di vista matematico si ha indubbiamente una forma di problemi relativamente semplice, con la possibilità di utilizzare tecniche ampiamente studiate e collaudate di risoluzione, che consentono di trattare casi con migliaia di vincoli e decine di migliaia di variabili, giungendo alla soluzione esatta.

Dal punto di vista della natura logica dei vincoli e dell'obiettivo, la linearità va accuratamente vagliata in sede di costruzione di un modello, perché implica ipotesi molto forti sul fenomeno che si sta studiando.

La linearità porta ad escludere **economie e diseconomie** di scala, così come impedisce effetti di **sinergia** e/o di **dispersione**.

Ogni variabile è moltiplicata per un coefficiente di proporzionalità, fisso al variare della grandezza della variabile stessa. Non vi è possibilità quindi di tenere conto di fattori di scala (ad esempio, sconti su quantità). D'altra parte, il risultato in termini di guadagno, costo o consumo di risorsa per due o più attività congiunte è sempre la somma dei singoli effetti (di guadagno, costo o consumo, rispettivamente) di ciascuna attività. La copresenza di più attività, che spesso nella realtà produce effetti più che additivi o meno che additivi su costi o guadagni, è qui del tutto neutrale.

E' quindi evidente come la linearità sia adeguata solo in certi casi o solo con una qualche approssimazione: in genere è valida quando i valori delle variabili di decisione sono piuttosto elevati.

Se la funzione oggetto e/o i vincoli non sono lineari - come pure quando le variabili decisionali non sono continue, ma possono assumere solo valori discreti - si entra nel campo della **programmazione non lineare**. Il caso che qui interesserà maggiormente è quello in cui le variabili decisionali possono avere solo valori interi o addirittura solo uno dei due valori: 0 oppure 1 (variabili **binarie** o **booleane**).

Si tratta di una situazione assai frequente: si producono unità di prodotti (e allora il modello più adeguato è la **programmazione intera**); si decide se fare, cioè 1, o non fare, cioè 0, una certa cosa e allora si tratta di usare la **programmazione booleana**. Ma è anche vero che quando si ragiona per quantità piuttosto elevate, le diseconomie e le economie di scala tendono a scomparire e d'altra parte adottando un'opportuna unità di misura (ad es., il migliaio di pezzi), considerare continua una variabile più propriamente discreta arreca poco disturbo.

2. LA PROGRAMMAZIONE LINEARE.

2.1 La Programmazione Lineare ed il metodo del simplesso.

2.1.1 Generalità: formulazioni equivalenti di programmi lineari.

Come si è già detto, in un modello di programmazione lineare (brevemente **PL**), sia nella funzione obiettivo sia nei vincoli, figurano solo polinomi di primo grado nelle variabili.

Ad esempio, un problema di PL nelle variabili x, y, w è il seguente:

$$\max \quad 2x - 5y + 4w$$

con i vincoli:

$$\begin{aligned} x - 3y + 8w &= 10 \\ -4x + 6y - 10w &= 100 \end{aligned}$$

$$x, y, w \geq 0.$$

I problemi di Programmazione Lineare possono assumere differenti forme presentando le seguenti varianti:

- la funzione oggetto può essere da **massimizzare** o da **minimizzare**;

- alcuni vincoli possono essere sotto forma di **eguaglianza**, mentre altri possono presentarsi come **diseguaglianza**;
- le variabili sono generalmente vincolate alla **non negatività**, ma non mancano casi di variabili senza requisiti di segno e quindi potenzialmente sia positive sia negative o nulle: si parla allora di variabili **libere**. A volte, soprattutto per esigenze di tipo formale, si introducono variabili **non positive**.

Si può agevolmente passare da un problema con funzione oggetto e vincoli di una certa categoria ad un altro equivalente, con funzione oggetto e vincoli di tipo diverso:

- ogni problema di massimo si può trasformare in un problema di minimo e viceversa, cambiando di segno la funzione oggetto; con questo non muta la soluzione ottima del problema, mentre cambia di segno il valore ottimo dell'obiettivo; ad esempio, se la funzione oggetto è:

$$\max \quad 3x - 5y + 8w$$

si può trasformare la stessa in

$$\min \quad -3x + 5y - 8w;$$

- ogni vincolo di diseguaglianza si può trasformare in vincolo di eguaglianza, con l'introduzione di **variabili ausiliarie** mentre un vincolo di eguaglianza è equivalente ad una coppia di vincoli di diseguaglianza.

Ad esempio, il vincolo

$$2x + 7y \leq 12$$

può essere trasformato in

$$2x + 7y + s = 12, \qquad s \geq 0,$$

introducendo una variabile non negativa s, che è detta **variabile scarto** o **variabile slack**; invece il vincolo

$$-3x + 5y \geq 9$$

può essere trasformato in

$$-3x + 5y - s = 9, \qquad s \geq 0,$$

introducendo una variabile, anche stavolta non negativa, s, che in questa circostanza è detta **variabile surplus**; infine, sempre per esemplificare, il vincolo

$$1/2\, x + 3\, y = 10$$

equivale alla coppia di vincoli:

$$1/2\ x + 3\ y \ \geq\ 10$$

e

$$1/2\ x + 3\ y \ \leq\ 10;$$

- un problema con variabili libere è equivalente ad uno con variabili non negative, ottenuto sostituendo ogni variabile libera con la differenza tra due variabili ≥ 0. Infatti, qualsiasi quantità, positiva o negativa che sia, può sempre essere pensata come la differenza tra due quantità non negative. Pertanto, per dare un esempio, il problema

$$\max\ 2x - 5y + 3w$$

con i vincoli

$$x - y + \tfrac{1}{2}\, w = 5,$$
$$-x + 4y - 6w = 1,$$
$$x,\ y \geq 0,\ w \text{ libera}$$

è equivalente al problema:

$$\max\ 2x - 5y + 3u - 3v$$

con i vincoli

$$x - y + \tfrac{1}{2}\, u - \tfrac{1}{2}\, v = 5,$$
$$-x + 4y - 6u + 6v = 1,$$
$$x,\ y,\ u,\ v \geq 0,$$

effettuando la sostituzione

$$w = u - v.$$

Alternativamente, sempre in presenza di variabili libere, le stesse possono essere eliminate, ricavandole da un vincolo e sostituendo l'espressione ottenuta sia negli altri vincoli sia nella funzione oggetto. Continuando a fare riferimento all'ultimo esempio, si può ricavare w dal primo vincolo, come segue:

$$w = 10 - 2x + 2y$$

e sostituire nell'altro vincolo e nella funzione oggetto, per cui il problema diventa:

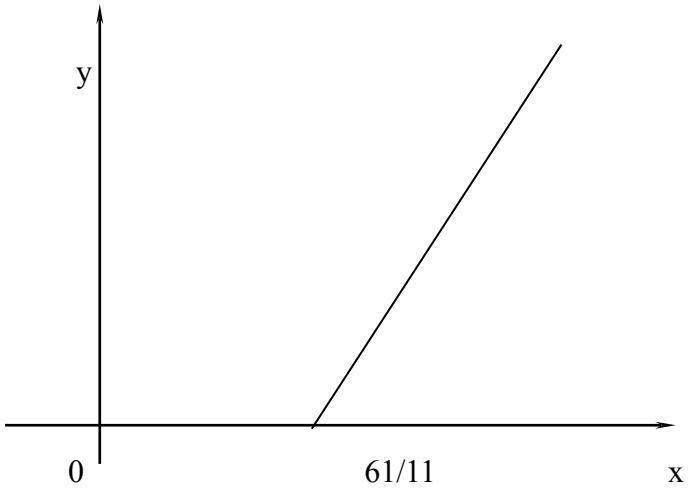
$$\max\ -4x + y + 30$$

soggetto a

$$11x - 8y = 61,$$
$$x,\ y \geq 0.$$

Una volta risolto questo problema (si può vedere per via grafica che la soluzione ottima è il punto di coordinate $x^* = 61/11$, $y^* = 0$ mentre il valore

ottimo della funzione oggetto è z* = -244/11 + 30 = 86/11), si può ricavare il valore ottimo di w dalla sua espressione, e cioè: w* = 10 - 122/11 = -12/11.



Prendendo ancora spunto da quest'ultimo esempio, si può osservare come una costante, nella funzione oggetto, non abbia riflessi sulle tecniche risolutive di un PL: in fase di soluzione la costante stessa può essere ignorata e la si può reintrodurre quando si deve indicare il valore ottimo della funzione oggetto stessa. Nell'esempio, si può osservare come il valore della funzione oggetto (che risulta 86/11, in corrispondenza della soluzione ottima x* = 61/11, y* = 0, w* = -12/11 nella formulazione originaria) coincida con il valore della espressione -4x + y + 30 in corrispondenza della soluzione x* = -61/11, y* = 0 del problema equivalente.

2.1.2 La forma standard e le forme canoniche.

Alcune forme di problemi di programmazione lineare hanno una propria denominazione. Adotteremo in questa esposizione le seguenti convenzioni:

a) con funzione oggetto di massimo e vincoli tutti di eguaglianza (oltre alle condizioni di non negatività) si ha la forma **standard**;

b) con funzione oggetto di massimo e vincoli di tipo ≤ oppure funzione oggetto di minimo e vincoli di tipo ≥ (e ancora con le condizioni di non negatività) si hanno le forme **canoniche**.

In simboli, un problema in forma standard si può scrivere:

$$\max \; c_1 x_1 + c_2 x_2 + \ldots + c_n x_n$$

con i vincoli (subject to)

$$a_{11} x_1 + a_{12} x_2 + \ldots + a_{1n} x_n = b_1,$$
$$a_{21} x_1 + a_{22} x_2 + \ldots + a_{2n} x_n = b_2,$$
$$\ldots\ldots\ldots\ldots$$
$$a_{m1} x_1 + a_{m2} x_2 + \ldots + a_{mn} x_n = b_m,$$

$$x_1, x_2, ..., x_n \geq 0.$$

I problemi in forma canonica presentano la seguente formulazione generale:

$$\max \; c_1 x_1 + c_2 x_2 + ... + c_n x_n$$

con i vincoli (s. t.)

$$\begin{aligned} a_{11} x_1 + a_{12} x_2 + ... + a_{1n} x_n &\leq b_1 \\ a_{21} x_1 + a_{22} x_2 + ... + a_{2n} x_n &\leq b_2 \\ &. . . . . \\ a_{m1} x_1 + a_{m2} x_2 + ... + a_{mn} x_n &\leq b_m, \end{aligned}$$

$$x_1, x_2, ..., x_n \geq 0$$

oppure

$$\min \; c_1 x_1 + c_2 x_2 + ... + c_n x_n$$

(s. t.)

$$\begin{aligned} a_{11} x_1 + a_{12} x_2 + ... + a_{1n} x_n &\geq b_1 \\ a_{21} x_1 + a_{22} x_2 + ... + a_{2n} x_n &\geq b_2 \\ &. . . . . \\ a_{m1} x_1 + a_{m2} x_2 + ... + a_{mn} x_n &\geq b_m, \end{aligned}$$

$$x_1, x_2, ..., x_n \geq 0.$$

Nella forma standard, se necessario con un ovvio cambiamento di segno, si può supporre anche che sia $b_i \geq 0$, per ogni i = 1, 2, .., m. Questo risulterà utile, in particolare, per impostare la risoluzione del problema con il metodo del simplesso, come vedremo successivamente.

Si osservi come nelle forme canoniche i vincoli appaiano 'contrastanti' rispetto alla funzione oggetto: questa proprietà sarà ripresa in considerazione quando sarà sviluppata la teoria della dualità.

Mentre le forme canoniche sono generalmente riconosciute come tali in tutta le letteratura, per la forma standard spesso è indicata una funzione oggetto di minimo. La formulazione qui adottata nasce dalla trasformazione della forma canonica di massimo mediante l'introduzione di variabili slack: ciò comporta alcuni vantaggi, assieme anche ad alcuni inconvenienti, che però non invalidano lo sviluppo logico del discorso. D'altra parte, il passaggio dall'una all'altra variante della forma standard è immediato: le conseguenze dell'una o dell'altra scelta saranno indicate ogni volta che ciò si renderà utile.

Più sinteticamente, si possono scrivere i problemi precedenti in forma matriciale. Il problema standard si può scrivere:

$$\max \quad \mathbf{c}\,\mathbf{x}$$

con i vincoli

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

$$\mathbf{x} \geq \mathbf{0}.$$

I problemi in forma canonica si possono formulare come:

$$\begin{array}{rl} \max & \mathbf{c}\,\mathbf{x} \\ \text{s.t.} & \\ & \mathbf{A}\,\mathbf{x} \leq \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0} \end{array} \qquad\qquad \begin{array}{rl} \min & \mathbf{c}\,\mathbf{x} \\ \text{s.t.} & \\ & \mathbf{A}\,\mathbf{x} \geq \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0} \end{array}$$

In queste impostazioni (nelle quali si sono voluti evitare i simboli di trasposizione), **c** è un vettore riga n-dimensionale, **x** un vettore colonna n-dimensionale, **b** un vettore colonna m-dimensionale, **A** una matrice mxn. Nella forma standard si ha m < n e la caratteristica di A è m, per cui le soluzioni del sistema sono ∞^{n-m}: un problema di Programmazione Lineare ha senso quando vi sono infinite soluzioni ammissibili, e questo richiede a sua volta che la matrice dei coefficienti dei vincoli **A**, come si è detto di caratteristica m, non sia quadrata.

Ogni vettore **x** che soddisfa il sistema dei vincoli si dice **soluzione** del problema di PL; se il vettore **x** soddisfa anche i vincoli di non negatività allora è detto **soluzione ammissibile**, mentre ogni vettore ammissibile che rende massima o minima, a seconda dei casi, la funzione oggetto è detto **soluzione ottima**.

L'insieme delle soluzioni ammissibili è detto **Regione Ammissibile**.

La regione ammissibile può essere vuota ed in tal caso il PL si dice **inammissibile**: allora, ovviamente, non esiste a maggior ragione la soluzione ottima. Si può anche dare il caso di problemi in cui la regione ammissibile non è vuota ma non esiste una soluzione migliore di tutte le altre. In questi casi, la regione ammissibile è necessariamente **illimitata**. Tuttavia occorre prestare attenzione al fatto che non vale il viceversa: in presenza di una regione ammissibile illimitata si può avere ottimo finito in un unico punto della regione ammissibile o anche si possono avere infinite soluzioni ottime, tutte con lo stesso valore ottimo (sempre finito!) z* della funzione oggetto.

Il sistema dei vincoli nella forma standard può essere scritto anche in forma vettoriale, introducendo esplicitamente le colonne della matrice A, $\mathbf{a}_j$, j = 1, 2, …, n, come segue:

$$x_1\,\mathbf{a}_1 + x_2\,\mathbf{a}_2 + \ldots + x_n\,\mathbf{a}_n = \mathbf{b}.$$

Questa scrittura mette in evidenza come le componenti di un vettore soluzione **x** si possano interpretare come i coefficienti di una combinazione lineare che dà il vettore **b**. Viste così le cose, un problema di PL può essere considerato come la ricerca del migliore insieme di coefficienti che diano il vettore **b**, utilizzando i vettori colonna della matrice A.

Come si è detto, la forma standard, con $\mathbf{b} \geq \mathbf{0}$, è quella cui si deve ricorrere quando si tratta di risolvere un problema di programmazione lineare con la tecnica del simplesso (che è la tecnica di gran lunga più adoperata); le forme canoniche, al di là della loro reciproca simmetria, sono utili per introdurre il concetto di dualità.

Nel prossimo paragrafo saranno descritti alcuni esempi di problemi di PL che si presentano naturalmente nelle forme canoniche e nella forma standard.

2.1.3 Esempi introduttivi alla programmazione lineare.

Come primo esempio di problema di programmazione lineare consideriamo la definizione di un **piano (ottimale) di produzione di guadagno massimo**. Per un caso particolare di questo problema, si daranno in un successivo paragrafo un'impostazione ed una soluzione grafica.

Un'azienda può produrre un fissato numero n di articoli P_1, P_2, ..., P_n, in quantità da decidere. Un'unità prodotta del generico articolo P_j comporta un guadagno c_j.

Nella produzione sono impiegate delle risorse, R_1, R_2, ..., R_m, in numero appunto di m, disponibili in quantità limitata. Si dirà b_i l'ammontare (positivo) disponibile della risorsa R_i. Sono noti inoltre dei coefficienti a_{ij}, il generico dei quali indica l'ammontare di risorsa i-esima necessaria per produrre un'unità del prodotto j-esimo.

Occorre decidere le quantità x_1, x_2, ..., x_n dei vari prodotti che conviene produrre. Si tratta ovviamente di quantità non negative che devono rispettare i vincoli dovuti alla scarsità di risorse.

Nei casi concreti le risorse potrebbero essere materie prime, ma anche ore-uomo disponibili nei vari reparti che il prodotto deve attraversare.

Da un punto di vista matematico, il guadagno da massimizzare può essere scritto come

$$c_1 x_1 + c_2 x_2 + \ldots + c_n x_n,$$

mentre il generico vincolo, relativo alla i-esima risorsa, può essere scritto come:

$$a_{i1} x_1 + a_{i2} x_2 + \ldots + a_{in} x_n \leq b_i.$$

Si cade pertanto nella forma canonica di massimo vista nel paragrafo precedente.

Se i coefficienti della matrice dei vincoli sono tutti positivi o nulli, e in particolare ogni prodotto consuma una quantità positiva di almeno una delle risorse, il problema ha soluzione ottima finita, nel senso che è garantita l'esistenza di una politica di produzione che produce il massimo guadagno. In casi particolari possono esistere più politiche che forniscono tutte l'ottimo (ed in tal caso si dimostra che sono in numero infinito).

Se (almeno) un coefficiente della matrice risulta negativo, il problema potrebbe non avere soluzione finita, nel senso che, pur essendovi politiche di produzione che rispettano i vincoli, non è detto ve ne sia una migliore di tutte le altre. Impropriamente, si può dire, in tale circostanza, che il guadagno ottimo è 'infinito'. Si osservi che la presenza di coefficienti negativi implica che la fabbricazione di un determinato prodotto, anziché consumare una delle risorse, a

sua volta ne produce, cosa che, in processi produttivi particolari, effettivamente può accadere.

Il problema descritto è di tipo **speculativo**, nel senso che si tratta di decidere il volume di attività che comporta un guadagno massimo.

Si consideri viceversa **il problema di ottenere dei prodotti petroliferi** (quali possono essere benzina per autotrazione, gasolio, benzina avio, bitume, ecc.), in quantitativi minimi prefissati, **al costo più basso possibile**, usufruendo di materie prime alternative P_1, P_2, ..., P_n di differente composizione (petroli di diversa provenienza). Per ogni tipo di petrolio sono noti il costo unitario, c_j, e dei coefficienti a_{ij}, il generico dei quali indica quanto prodotto i-esimo è ottenibile da un'unità del petrolio j-esimo.

Il problema (che, evidentemente, non è speculativo) è quello di individuare i quantitativi da acquistare dei vari petroli, rispettivamente x_1, x_2, …, x_n, che rendano minimo il costo, dato da:

$$c_1 x_1 + c_2 x_2 + \ldots + c_n x_n,$$

e che permettano di ricavare i prodotti (almeno) nelle quantità richieste. Per il generico prodotto i-esimo va rispettato il vincolo:

$$a_{i1} x_1 + a_{i2} x_2 + \ldots + a_{in} x_n \geq b_i,$$

dove con b_i si è indicata appunto la quantità minima di prodotto i-esimo che si vuole ottenere, e pertanto si cade nella formulazione di un problema in forma canonica di minimo.

Un terzo esempio, di rilievo per le applicazioni che saranno sviluppate nella seconda parte, va sotto il nome di **problema di trasporto**.

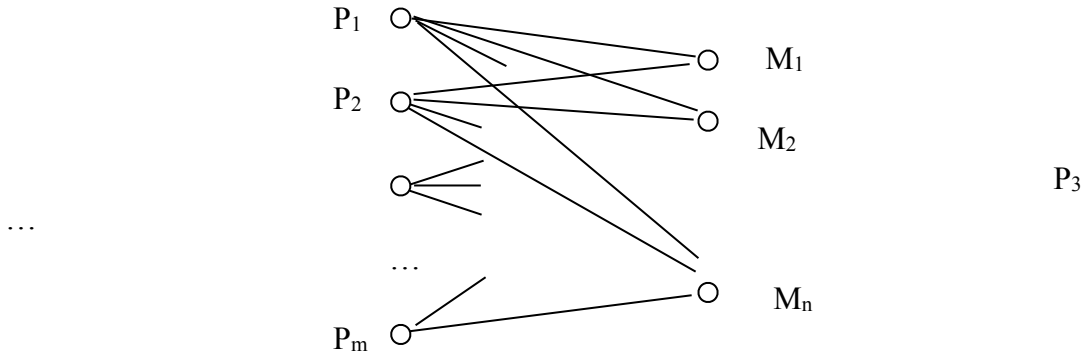
Sono date m origini, P_1, P_2, ..., P_m, luoghi nei quali è prodotto uno stesso tipo di articolo nei quantitativi, rispettivamente, a_1, a_2, ..., a_m, e sono date n destinazioni o luoghi di consumo (mercati) M_1, M_2, ..., M_n. Il generico mercato j-esimo richiede la quantità di articolo b_j.

Si può ipotizzare, senza perdita di generalità, ed eventualmente a costo di semplici trasformazioni, che la somma delle quantità prodotte sia eguale a quella delle quantità richieste:

$$\Sigma_i a_i = \Sigma_j b_j.$$

Infatti, trattandosi di un modello non speculativo, non vi sarebbe soluzione se le quantità prodotte fossero in misura inferiore alla domanda, mentre se si produce di più di quello che serve, si può ipotizzare che esista un ulteriore mercato fittizio che assorba l'eccesso di produzione a costo di trasporto zero.

Ogni luogo di produzione è collegato a ciascun luogo di consumo e viceversa, per cui si può rappresentare graficamente la situazione come segue:



Ogni tratto che congiunge un'origine ad una destinazione rappresenta un percorso. Il generico di questi, da P_i a M_j, comporta un costo unitario di trasporto che indicheremo con c_{ij}. Si tratta di decidere le quantità, la generica delle quali sarà indicata con x_{ij}, che conviene spedire dalle varie origini alle varie destinazioni al minimo costo, nel rispetto di due tipi di vincoli:

- da ogni origine non può uscire più di quanto è prodotto (anzi: deve uscire esattamente quanto è prodotto, vista l'ipotesi $\Sigma_i a_i = \Sigma_j b_j$);
- in ogni destinazione deve arrivare quanto richiesto.

La funzione oggetto, da minimizzare, assume in questo caso un aspetto diverso da quello dei due esempi precedenti perché risulta conveniente attribuire alle variabili di decisione (ed ai loro coefficienti di costo) due indici, uno relativo all'origine e l'altro alla destinazione. Essa risulta:

$$\begin{aligned} &c_{11}\, x_{11} + c_{12}\, x_{12} + \ldots + c_{1n}\, x_{1n} + \\ &c_{21}\, x_{21} + c_{22}\, x_{22} + \ldots + c_{2n}\, x_{2n} + \\ &\quad + \ldots\ldots\ldots\ldots\ldots\ldots\ldots + \\ &c_{m1}\, x_{m1} + c_{m2}\, x_{m2} + \ldots + c_{mn}\, x_{mn}. \end{aligned}$$

Sempre nell'ipotesi che domanda ed offerta complessiva coincidano, i vincoli relativi alle origini sono del tipo:

$$x_{i1} + x_{i2} + \ldots + x_{in} = a_i, \quad \forall\, i,$$

mentre quelli relativi alle destinazioni sono:

$$x_{i1j} + x_{i2j} + \ldots + x_{mj} = b_j \qquad \forall\, j.$$

E' il caso di evidenziare altri aspetti del problema di trasporto.
Innanzi tutto, esso contiene $n \times m$ variabili di decisione, mentre i vincoli sono in numero di $m + n$. La matrice dei coefficienti del sistema di equazioni che formano

i vincoli è formata da elementi che valgono 0 oppure 1. Ogni variabile di decisione compare nei vincoli solo due volte: precisamente, compare in uno (solo) dei vincoli di origine ed in uno (solo) dei vincoli di destinazione. Ciò fa sì che la matrice dei coefficienti abbia due soli elementi eguali ad 1 per ciascuna colonna. Infine va osservato che i vincoli non sono tra di loro indipendenti: sommando i primi m e sottraendo i successivi n-1 si ottiene l'ultimo vincolo.

La matrice dei coefficienti risulta quindi del tipo :

$$
\begin{array}{cccccccccccccccccc}
1 & 1 & 1 & \ldots & 1 & 0 & 0 & 0 & \ldots & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\
0 & 0 & 0 & \ldots & 0 & 1 & 1 & 1 & \ldots & 1 & \ldots & 0 & 0 & 0 & \ldots & 0 \\
\multicolumn{16}{c}{\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots} \\
0 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 & \ldots & 0 & \ldots & 1 & 1 & 1 & \ldots & 1 \\
1 & 0 & 0 & \ldots & 0 & 1 & 0 & 0 & \ldots & 0 & \ldots & 1 & 0 & 0 & \ldots & 0 \\
0 & 1 & 0 & \ldots & 0 & 0 & 1 & 0 & \ldots & 0 & \ldots & 0 & 1 & 0 & \ldots & 0 \\
0 & 0 & 1 & \ldots & 0 & 0 & 0 & 1 & \ldots & 0 & \ldots & 0 & 0 & 1 & \ldots & 0 \\
\multicolumn{16}{c}{\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots} \\
0 & 0 & 0 & \ldots & 1 & 0 & 0 & 0 & \ldots & 1 & \ldots & 0 & 0 & 0 & \ldots & 1
\end{array}
$$

e può essere ripartita in due sottomatrici, la prima (sottomatrice superiore) contenente le prime m righe, la seconda (sottomatrice inferiore) le successive n. La prima sottomatrice contiene m blocchi (stringhe) di elementi eguali a 1 (n elementi per ciascun blocco), collocati diagonalmente, a partire dalle posizioni in alto a sinistra; la sottomatrice inferiore è costituita a sua volta da m sottomatrici identiche affiancate, ciascuna di ordine n.

Va detto anche, per completezza, che il problema di trasporto, anche se si può risolvere con il metodo del simplesso - che sarà illustrato più avanti - e quindi con lo stesso software generalmente in uso per qualsiasi problema di programmazione lineare, è risolubile con tecniche ad hoc che utilizzano la sua particolare struttura.

Un ulteriore esempio di problema di programmazione lineare è costituito dal classico **problema della dieta**. Si suppone di dover individuare una dieta ammissibile di costo minimo, cioè i quantitativi di alcuni alimenti disponibili, di cui sono noti i prezzi e il contenuto in fattori nutritivi, da acquistare, ad esempio giornalmente, per costituire una dieta.

Formalmente, siano F_1, F_2, …., F_n n alimenti che si possono acquistare sul mercato, il generico dei quali F_j ha un prezzo c_j. Sono dati anche m fattori nutritivi Q_1, Q_2, …., Q_m: il quantitativo minimo necessario del fattore nutritivo Q_i per una corretta alimentazione è b_i. E' nota anche una matrice di coefficienti, il generico dei quali a_{ij} esprime il quantitativo di fattore nutritivo Q_i presente in un'unità di alimento F_j.

Indicando con x_j il quantitativo da acquistare dell'alimento F_j, si tratta di risolvere il problema:

$$\min \; c_1 x_1 + c_2 x_2 + \ldots + c_n x_n$$

con i vincoli

$$
\begin{aligned}
a_{11} x_1 + a_{12} x_2 + \ldots + a_{1n} x_n &\geq b_1 \\
a_{21} x_1 + a_{22} x_2 + \ldots + a_{2n} x_n &\geq b_2 \\
&\ldots\ldots\ldots \\
a_{m1} x_1 + a_{m2} x_2 + \ldots + a_{mn} x_n &\geq b_m,
\end{aligned}
$$

ed i vincoli di non negatività

$$x_1, x_2, \ldots, x_n \geq 0.$$

Il problema ha certamente soluzione se ogni elemento nutritivo è contenuto in almeno un alimento.

Come modello, quello appena formulato presenta parecchi punti deboli.
Innanzi tutto non tiene conto evidentemente del mix di alimenti che si potrebbero venire a creare e che potrebbero essere improponibili nei casi concreti: da questo punto di vista si potrebbe considerare più adatto a situazioni come quelle di un allevamento di bestiame. Inoltre il modello ha carattere statico: proporre una dieta costante nel tempo è ancora poco accettabile. Per ovviare a questi inconvenienti, diversi autori hanno proposto in letteratura altri modelli che prevedono, piuttosto che l'individuazione di quantitativi di alimenti, la costruzione di veri e propri menù.

Altro elemento poco realistico in casi concreti è dato dall'assumere l'additività nei quantitativi di elementi nutritivi e la loro indipendenza: è noto infatti che gli effetti congiunti dell'assunzione simultanea di più alimenti può ingenerare effetti sinergici o viceversa di blocco, per cui la presenza di certi quantitativi di un fattore nutritivo esalta o al contrario inibisce l'utilizzo da parte dell'organismo di altri fattori. Sotto questo riguardo è la linearità del modello che potrebbe essere messa in discussione, come può succedere anche per gli altri esempi sopra citati.

In effetti, nel problema del piano produttivo di guadagno massimo, non si tiene conto del fatto che spesso non vi è linearità nei prezzi praticati (è molto facile che si presenti il fenomeno degli sconti sulle quantità, e, d'altra parte, non è nemmeno detto che vi sia completa additività nel consumo delle risorse). Nel caso poi del problema dei trasporti, a critiche di natura simile alle precedenti si può aggiungere quella che il supporre il costo di trasporto proporzionale alle quantità movimentate va contro il fatto ben noto che il costo dello spostamento del mezzo che effettua il trasporto spesso è quasi indipendente dal carico: si può tuttavia considerare il modello adatto a situazioni in cui le variabili di decisione assumono valori elevati, nel senso che si tratta di effettuare spedizioni tra origini e destinazioni che coinvolgono un rilevante numero di autoveicoli, in modo tale che, almeno con una certa approssimazione, la proporzionalità tra merce trasportata e costo sia ristabilita.

2.1.4 Risoluzione grafica di un problema di produzione in due variabili.

Allo scopo di illustrare alcune caratteristiche dei problemi di PL, in particolare gli aspetti geometrici e la tecnica risolutiva comunemente usata, cioè la tecnica del simplesso, si studierà in questo paragrafo un semplice esempio numerico che può essere considerato un caso (didattico) di problema speculativo di produzione.

Si supponga che l'azienda Ciclo S.p.A. possa produrre due tipi di articoli: **biciclette da corsa** e **biciclette sportive**.

Queste comportano un guadagno unitario per l'azienda, rispettivamente di 8 e 10 unità.

Si abbiano poi i seguenti dati tecnici di produzione:

- per produrre una bicicletta da corsa occorrono 10 ore di officina stampaggio, 5 di montaggio e 4 per le verniciature;
- per una bicicletta sportiva occorrono 5 ore di stampaggio, 6 ore di montaggio e 10 per le verniciature;
- i reparti di stampaggio, montaggio e verniciatura dispongono rispettivamente di 200, 120 e 160 ore macchina mensili.

Occorre decidere quante biciclette produrre di ciascun tipo nel periodo di riferimento per massimizzare il guadagno.

Si suppone, nel modello, che qualunque numero di biciclette dei due tipi sarà in ogni caso vendibile sul mercato e che eventuali soluzioni frazionarie indichino in realtà il numero medio di prodotti da fabbricare per unità di tempo lavorativo, cioè in un mese (in caso contrario, se si vogliono ottenere soluzioni intere, il modello di PL non è adeguato, e si dovrà ricorrere, appunto, alla Programmazione Intera).

I dati possono essere raccolti nella seguente tabella :

	biciclette da corsa	biciclette sportive	ore complessive disponibili
guadagni unitari	8	10	
ore di stampaggio	10	5	200
ore di montaggio	5	6	120
ore di verniciatura	4	10	160

Indicando con x la quantità da produrre di biciclette da corsa e con y quella delle biciclette sportive, si può impostare il seguente problema :

$$\max \ (8\,x + 10\,y)$$

con i vincoli

$$\begin{aligned} 10\,x + 5\,y &\leq 200, \\ 5\,x + 6\,y &\leq 120, \\ 4\,x + 10\,y &\leq 160, \end{aligned}$$

$$x, y \geq 0.$$

Il problema può essere affrontato per via grafica, tracciando innanzi tutto su un piano cartesiano Oxy le rette

$$10\,x + 5\,y = 200 \ ; \quad 5\,x + 6\,y = 120 \ ; \quad 4\,x + 10\,y = 160.$$

I vincoli individuano, insieme con gli assi coordinati, la regione ammissibile RA: si tratta di un pentagono irregolare avente come vertici l'origine O (0, 0), i punti S (0, 16) e R (20, 0), rispettivamente sull'asse delle ordinate e delle ascisse ed i punti P (120/13, 160/13) e Q (120/7, 40/7). Si tratta allora di individuare quel punto (o quei punti) della regione ammissibile in corrispondenza del quale si ha il valore massimo per la funzione oggetto.

A tale scopo si considerino la funzione $z = 8\,x + 10\,y$ e l'equazione

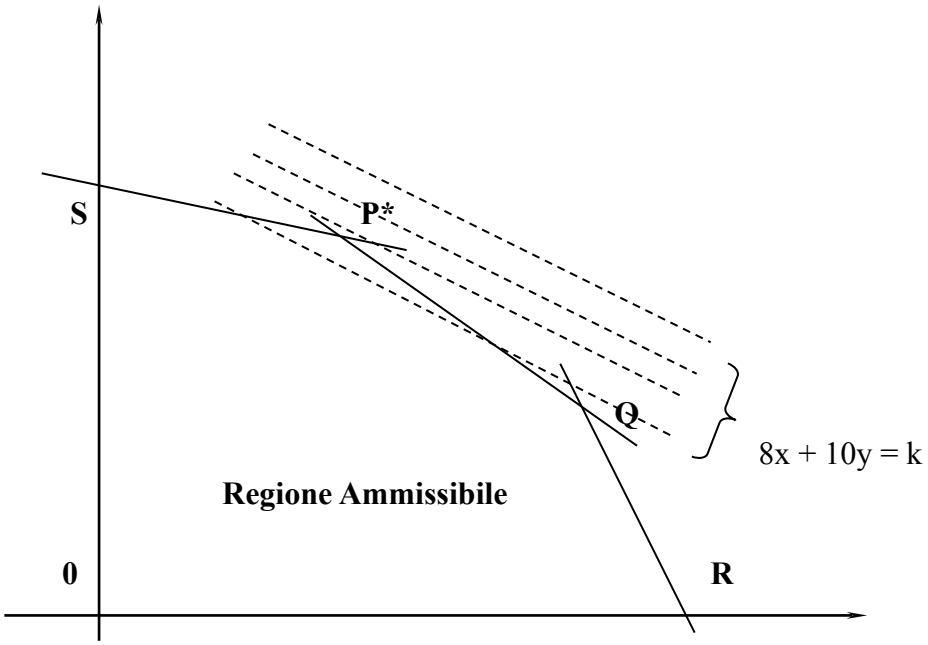
$$8\,x + 10\,y = k,$$

che, fissato il valore di k, rappresenta una retta e precisamente la retta contenente tutte le combinazioni di prodotti x e y che danno lo stesso guadagno k. Variando k si ottengono rette tra di loro parallele dette **rette di livello**. Di queste interessa

quella che, avendo il valore di k più elevato, ha intersezione non vuota con la regione ammissibile.

Nel nostro caso si tratta della retta tratteggiata la quale ha in comune con la regione ammissibile solo il punto P, in corrispondenza del quale il valore della funzione oggetto risulta $z^* = 2560/13 = 196.92$.

Come verifica, si può procedere al calcolo della funzione oggetto negli altri vertici del pentagono (si ottengono sempre valori inferiori a z^*: ad es., in Q risulta $z = 194.28$ e così via).



E' il caso di osservare che la soluzione che si ottiene, in questo caso ma anche più in generale, non ha coordinate intere. Come accennato in precedenza, se da un lato può apparire privo di significato, trattandosi della produzione di biciclette, d'altra parte è egualmente accettabile: essendo la soluzione formata da numeri razionali, cambiando il periodo di riferimento, è possibile farli diventare interi, riferendoli ad un arco di tempo adeguato. In ogni caso il dato è indicativo della politica ottima, almeno con una certa approssimazione.

In effetti, la risoluzione del problema appena proposto, con l'ulteriore requisito che la soluzione sia costituita da componenti intere, rende il problema stesso molto più difficile, con un vantaggio pratico discutibile.

2.1.5 La soluzione del problema 'biciclette' con la tecnica del simplesso.

La risoluzione per via grafica del caso prospettato nel paragrafo precedente è stata possibile per la presenza di due sole variabili di decisione. Con un numero di variabili superiore, è necessario ricorrere ad algoritmi adeguati. E' utile comunque, per chiarire come procede l'algoritmo del simplesso, illustrarne i vari passaggi, sempre riferiti al caso del problema della produzione delle biciclette.

Innanzi tutto occorre trasformare il problema nella forma standard, e ciò richiede l'introduzione di tre variabili slack, che diremo s_1, s_2, s_3. Si ottiene:

$$\max\ (8\,x + 10\,y + 0\,s_1 + 0\,s_2, + 0\,s_3)$$

con i vincoli

$$\begin{aligned} 10\,x + 5\,y + s_1 &= 200, \\ 5\,x + 6\,y + s_2 &= 120, \\ 4\,x + 10\,y + s_3 &= 160, \\ x, y, s_1, s_2, s_3 &\geq 0. \end{aligned}$$

Il procedimento prosegue trasformando opportunamente la funzione oggetto ed il sistema dei vincoli in forme equivalenti in modo tale che, dopo ogni trasformazione, vi sono una soluzione corrente ed un valore z della funzione obiettivo facilmente leggibili e una struttura che consente di capire se il procedimento ha prodotto la soluzione ottima oppure deve ancora continuare. Infatti:

- la funzione oggetto contiene solo le (due) variabili che al momento attuale (cioè nella soluzione corrente) assumono valore 0 (le altre tre variabili nella funzione oggetto hanno coefficiente nullo);

- il sistema dei vincoli consente la lettura di una soluzione in via immediata, in quanto vi sono sempre tre variabili ciascuna delle quali compare in un vincolo solo e quindi, ponendo a zero le rimanenti due, il valore di queste tre variabili coincide con i termini noti delle equazioni vincolari;

- le variabili che nella soluzione corrente sono eguali a zero presentano nella funzione oggetto dei coefficienti che indicano se risulta conveniente fare assumere loro valore positivo (e quindi procedere ad un'altra trasformazione di simplesso).

Inizialmente (nella forma standard appena scritta) si può costatare che:

- facendo assumere valore 0 alle variabili x e y (che compaiono in tutti e tre i vincoli e anche nella funzione oggetto) si ottiene la prima soluzione:

$$x = 0;\ y = 0;\ s_1 = 200;\ s_2 = 120;\ s_3 = 160$$

ed in corrispondenza il valore della funzione oggetto è $z = 0$;

- dalla forma della funzione oggetto si intuisce che, essendo positivi sia il coefficiente di x che quello di y, converrà cercare di incrementare il valore di (almeno) una di queste due variabili.

Il passo successivo è quello di scegliere una tra le variabili che si ritiene conveniente incrementare: un criterio può essere quello di individuare il coefficiente positivo maggiore (nel nostro caso, quello della y), ma si può operare anche una scelta a piacere.

Si incrementa adesso il valore della sola variabile scelta (le altre rimangono nulle: pertanto, se si decide di incrementare y, la variabile x rimane a quota zero). La variabile y può aumentare solo a scapito delle altre variabili che sono attualmente positive: lo può fare solo sino al punto in cui una di queste diventa zero (si ricordi che tutte le variabili devono restare non negative).

In particolare, nei riguardi di s_1 (nel primo vincolo) y può crescere sino al valore di 40; nel secondo vincolo, a spese di s_2, y può crescere sino a 20; infine nel terzo vincolo, y può crescere sino a 16.

In definitiva, è proprio questo terzo vincolo a determinare la crescita massima possibile di y: in corrispondenza, la variabile s_1 diventa nulla, mentre s_2 ed s_3 diminuiscono di valore, ma restano positive.

Si procede allora come segue:

- si fa diventare 1 il coefficiente di y nel terzo vincolo (quello per il quale un'altra variabile diventa 0): a tale scopo basta dividere entrambi i membri per 10;

- si elimina y dai vincoli restanti, e questo può essere fatto sia per sostituzione che sommando membro a membro i vincoli in maniera adeguata: al primo vincolo si può sommare la nuova formulazione del terzo, moltiplicandone I° e II° membro per -5; al secondo vincolo si può sommare ancora la nuova formulazione del terzo, moltiplicandone ambo i membri per -6;

- si elimina y anche dalla funzione oggetto, sempre ricavando l'espressione di y dal terzo vincolo e sostituendo nella funzione oggetto stessa.

Effettuate tutte queste operazioni, il problema assume l'aspetto:

$$\max\ (4x + 0y + 0s_1 + 0s_2 - s_3 + 160)$$

con i vincoli

$$\begin{aligned} 8x + s_1 - \tfrac{1}{2}s_3 &= 120, \\ \tfrac{13}{5}x + s_2 - \tfrac{3}{5}s_3 &= 24, \\ \tfrac{4}{10}x + y + \tfrac{1}{10}s_3 &= 16, \end{aligned}$$

oltre ai vincoli di non negatività.

La nuova formulazione consente di affermare che:

- vi è una soluzione nel punto $x = 0$, $y = 16$, $s_1 = 120$, $s_2 = 24$, $s_3 = 0$, punto in corrispondenza del quale la funzione oggetto vale 160;
- la presenza di un coefficiente positivo nella funzione oggetto per la variabile x (che al momento vale 0) suggerisce che conviene incrementare questa variabile.

E' da osservare, a questo punto, il nuovo valore che nella funzione oggetto assume il coefficiente di guadagno per ogni bicicletta da corsa: questo inizialmente era 8, mentre ora è solo 4. Ciò si giustifica con il fatto che, mentre nelle condizioni iniziali, in cui ancora non si era deciso di produrre nulla, il guadagno unitario era 8, adesso che si è ipotizzato di portare a 16 la produzione di biciclette sportive, la produzione anche di biciclette da corsa può essere effettuata solo a scapito di qualche bicicletta sportiva, data la limitatezza delle risorse a disposizione. Quindi, non si può tenere conto solo del guadagno aggiuntivo che deriva da questa nuova produzione, ma occorre anche considerare l'inevitabile riduzione nella produzione dell'altro tipo di biciclette! Il coefficiente 4, che prende il nome di **coefficiente di guadagno ridotto**, indica comunque che tale sostituzione nella produzione, entro certi limiti, risulta ancora conveniente.

Si procede ora analogamente al passo precedente: si individua quanto vale l'incremento massimo che può subire la variabile di decisione x; determinato il vincolo che pone tale limitazione, si riscrive il vincolo stesso con coefficiente della x posto eguale a 1 e si elimina la variabile x dagli altri vincoli e dalla funzione oggetto.

Alla fine del procedimento - avendo constatato che, al crescere di x, il vincolo in cui per prima si azzera una variabile, precedentemente positiva, è il secondo - il problema diventa:

$$\max\ (0\,x + 0\,y + 0\,s_1 - 20/13\,s_2 - 1/13\,s_3 + 2560/13)$$

con i vincoli

$$\begin{aligned} s_1 &- 40/13\,s_2 + 35/26\,s_3 &&= 600/13,\\ x &+ 5/13\,s_2 - 3/13\,s_3 &&= 120/13,\\ y &- 2/13\,s_2 + 5/26\,s_3 &&= 160/13, \end{aligned}$$

oltre ai vincoli di non negatività.

La formulazione appena data consente di stabilire che il procedimento è terminato e che si è individuata la politica ottima. Infatti, nella funzione oggetto non compaiono variabili di decisione con coefficienti positivi, mentre i vincoli, stabilito che il valore delle ultime due variabili slack è 0, consentono di individuare la soluzione:

$$x = 120/13,\ y = 160/13,\ s_1 = 600/13,\ s_2 = 0,\ s_3 = 0,$$

ed il valore ottimo della funzione oggetto è 2560/13 = 196.92.

Sulla procedura, da un punto di vista teorico, torneremo nei prossimi paragrafi.

2.1.6 Aspetti algoritmici: generalità sul metodo del simplesso.

Dalla rappresentazione grafica relativa all'esempio precedente, che fa riferimento ad un problema in forma canonica, si può intuire come la soluzione ottima di un problema di programmazione lineare, se esiste, cade necessariamente

sulla **frontiera** della regione ammissibile. Anzi, in generale cade su un punto che è intersezione di un adeguato numero di vincoli. Finché le variabili in gioco sono solo due, tale intersezione è quella tra due rette (ed è identificabile anche per via geometrica); se le variabili sono tre, il punto ottimo è normalmente l'intersezione di tre piani (e la soluzione geometrica diventa problematica); con più di tre variabili, l'ottimo, se unico, è intersezione di un congruo numero di iperpiani e per individuarlo occorre ricorrere a strumenti analitici. Si deve considerare poi che vi sono casi in cui le soluzioni ottime sono in numero infinito (in due dimensioni, tutti i punti di un segmento o di una semiretta) ed anche casi in cui la regione ammissibile è vuota: occorrono evidentemente strumenti idonei a distinguere tra tutte queste possibilità.

L'esempio precedente potrebbe in effetti lasciare spazio all'idea che una via di ricerca della soluzione ottima potrebbe essere quella di testare tutte le intersezioni tra vincoli, o, meglio ancora, solo le intersezioni tra vincoli che stanno sulla frontiera della regione ammissibile e individuare la migliore (o le migliori), ma ciò è vero solo se la regione ammissibile è limitata.

Sempre sull'esempio numerico trattato, si può notare come il problema, impostato in forma canonica, sia stato trasformato come primo passo nella forma standard:

$$\max \ \mathbf{c}\,\mathbf{x} \qquad \text{con i vincoli} \quad \mathbf{A}\,\mathbf{x} = \mathbf{b}, \qquad \mathbf{x} \geq \mathbf{0}.$$

Inoltre dall'esempio stesso, si può costatare come le intersezioni tra le rette che esprimono i vincoli e tra le rette stesse e gli assi coordinati (nella forma canonica) corrispondano a quelle che in algebra lineare sono definite come le **soluzioni di base** del sistema nella forma standard.

Si può notare, dalla figura dell'esempio relativo alla produzione di biciclette, come le intersezioni tra le rette e tra le rette e gli assi si ottengono ponendo volta a volta a zero due delle cinque variabili in gioco (che sono, ricordiamolo, le variabili x e y e le tre variabili scarto): ad esempio, l'intersezione degli assi corrisponde a porre a zero x e y e a determinare il valore delle tre slack; l'intersezione tra il primo vincolo e l'asse y corrisponde all'azzeramento della variabile slack dello stesso vincolo e della x mentre sono diverse da zero y e le altre due slack; in generale, l'intersezione di due vincoli si ottiene azzerando le slack dei vincoli stessi e calcolando il valore delle tre variabili rimanenti.

La regola diventa generale se si considerano anche i due assi come rette che esprimono vincoli. Le intersezioni che coincidono con i vertici della regione ammissibile sono soluzioni ammissibili (e si vedrà che si tratta di soluzioni di base), mentre le intersezioni che cadono al di fuori della regione ammissibile corrispondono a soluzioni (di base) non ammissibili.

Algebricamente, per ottenere tutte le soluzioni di base occorre individuare (nel problema in forma standard) tutte le sottomatrici B della matrice A che possono costituire una base per lo spazio vettoriale a m dimensioni (v. oltre) e risolvere sistemi quadrati del tipo

$$\mathrm{B}\,\mathbf{x}_\mathrm{B} = \mathbf{b},$$

dove $\mathbf{x}_\mathrm{B}$ è un vettore a m componenti formato esattamente dalle m componenti di $\mathbf{x}$ che corrispondono alle colonne di B.

Occorre poi individuare una soluzione che risulti **ammissibile** (con le componenti tutte non negative) e che fornisca il **miglior valore** per la funzione oggetto.

Ciò però, è il caso di ribadire, ancora non sarebbe sufficiente, perché rimarrebbe un dubbio: in realtà la soluzione ottima potrebbe risultare inesistente (cioè, impropriamente, l'ottimo potrebbe essere infinito) e quella individuata potrebbe essere solo la migliore tra le soluzioni di base, mentre la regione ammissibile, illimitata, consentirebbe anche soluzioni migliori.

Il metodo del simplesso è una tecnica di ricerca intelligente della soluzione ottima che, esaminando solo (alcune) soluzioni corrispondenti alle intersezioni tra vincoli, e cioè soluzioni di base, consente con opportune regole di arresto di rispondere anche al quesito sulla eventuale non ammissibilità del problema o sul fatto che l'ottimo sia infinito.

Per delineare il metodo con maggiore dettaglio, è opportuno richiamare alcune definizioni sugli insiemi convessi ed alcune nozioni di algebra lineare.

Si farà nel seguito sempre riferimento ad un problema in forma standard espresso in forma matriciale.

2.1.7 Soluzioni di base e insiemi convessi.

2.1.7.1 Soluzioni di base

Dato un sistema di equazioni lineari

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

supponiamo che la matrice A, del tipo m×n, con m<n, abbia caratteristica m. Ciò implica che tra le colonne di A ve ne sono m di linearmente indipendenti: anzi, in generale, la matrice A conterrà più insiemi di m colonne linearmente indipendenti. Scelto uno di tali insiemi, diciamo B la sottomatrice quadrata formata dalle colonne stesse. Ponendo a zero le incognite corrispondenti alle colonne che non costituiscono B, si ottiene un sistema quadrato che ammette un'unica soluzione (che si può calcolare con varie tecniche, quali ad esempio la regola di Cramer o procedendo a triangolarizzare la matrice B). Completando tale soluzione (di m componenti) con le componenti precedentemente poste a zero, si ha una **soluzione di base** del sistema originario. Le m componenti di **x** corrispondenti alle colonne della matrice B vengono dette **componenti in base**, mentre le altre sono dette **fuori base**.

Può capitare che anche tra le componenti in base ve ne siano di nulle: in tal caso la soluzione di base viene detta **degenere**.

La denominazione di **soluzione di base** deriva dal fatto che, nello spazio vettoriale dei vettori colonna a m componenti, un qualsiasi insieme di m vettori linearmente indipendenti costituisce una base. Il concetto di **base**, a sua volta, sta ad indicare il fatto che dato un qualsiasi vettore (colonna) **b** dello spazio, esso è ottenibile come combinazione lineare (con coefficienti univocamente individuati) dei vettori della base. Le soluzioni di base sono non più del numero delle

combinazioni delle n colonne, considerate m alla volta, ma possono ridursi perché non è detto che **tutti** i possibili sottoinsiemi di m vettori siano linearmente indipendenti.

Se si suppone, per semplicità, che siano linearmente indipendenti le prime m colonne, indicando con B la sottomatrice formata dalle stesse, una soluzione di base si ottiene ponendo eguali a zero le ultime n-m componenti del vettore **x** e risolvendo un sistema che si scrive come

$$B\,\mathbf{x}_B = \mathbf{b}.$$

La soluzione di base corrispondente è data poi dal vettore $(\mathbf{x}_B, \mathbf{0})$, con $\mathbf{x}_B = B^{-1}\,\mathbf{b}$, avente ancora n componenti.

Si osservi che, data una soluzione di base non degenere, le componenti diverse da zero individuano m colonne costituenti una base e tra questi due insiemi (soluzioni di base e basi) si instaura una corrispondenza biunivoca, nel senso che ad ogni soluzione non degenere corrisponde una base e viceversa.

Nel caso di una soluzione di base degenere, con $p < m$ componenti positive, tale corrispondenza viene a mancare perché la stessa soluzione può essere messa in corrispondenza con tutte le basi contenenti le p colonne che corrispondono alle componenti positive di **x**. In altre parole, in caso di degenerazione di una soluzione, vi è una certa libertà nel costruire una base associata, scegliendo a piacere m-p colonne tra quelle che corrispondono a componenti nulle di **x** (purché, ovviamente, il complesso delle m colonne scelte goda della proprietà della indipendenza lineare).

Per esemplificare quanto detto, si consideri il sistema:

$$\begin{aligned} 2\,x + 3y + 4w &= 7 \\ x - y + 3w &= -1. \end{aligned}$$

Ricorrendo alla posizione

$$\mathbf{p} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad \mathbf{q} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}, \quad \mathbf{r} = \begin{pmatrix} 4 \\ 3 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 7 \\ -1 \end{pmatrix},$$

lo stesso sistema può essere scritto

$$x\,\mathbf{p} + y\,\mathbf{q} + w\,\mathbf{r} = \mathbf{b}.$$

Si può verificare facilmente che le tre coppie di vettori colonna (**p**, **q**), (**p**, **r**), (**q**, **r**) costituiscono ciascuna una base per lo spazio vettoriale di dimensione due, in quanto tutte e tre sono coppie di vettori linearmente indipendenti (il determinante della matrice formata da tali vettori colonna è diverso da zero): si può pertanto ottenere **b** come combinazione lineare di due soli dei tre vettori, **p**, **q**, **r**, e porre a zero il coefficiente del vettore rimanente.

Pertanto, ipotizzando $w = 0$ e risolvendo quindi rispetto a x e y, si ha la soluzione (di base):

$$x = 4/5, \quad y = 9/5, \quad w = 0.$$

Una seconda soluzione di base si può ottenere ponendo $y = 0$ e risolvendo allora il sistema nelle incognite x e w. Si ha:

$$x = 25/2, \quad y = 0, \quad w = -9/2.$$

Infine, e lasciamo il calcolo al lettore, si può ottenere una terza soluzione di base ponendo a 0 il valore di x e risolvendo rispetto a y e a w.

In generale, per un sistema A**x** = **b**, e nelle ipotesi fatte all'inizio di questo paragrafo, le soluzioni di base sono, al più, tante quante le combinazioni delle n colonne, prese a m a m: $\binom{n}{m}$.

Una soluzione di base nella quale tutte le componenti siano non negative (quindi, positive o nulle) si dice **soluzione di base ammissibile**.

Nell'esempio precedente, la prima soluzione di base ottenuta è ammissibile, la seconda non lo è.

Come è stato già anticipato, e verrà precisato nei prossimi paragrafi, in un PL la soluzione ottima viene ricercata tra le sole soluzioni di base ammissibili. Il numero di queste non è definibile in generale, ma si tratta in ogni caso di un valore inferiore a $\binom{n}{m}$.

2.1.7.2 Insiemi convessi.

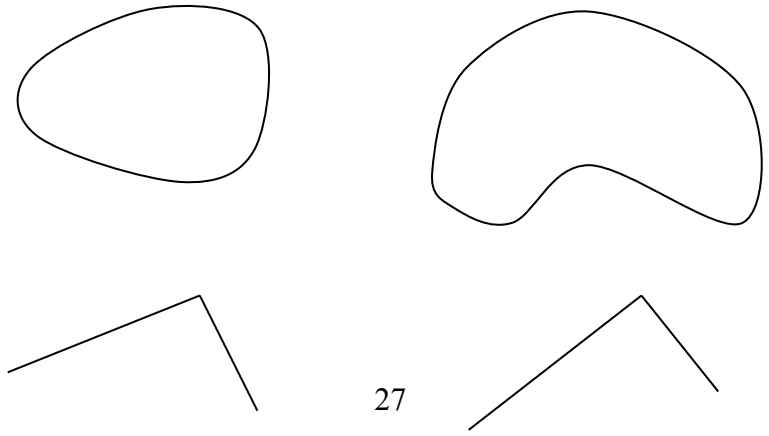
Gli **insiemi convessi** (di punti) sono caratterizzati dal fatto che **il segmento che congiunge due qualsiasi punti dell'insieme appartiene interamente all'insieme stesso**.

Da un punto di vista vettoriale, un insieme I di vettori è convesso se dati due vettori **x** e **y** appartenenti all'insieme, appartiene all'insieme anche qualsiasi **combinazione lineare convessa**, cioè qualsiasi combinazione del tipo

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

con $0 \leq \alpha \leq 1$.

In un insieme convesso (di punti o di vettori) si distinguono particolari elementi che (se punti) non risultano interni ad alcun segmento appartenente all'insieme, oppure (se vettori) non sono ottenibili come combinazioni lineari convesse di altri vettori. Tali punti (e tali vettori) sono detti **vertici** dell'insieme convesso. Algebricamente, un vertice non è ottenibile da una combinazione convessa come quella sopra scritta con **x** e **y** appartenenti all'insieme e α strettamente compreso tra 0 e 1.



(i) (ii)

Esempi di insiemi convessi (i) e non convessi (ii)

2.1.8 Aspetti geometrici ed algebrici delle soluzioni di un PL.

I concetti appena richiamati di soluzione di base e di insieme convesso sono il fondamento di alcuni teoremi che caratterizzano la regione ammissibile di un qualsiasi problema di PL e costituiscono le basi teoriche dell'algoritmo del simplesso.

Più precisamente, valgono i seguenti teoremi.

La regione ammissibile di un problema di programmazione lineare, se non è vuota, è un insieme convesso.

Dimostrazione. Per dimostrare l'enunciato si considerino due vettori, **x** e **w**, soluzioni ammissibili per un PL in forma standard. Occorre provare che anche la generica combinazione lineare convessa $\mathbf{y} = \alpha\mathbf{x}+(1-\alpha)\mathbf{w}$ è soluzione ammissibile per lo stesso PL. In effetti, per il modo come è costruito, il vettore **y** soddisfa i vincoli di non negatività (come **x** e **w**); inoltre, poiché da $A\mathbf{x}=\mathbf{b}$, segue anche $\alpha A\mathbf{x} = \alpha\mathbf{b}$ e da $A\mathbf{w} = \mathbf{b}$ segue anche $(1-\alpha)A\mathbf{w} = (1-\alpha)\mathbf{b}$, sommando a membro a membro si ottiene:

$$\alpha\, A\mathbf{x}+(1-\alpha)A\,\mathbf{w} = A\mathbf{y} = \alpha\mathbf{b}+(1-\alpha)\mathbf{b} = \mathbf{b},$$

cioè anche **y** è soluzione ammissibile.

Le implicazioni di tale teorema sono notevoli: la regione ammissibile non può essere costituita da due o più sottoregioni disgiunte. Se si tratta di un insieme limitato, assume l'aspetto di un poliedro convesso (in due dimensioni, si ha un poligono convesso).

In un PL, le soluzioni di base ammissibili del sistema dei vincoli corrispondono ai vertici della regione ammissibile.

Dimostrazione. Occorre dimostrare, da un lato, che dato un vertice della Regione Ammissibile esso corrisponde ad una soluzione di base ammissibile, d'altra parte che data una soluzione di base (ammissibile) essa corrisponde ad un vertice di RA. Entrambe le dimostrazioni procedono per assurdo.

Per quanto riguarda la prima parte, sia **x** un vettore soluzione ammissibile corrispondente ad un vertice e supponiamo che abbia k ($\leq$n) componenti strettamente positive e supponiamo anche, per comodità, che si tratti delle prime k. Ciò significa che vale l'eguaglianza

$$x_1\,\mathbf{a}_1 + x_2\,\mathbf{a}_2 + \ldots + x_k\,\mathbf{a}_k = \mathbf{b},$$

dove i k vettori $\mathbf{a}_j$ (j=1, 2, …, k) costituiscono le prime k colonne di A. Affermare che **x** è una soluzione di base equivale ad affermare che i k vettori colonna in questione sono linearmente indipendenti. In effetti, se per assurdo questi vettori fossero linearmente dipendenti, allora esisterebbe una loro combinazione lineare nulla con coefficienti non tutti nulli

$$y_1 \mathbf{a}_1 + y_2 \mathbf{a}_2 + \ldots + y_k \mathbf{a}_k = \mathbf{0}.$$

Se diciamo allora **y** il vettore (n-dimensionale)

$$\mathbf{y} = (y_1 \quad y_2 \quad \ldots \quad y_k \quad 0 \quad 0 \quad \ldots \quad 0),$$

prendendo un numero $\varepsilon > 0$ sufficientemente piccolo, i due vettori $\mathbf{p} = \mathbf{x} - \varepsilon\ \mathbf{y}$ e $\mathbf{q} = \mathbf{x} + \varepsilon\ \mathbf{y}$ risulterebbero soluzioni ammissibili - differenti dal vettore **x** - (è pressoché banale verificare che hanno componenti non negative e risolvono il sistema $A\mathbf{x} = \mathbf{b}$; d'altra parte qualche componente di **y** è strettamente diversa da zero quindi **p** e **q** non possono coincidere con **x**): **x** sarebbe a questo punto una loro combinazione lineare convessa ($\mathbf{x} = ½\ \mathbf{p} + ½\ \mathbf{q}$) e quindi il punto di mezzo del segmento che congiunge i due punti ammissibili **p** e **q**, contro l'ipotesi che **x** sia un vertice. Pertanto non è possibile che i k vettori colonna $\mathbf{a}_j$, (j=1, 2, …, k) siano linearmente dipendenti: essi viceversa sono linearmente indipendenti e quindi **x** è una soluzione di base.

Per dimostrare la seconda parte procediamo ancora per assurdo. Consideriamo allora una soluzione di base ammissibile **x** e supponiamo che essa sia combinazione convessa di altri due vettori, anch'essi ammissibili, **y** e **w**: $\mathbf{x} = \alpha\ \mathbf{y} + (1\text{-}\alpha)\ \mathbf{w}$. Si vedrà, in realtà, che ciò non è possibile e che questi due vettori devono forzatamente coincidere con **x**. Supponiamo, per comodità, che le componenti di **x** fuori base, e pertanto nulle, siano le ultime n-m. Si può allora scrivere il seguente insieme di eguaglianze:

$$x_1 = \alpha\ y_1 + (1\text{-}\alpha)\ w_1;\ x_2 = \alpha\ y_2 + (1\text{-}\alpha)\ w_2;\ \ldots\ ;\ x_m = \alpha\ y_m + (1\text{-}\alpha)\ w_m;$$
$$0 = x_{m+1} = \alpha\ y_{m+1} + (1\text{-}\alpha)\ w_{m+1};\ \ldots.\ ;\ 0 = x_n = \alpha\ y_n + (1\text{-}\alpha)\ w_n.$$

Poiché **y** e **w** sono ammissibili, anche le loro ultime n-m componenti non possono che essere 0 (si osservi che altrimenti non sarebbe possibile ottenere 0 come combinazione lineare di elementi non negativi). Ma allora sia **x** sia **y** sia **w** hanno nulle le stesse ultime n-m componenti e quindi sono soluzioni (di base) corrispondenti alla stessa matrice di base: in altre parole, risolvono lo stesso sistema (quadrato), perciò devono coincidere. Se ne può concludere, come si è già detto, che è impossibile ottenere **x** come combinazione lineare convessa di vettori diversi da **x** stesso e quindi **x** è un vertice della Regione Ammissibile.

Premessi questi teoremi, si può enunciare il

Teorema fondamentale della Programmazione Lineare.

Dato un problema di Programmazione Lineare PL in forma standard:

$$\mathbf{max\ cx \quad s.t. \quad Ax = b; \quad x \geq 0,}$$

in cui la matrice A, con m righe ed n colonne, abbia rango m, se esistono soluzioni ammissibili per PL esistono anche soluzioni di base ammissibili; se esistono soluzioni ottime, esistono anche soluzioni di base (ammissibili) ottime.

Un'utile interpretazione del teorema fondamentale può essere ottenuta considerando le sue due tesi in ordine inverso: la seconda stabilisce che la soluzione ottima può essere ricercata limitando lo studio alle soluzioni di base e questo riduce l'analisi ad un numero finito di alternative (a priori le possibilità costituite dalla Regione Ammissibile sono infinite). La prima parte del teorema risolve a questo punto un problema di esistenza, perché una volta stabilito che la soluzione ottima è anche di base è ragionevole chiedersi che cosa garantisce l'esistenza di soluzioni di base: la prima tesi dice in effetti che queste soluzioni di base ci sono purché la Regione Ammissibile risulti non vuota.

L'ipotesi che la matrice A abbia caratteristica m merita un'osservazione. Tale proprietà significa, tra l'altro, che sono linearmente indipendenti i vettori riga di A, cioè i coefficienti delle singole equazioni che costituiscono i vincoli. In effetti, questo è tipico di un problema costruito con

una certa attenzione, nel senso che non vi sono vincoli ridondanti, che sono combinazione di altri vincoli. Se vincoli di questo tipo ve ne fossero, sarebbe di notevole interesse eliminarli, e questo può costituire un problema di non facile soluzione. D'altra parte è anche comprensibile che in un problema reale dove i vincoli possono sorgere da più settori di un'azienda o di un sistema qualsiasi, e formulati quindi da persone differenti, non appare strano che da fonti diversi vengano segnalazioni di vincoli tra loro dipendenti (se non dello stesso vincolo, più volte). Pertanto l'ipotesi semplifica la trattazione teorica ma apre un altro capitolo che è quello della costruzione di problemi senza ridondanze.

Dimostrazione. Per quanto riguarda la prima parte, si farà vedere che, se esiste una soluzione ammissibile, allora si può costruire una soluzione di base dello stesso sistema di vincoli. In effetti, sia **x** una soluzione (in generale, non di base) a componenti non negative: le componenti strettamente positive di **x** siano le prime p. Facendo ricorso alla formulazione del sistema dei vincoli utilizzando le colonne della matrice A, si potrà allora scrivere:

$$\mathbf{b} = x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + \ldots + x_p \mathbf{a}_p.$$

Se $p \leq m$ e le colonne sono linearmente indipendenti, allora la soluzione è già di base, eventualmente degenere. Se le colonne della combinazione lineare scritta sono linearmente dipendenti (sia quando $p < m$ sia con $p \geq m$), si procede costruendo con un processo iterativo una soluzione con un minor numero di componenti positive che alla fine della procedura saranno linearmente indipendenti, come segue.
Scriviamo una combinazione lineare nulla delle colonne in oggetto:

$$y_1 \mathbf{a}_1 + y_2 \mathbf{a}_2 + \ldots + y_p \mathbf{a}_p = \mathbf{0} .$$

Possiamo supporre, eventualmente dopo un cambiamento di segno, che almeno uno dei coefficienti y sia positivo. Analogamente a quanto fatto in una precedente dimostrazione, diciamo **y** il vettore n-dimensionale

$$\mathbf{y} = (y_1 \ y_2 \ \ldots \ y_p \ 0 \ 0 \ \ldots \ 0).$$

Il vettore $\mathbf{x} - \varepsilon \mathbf{y}$ risulta, per $\varepsilon \geq 0$, ma sufficientemente piccolo, ancora ammissibile: in particolare, per $\varepsilon = 0$ coincide con **x**. Se supponiamo di far crescere ε a partire da 0, le componenti del vettore $\mathbf{x} - \varepsilon \mathbf{y}$ aumentano in corrispondenza delle componenti $y_k < 0$ mentre diminuiscono in corrispondenza delle componenti del vettore **y** che risultano positive (di queste, come si è detto, ve n'è almeno una). Sia h l'indice della componente del vettore $\mathbf{x} - \varepsilon \mathbf{y}$ che si annulla per prima. Ponendo allora proprio $\varepsilon = x_k / y_k$, il vettore $\mathbf{x} - \varepsilon \mathbf{y}$ contiene (almeno) una componente positiva in meno rispetto a **x**. Si ripropone allora la questione iniziale: se le colonne corrispondenti alle componenti (rimaste) positive sono linearmente indipendenti, la soluzione è di base e il procedimento è terminato. Se invece sono ancora linearmente dipendenti, si applica nuovamente la procedura appena descritta, riducendo ulteriormente le colonne usate nella soluzione, che, in definitiva, risulteranno in un numero finito di passi linearmente indipendenti.

Per quanto riguarda la seconda parte, relativa all'ottimalità, si consideri una soluzione ottima non di base **x***. Per definizione di soluzione ottima, ciò consente di affermare che sarà anche

$$\mathbf{c}\,\mathbf{x} \leq \mathbf{c}\,\mathbf{x}^*$$

per ogni soluzione **x** ammissibile. Ricorrendo al procedimento già usato nella prima parte della dimostrazione, riduciamo il numero di componenti > 0 di **x***, introducendo i vettori del tipo $\mathbf{x}^* - \varepsilon\mathbf{y}$. Si può affermare che in ogni caso $\mathbf{cy} = 0$. Infatti, poiché $\mathbf{x}^* - \varepsilon\mathbf{y}$ è ammissibile per ε sufficientemente piccolo **sia positivo sia negativo**, se fosse $\mathbf{c}\,\mathbf{y} \neq 0$, per assurdo, **x*** non potrebbe essere soluzione ottima: pertanto il valore della funzione oggetto rimane lo stesso dopo ogni modifica della soluzione e in realtà quelle che si ottengono man mano sono tutte soluzioni ammissibili egualmente ottime, e l'ultima del procedimento è anche di base.

La prima affermazione del Teorema Fondamentale ha un'immediata interpretazione geometrica che si può formulare con il seguente enunciato:

La regione ammissibile di un problema di programmazione lineare se non è vuota contiene sempre almeno un vertice.

La seconda affermazione del teorema fondamentale, relativa all'ottimalità di un vertice, può essere dimostrata agevolmente per altra via nel caso in cui la regione ammissibile sia limitata. Infatti, in tale situazione ogni punto **x** della regione ammissibile stessa è ottenibile (generalmente in più modi) come combinazione lineare convessa dei vertici, **v**(1), **v**(2), …, **v**(n):

$$\mathbf{x} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \ldots + \alpha_n \mathbf{v}_n,$$

e il valore della funzione oggetto nel punto stesso **x** si può esprimere come combinazione lineare convessa (con gli stessi coefficienti α_k) dei valori nei vari vertici. Poiché una combinazione lineare convessa di scalari è una loro media ponderata, e pertanto non superiore al più grande né inferiore al più piccolo dei valori di cui si effettua la media, ne segue che il valore in **x** non può risultare migliore rispetto a quello in ciascuno dei vertici.

2.1.9 Il metodo del simplesso: generalità.

Il metodo del simplesso è la tecnica risolutiva più nota e maggiormente adoperata per i problemi di programmazione lineare. Sono state messe a punto anche altre procedure, ma occorre dire che esse, se è vero che sono risultate di notevole interesse dal punto di vista teorico, avendo consentito di comprendere più a fondo la natura e la difficoltà dei problemi di PL, sinora non hanno avuto altrettanto successo dal punto di vista applicativo.

Il metodo del simplesso è codificato in numerosi pacchetti software che ne consentono un utilizzo anche a chi ha una conoscenza parziale dei suoi fondamenti teorici. In questo paragrafo il metodo sarà illustrato nelle sue linee generali, e ciò potrà essere sufficiente per chi debba farne uso in qualche situazione concreta. Nei paragrafi successivi esso sarà invece illustrato nei dettagli che, in definitiva, non sono altro che la formalizzazione ed il completamento delle procedure già viste nell'esempio (didattico) di calcolo relativo alla produzione delle biciclette.

Il simplesso esplora soltanto (alcune) soluzioni di base ammissibili per il problema di PL allo studio, secondo una successione di operazioni (**iterazioni**) che, **ipotizzando una funzione obiettivo di massimo**, si possono schematizzare come segue:

1) ricerca di una soluzione ammissibile di base iniziale: se la soluzione esiste, andare al passo 2); se non esiste, STOP;
2) verifica dell'ottimalità della soluzione di base ottenuta: se la soluzione è ottima STOP; se non è ottima, ma si costata che la funzione oggetto nella regione ammissibile è illimitata superiormente; STOP; in caso contrario, andare al passo 3;
3) costruzione di una nuova soluzione di base ammissibile per la quale il valore della funzione oggetto non è inferiore al valore della soluzione precedente; ripetere il passo 2.

L'alternarsi dei passi 2 e 3 porta ad effettuare le cosiddette 'iterazioni del simplesso', il cui numero determina evidentemente il tempo di esecuzione dell'algoritmo. A questo proposito va detto che, sebbene siano stati costruiti esempi nei quali la tecnica esamina tutte le soluzioni di base ammissibili (e queste possono essere in numero così elevato da non essere gestibile in tempi realistici), nei casi pratici le soluzioni considerate sono relativamente poche (dell'ordine di grandezza del numero delle colonne).

Da un punto di vista geometrico, si tratta di individuare (passo 1) un primo vertice della regione ammissibile e quindi, se non si tratta della soluzione ottimale (verifica al passo 2), determinare un vertice adiacente che dia un valore migliore per la funzione oggetto (passo 3).

Nei paragrafi che seguono saranno poste le basi teoriche dell'algoritmo analizzando, nell'ordine, le seguenti operazioni:

- **costruzione di una nuova soluzione di base** a partire da una soluzione data, cambiando la base di riferimento mediante lo scambio di due vettori, un vettore che lascia la base ed uno che lo sostituisce (operazione di **pivot**);

- **mantenimento dell'ammissibilità**: regole da seguire affinché la nuova soluzione di base ottenuta sia ammissibile, dato che lo è la soluzione di base di partenza;

- **verifica dell'ottimalità** della soluzione ottenuta (regole di arresto) e scelta eventuale di una nuova variabile da far entrare in base;

- **costruzione di una soluzione di base iniziale**.

Quest'ordine di esposizione non coincide con quello delle operazioni pratiche, ma è obbligato perché la costruzione di una soluzione iniziale si effettua con un procedimento che richiede, dal punto di vista teorico, la conoscenza della procedura di passaggio da una soluzione ammissibile di base ad un'altra.

Si tenga presente che in letteratura spesso l'algoritmo del simplesso è formulato in termini di funzione oggetto di minimo: i cambiamenti da adottare nelle tabelle sono inessenziali per la comprensione della procedura.

2.1.10 L'operazione di pivot.

Con l'operazione di 'cardine' o 'pivot' si ottiene una nuova soluzione di base a partire da una soluzione di base assegnata mediante lo scambio tra due vettori, uno che è escluso dalla vecchia base B ed un altro che lo sostituisce, ottenendo una nuova base che diremo B'.

Nel metodo del simplesso, ad ogni iterazione la soluzione di base è facilmente 'leggibile' in quanto il sistema si trova nella 'forma canonica'.

Si ha la **forma canonica** quando **le variabili di base compaiono ciascuna in una sola delle equazioni e con coefficiente unitario** e quindi, assumendo che le variabili fuori base assumano valore nullo, **il valore di ciascuna componente di base della soluzione x coincide con il termine noto della equazione in cui la variabile stessa compare**.

Un sistema in forma canonica, a meno di permutazioni sulle colonne, (e quindi supponendo, per comodità, che le componenti in base siano le prime m) si presenta come segue:

$$
\begin{aligned}
x_1 \qquad\qquad\qquad\qquad & + a_{1,m+1}\, x_{m+1} + \ldots + a_{1,n}\, x_n = b_1, \\
x_2 \qquad\qquad\qquad\qquad & + a_{2,m+1}\, x_{m+1} + \ldots + a_{2,n}\, x_n = b_2, \\
\ldots\ldots\ldots\ldots\ldots\ldots & \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
x_m & + a_{m,m+1}\, x_{m+1} + \ldots + a_{m,n}\, x_n = b_m.
\end{aligned}
$$

dove è evidente come la sottomatrice di base B di riferimento sia data dalla matrice identica: B = I.

Si osservi che in questa formulazione, il vettore **b** è rappresentabile con una combinazione lineare delle prime m colonne di A, (cioè di B), nella quale i combinatori, proprio perché B è la matrice identica, coincidono con le componenti di **b**.

In generale, si avrà che, se le prime m colonne sono linearmente indipendenti ma non costituiscono una matrice identica, **b** è sempre ottenibile come loro combinazione lineare, ma per individuare i coefficienti occorre porre a zero il valore delle n-m variabili fuori base e risolvere il sistema che rimane.

Le trasformazioni che il sistema subisce per essere portato nella forma canonica (che permette una visione immediata della soluzione), equivalgono ad un insieme di rotazioni e cambiamenti di unità di misura degli assi coordinati (cioè a variazioni nel sistema di riferimento). Tali trasformazioni comportano a loro volta cambiamenti nelle coordinate del vettore **b**. Tuttavia, se si considerano sia i vettori colonna di A che il vettore **b** in maniera astratta, nella nuova rappresentazione (che porta le prime m colonne di A nei vettori unità dei vari assi coordinati), sono le stesse componenti di **b** a indicare quale sia la combinazione lineare che **in una qualsiasi trasformazione lineare dello spazio vettoriale a m dimensioni** fornisce **b** in funzione delle prime m colonne di A.

Si voglia ora effettuare un cambiamento di base ed esprimere **b** non più come combinazione lineare delle prime m colonne, (cioè di B), bensì, eliminata dalla base una di queste colonne, che poniamo sia la p-esima, $\mathbf{a}_p$, sostituirla con il vettore colonna q-esimo, $\mathbf{a}_q$.

In questo modo si vuole ottenere **b** come combinazione lineare del nuovo insieme di vettori colonna

$$\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{p-1}, \mathbf{a}_{p+1}, \ldots, \mathbf{a}_{m-1}, \mathbf{a}_m, \mathbf{a}_q.$$

E' facile costatare che:

questo nuovo insieme di vettori costituisce una nuova base B' se e solo se l'elemento a_{pq} della matrice A è $\neq 0$.

Infatti, è agevole vedere come in tale ipotesi la caratteristica della matrice B' formata da queste colonne rimane m e, d'altra parte, una soluzione di un qualsiasi sistema del tipo $B'\mathbf{x}_{B'} = \mathbf{b}$ si può ottenere facilmente ricavando dapprima $x_q = b_p/a_{pq}$ e quindi, calcolato il vettore

$$\mathbf{b'} = \mathbf{b} - x_q\, \mathbf{a}_q$$

(che ha la p-esima componente eguale a zero), porre

$$x_1 = b'_1, \quad x_2 = b'_2, \ldots, x_{p-1} = b'_{p-1}, \quad x_{p+1} = b'_{p+1}, \ldots, x_m = b'_m.$$

Per mantenere la forma canonica nel sistema dei vincoli, si tratta di effettuare delle trasformazioni che conducano alla eliminazione della variabile x_q da ogni equazione che non sia la p-esima, mentre in questa stessa equazione il coefficiente deve essere 1. A tale scopo si procede come segue:

- si dividono ambo i membri della equazione p-esima per a_{pq}, in modo tale che i coefficienti della riga p-esima diventano:

$$0 \quad 0 \quad \ldots. \quad 1 / a_{pq} \quad \ldots. \; 0 \quad a_{p,m+1} / a_{pq} \quad \ldots. \quad a_{pq} / a_{pq} \quad \ldots. \; a_{pn} / a_{pq}$$

  a primo membro, mentre il secondo membro diviene:

$$b_p / a_{pq};$$

- si somma a membro a membro l'equazione p-esima (ottenuta dalla trasformazione appena descritta) a ciascuna delle altre moltiplicandone ambo i membri per i coefficienti, rispettivamente, $-a_{1q}$, $-a_{2q}$, …, $-a_{mq}$, in modo tale che sulla colonna q-esima i coefficienti siano tutti nulli (tranne il p-esimo, che ora è 1).

Se si indicano con a'_{ij} e b'_j i nuovi coefficienti della matrice A, tra essi ed i coefficienti iniziali valgono le relazioni:

$$a'_{ij} = a_{ij} - a_{iq} * a_{pj} / a_{pq} \; ; \quad b'_i = b_i - a_{iq} * b_p / a_{pq}.$$

Infatti, dopo aver trasformato la riga p-esima, la matrice dei coefficienti ed i termini noti assumono la seguente forma:

$$\begin{array}{ccccccccccccc}
1 & 0 & 0 & \ldots & 0 & \ldots & 0 & a_{1,m+1} & \ldots & a_{1q} & \ldots & a_{1n} & b_1 \\
0 & 1 & 0 & \ldots & 0 & \ldots & 0 & a_{2,m+1} & \ldots & a_{2q} & \ldots & a_{2n} & b_2 \\
\multicolumn{13}{c}{\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots} \\
0 & 0 & 0 & \ldots & 1/a_{pq} & \ldots & 0 & a_{p,m+1} / a_{pq} & \ldots & 1 & \ldots & a_{pn} / a_{pq} & b_p / a_{pq} \\
\multicolumn{13}{c}{\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots} \\
0 & 0 & 0 & \ldots & 0 & \ldots & 1 & a_{m,m+1} & \ldots & a_{mq} & \ldots & a_{mn} & b_m
\end{array}$$

Successivamente, occorre portare a zero tutti i restanti elementi della colonna q-esima. Le trasformazioni che subiscono gli altri elementi della matrice e le componenti del vettore dei termini noti possono essere illustrate mettendo in evidenza due righe, la riga p-esima ed un'altra generica riga, che supporremo sia la i-esima e su queste gli elementi indicati nella figura che segue:

↑ → … a_{iq} ….. a_{ij} ….. b_i

$-a_{iq}$

$\leftarrow \quad \ldots \quad 1 = a_{pq} / a_{pq} \quad \ldots\ldots \quad a_{pj} / a_{pq} \quad \ldots\ldots \quad b_p / a_{pq}$

Come si vede, occorre moltiplicare la riga i-esima per il coefficiente $-a_{iq}$ e sommarla membro a membro alla riga i-esima e così si ottiene:

$\ldots \quad 0 \quad \ldots\ldots \quad a_{ij} - a_{iq} * a_{pj} / a_{pq} \quad \ldots\ldots \quad b_i - a_{iq} * b_p / a_{pq}$

$\ldots \quad 1 = a_{pq} / a_{pq} \quad \ldots\ldots \quad a_{pj} / a_{pq} \quad \ldots\ldots \quad b_p / a_{pq}$

Operativamente, ipotizzando una soluzione 'manuale', il metodo del simplesso è attuato lavorando su una successione di tabelle, contenenti solo i coefficienti della matrice A e le componenti del vettore **b** (come si vedrà in seguito, a queste è aggiunta un'ulteriore riga contenente i coefficienti della funzione oggetto ed il valore corrente della stessa). Un'operazione di cardine può essere condotta su questa tabella sostituendo innanzi tutto alla riga (p-esima) del pivot la nuova riga in cui il pivot è divenuto 1 e quindi sommando questa riga alle altre, moltiplicata per il coefficiente opportuno.

Si consideri il seguente esempio elementare (nel quale le colonne in base sono individuate da un asterisco):

*	*	*				
1	0	0	4	2	-1	7
0	1	0	3	-1	0	6
0	0	1	1	5	-8	4

In questa tabella, che evidenzia la soluzione di base data dal vettore

$$(7 \;\; 6 \;\; 4 \;\; 0 \;\; 0 \;\; 0),$$

possono essere pivot tutti gli elementi delle colonne 4, 5 e 6, tranne l'elemento a_{26} che è nullo. Se si decide di cambiare base, sostituendo ad esempio la seconda colonna con la quarta, il pivot è l'elemento $a_{24} = 3$. Facendo diventare 1 questo coefficiente, la tabella diventa:

1	0	0	4	2	-1	7
0	1/3	0	1*	-1/3	0	2
0	0	1	1	5	-8	4

(In questa tabella il pivot è stato evidenziato con un asterisco).

Occorre adesso portare a zero gli altri elementi della quarta colonna. Per ottenere questo risultato, basterà sommare la seconda riga alla prima, dopo averla moltiplicata per –4 e successivamente sommarla alla terza, dopo averla moltiplicata per –1. Il risultato finale è il seguente:

		*		*			
1	-4/3	0	0	10/3	-1		-1
0	1/3	0	1	-1/3	0		2
0	-1/3	1	0	16/3	-8		2

e la nuova soluzione di base (corrispondente alla nuova base, formata dalle colonne 1, 2 e 4) è il vettore

$$(-1 \quad 0 \quad 2 \quad 2 \quad 0 \quad 0).$$

Si può osservare come la soluzione di base iniziale è una soluzione ammissibile (con tutte le componenti non negative), mentre la nuova soluzione non lo è più. Il mantenimento dell'ammissibilità delle soluzioni di base è ottenuto scegliendo il pivot sulla base di regole che saranno motivate nel prossimo paragrafo.

2.1.11 Il mantenimento dell'ammissibilità.

Se la soluzione di base relativa ad una tabella del simplesso è ammissibile, ciò significa che le componenti del vettore **b** nella rappresentazione attuale sono non negative. Se poi la soluzione è non degenere, le componenti stesse sono tutte strettamente positive.

Si noti che le componenti del vettore soluzione, se non sono fuori base e quindi nulle per definizione, sono, salvo permutazioni, le componenti stesse del vettore **b**.

Per mantenere l'ammissibilità della nuova soluzione di base, occorre che le componenti del vettore dei termini noti siano ancora non negative dopo l'operazione di cardine. In particolare:

- dovendo rimanere non negativo l'elemento b'_p, occorre che $a_{pq} > 0$;
- per le componenti restanti, occorre che sia

$$b'_i = b_i - a_{iq} * b_p / a_{pq} \geq 0,$$

cioè

$$b_i \geq a_{iq} * b_p / a_{pq}.$$

In quest'ultima diseguaglianza, una volta rispettata la condizione che il pivot sia non negativo, si può affermare che il segno del secondo membro dipende dal termine a_{iq}: se tale coefficiente è **negativo o nullo**, certamente la nuova componente b'_i è non negativa e anzi, se risulta $a_{iq} < 0$ e $b_p > 0$ si ottiene un valore superiore al precedente (cioè, $b'_i > b_i$).
Se a_{iq} è una quantità positiva, allora la diseguaglianza è equivalente all'altra

$$b_i / a_{iq} \geq b_p / a_{pq}.$$

Questa relazione implica che il rapporto tra le componenti del vettore **b** e le componenti della colonna del pivot (con lo stesso indice di riga) deve essere minimo in corrispondenza del pivot. Di qui la regola di scelta del pivot che, riassumendo tutte le considerazioni fin qui svolte, si può esprimere come segue:

per mantenere l'ammissibilità della nuova soluzione di base, il pivot deve essere strettamente positivo ed essere situato sulla riga per la quale il rapporto b_i/a_{iq}, calcolato per i valori $a_{iq} > 0$ è minimo:

$$\mathbf{b_p/a_{pq} = \min \{ b_i / a_{iq} \mid a_{iq} > 0 \}.}$$

Si osservi che alla stessa conclusione si poteva giungere sulla base di un ragionamento che ricorda quello svolto nella dimostrazione del **Teorema Fondamentale** della Programmazione Lineare. Infatti, indicando con ε il valore della variabile x_q che si vuole fare entrare in base (tale valore inizialmente è 0), e ipotizzando di incrementare il valore stesso, il rispetto dei vincoli impone in generale che il valore delle altre componenti della soluzione cambino. Se si mantengono nulle tutte le variabili che rimangono fuori base, le altre variabili attualmente in base potranno restare costanti, aumentare o diminuire di valore sulla base della relazione (vincolo i-esimo):

$$x_i + \varepsilon\, a_{iq} = b_i.$$

Evidentemente, per $\varepsilon=0$, $x_i = b_i$, e si ha la vecchia soluzione di base; per $\varepsilon > 0$, si ha un nuovo valore

$$x'_i = b_i - \varepsilon\, a_{iq}.$$

Questo sarà certamente positivo se $a_{iq} < 0$ (e ciò non pone problemi di ammissibilità); sarà invece una quantità che decresce al crescere di ε (e quindi potenzialmente, anche negativa) se $a_{iq} > 0$. Si può pensare allora di aumentare progressivamente il valore ε della variabile x_q ma solo sino al punto in cui una delle altre variabili che diminuiscono raggiunge il valore 0 (un ulteriore incremento di x_q non è possibile perché questa stessa variabile diventerebbe negativa). L'incremento massimo possibile per x_q, cioè il suo nuovo valore, resta allora evidentemente determinato dal più piccolo tra i rapporti b_i /a_{iq}, calcolati per $a_{iq} > 0$. In corrispondenza a questo **rapporto minimo** una delle variabili che prima erano in base è diventata 0, e quindi si può considerare uscita dalla base, mentre le altre mantengono ancora valori positivi (con l'eccezione di più variabili che per prime si annullano contemporaneamente).

Facendo riferimento all'esempio del paragrafo precedente, e cioè alla tabella

*	*	*				
1	0	0	4	2	-1	7
0	1	0	3	-1	0	6
0	0	1	1	5	-8	4

se si vuole introdurre in base il quinto vettore colonna e contemporaneamente mantenere l'ammissibilità, il pivot non può essere l'elemento –1 (il pivot deve essere positivo) e va scelto tra gli altri due elementi, 2 e 5, calcolando i rapporti:

$$7/2 = 3.5 \qquad \text{e} \qquad 4/5 = 0.8.$$

Il più piccolo di questi determina il pivot, che sarà quindi l'elemento '5'. La tabella successiva si ottiene allora nelle due fasi:

$$\begin{array}{cccccc|c} 1 & 0 & 0 & 4 & 2 & -1 & 7 \\ 0 & 1 & 0 & 3 & -1 & 0 & 6 \\ 0 & 0 & 1/5 & 1/5 & 1 & -8/5 & 4/5 \end{array}$$

(divisione di tutti gli elementi della terza riga per 5) e successivamente

$$\begin{array}{cccccc|c} * & * & & & * & & \\ 1 & 0 & -2/5 & 18/5 & 0 & 11/5 & 27/5 \\ 0 & 1 & 1/5 & 16/5 & 0 & -8/5 & 34/5 \\ 0 & 0 & 1/5 & 1/5 & 1 & -8/5 & 4/5 \end{array}$$

da cui si può leggere la nuova soluzione di base (ancora ammissibile) data dal vettore

$$(27/5 \quad 34/5 \quad 0 \quad 0 \quad 4/5 \quad 0).$$

2.1.12 Verifica della ottimalità di una soluzione.

Il valore assunto dalla funzione oggetto in corrispondenza di una soluzione di base si può ottenere in modo ovvio per sostituzione nella sua espressione dei valori delle componenti della soluzione stessa. Questo però non consente di comprendere se la soluzione ottenuta è ottima oppure se è migliorabile (e neppure se la funzione oggetto è superiormente illimitata).

Per avere queste informazioni, il metodo del simplesso prevede che ad ogni iterazione anche la funzione oggetto sia trasformata: in particolare, si fa in modo che essa contenga solo le variabili fuori base e saranno i coefficienti di queste ultime ad indicare se si è giunti o no alla soluzione ottima.

A tale scopo, consideriamo nuovamente il sistema in forma canonica:

$$\begin{aligned} x_1 \qquad\qquad & + a_{1,m+1}\, x_{m+1} + \ldots + a_{1,n}\, x_n = b_1, \\ x_2 \qquad\qquad & + a_{2,m+1}\, x_{m+1} + \ldots + a_{2,n}\, x_n = b_2, \\ & \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_m & + a_{m,m+1}\, x_{m+1} + \ldots + a_{m,n}\, x_n = b_m. \end{aligned}$$

e ricaviamo da ogni equazione la variabile in base che vi compare esprimendola in funzione del termine noto e delle restanti variabili fuori base:

$$\begin{aligned} x_1 &= b_1 - a_{1,m+1}\, x_{m+1} - \ldots - a_{1,n}\, x_n, \\ x_2 &= b_2 - a_{2,m+1}\, x_{m+1} - \ldots - a_{2,n}\, x_n, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_m &= b_m - a_{m,m+1}\, x_{m+1} - \ldots - a_{m,n}\, x_n. \end{aligned}$$

Diremo d'ora in poi **x'** la soluzione di base (corrente) con tali componenti in base.

Consideriamo ora la funzione oggetto

$$z = \mathbf{c}\, \mathbf{x'} =$$

$$= c_1 x_1 + c_2 x_2 + \ldots + c_m x_m + c_{m+1} x_{m+1} + \ldots + c_n x_n.$$

Sostituendo alle variabili in base le loro espressioni ricavate in precedenza, si ottiene la seguente formula:

$$\begin{aligned} z = {} & c_1 b_1 + c_2 b_2 + \ldots + c_m b_m + \\ & + x_{m+1} \{c_{m+1} - c_1 a_{1,m+1} - c_2 a_{2,m+1} - \ldots - c_m a_{m,m+1}\} + \\ & + x_{m+2} \{c_{m+2} - c_1 a_{1,m+2} - c_2 a_{2,m+2} - \ldots - c_m a_{m,m+2}\} + \\ & + \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots + \\ & + x_n \{c_n - c_1 a_{1,n} - c_2 a_{2,n} - \ldots - c_m a_{m,n}\}. \end{aligned}$$

In questa compaiono il valore corrente della funzione oggetto (dato dalla $\Sigma\, c_i b_i$) e solo le ultime n-m variabili di decisione, cioè le variabili fuori base. I coefficienti di queste ultime prendono il nome di **coefficienti di guadagno ridotto** e costituiscono l'elemento di valutazione della soluzione di base **x'** (in caso di un problema di minimo si parlerà di **coefficienti di costo ridotto**).

Per illustrare il significato di questi coefficienti, possiamo riprendere in considerazione l'esempio della determinazione di una politica di produzione di guadagno massimo in presenza di risorse limitate. Il problema, nell'esempio, è stato formulato nella forma canonica di massimo, che può essere trasformata nella forma standard introducendo variabili scarto, corrispondenti in pratica ai quantitativi delle varie risorse che non sono utilizzati. La trasformazione poi del sistema della forma standard nel sistema in forma canonica richiede una procedura che sarà illustrata più avanti.

Supponiamo per semplicità che vi siano m risorse e più di m possibili prodotti e di trovarci di fronte ad una soluzione di base

$$x_1 = b_1,\ x_2 = b_2,\ \ldots,\ x_m = b_m,\ x_{m+1} = 0,\ \ldots,\ x_n = 0$$

che contempla la produzione dei primi m articoli, mentre è nulla quella dei rimanenti: con questa formulazione è come se ognuno dei primi m articoli richiedesse una sola risorsa e la consumasse tutta, per cui sono nulle sia le variabili corrispondenti ai prodotti successivi all'm-esimo sia tutte le variabili slack. Per produrre i successivi articoli (da P_{m+1} a P_n), invece, supponiamo che in generale occorrano tutte le risorse.

Con la soluzione considerata, si ha evidentemente un guadagno dato da $\Sigma\, c_j b_j$, dove la sommatoria è estesa appunto ai primi m prodotti. Supponiamo adesso di cambiare la politica produttiva e introdurre un nuovo articolo, ad es. P_{m+1}. Per valutare il guadagno unitario netto che ne consegue, occorre considerare che tale nuova produzione, proprio per la scarsità delle risorse, andrà almeno parzialmente a scapito della produzione dei primi m articoli. Pertanto non sarebbe corretto ritenere che ogni unità prodotta del nuovo articolo P_{m+1} implichi un guadagno dato dal suo coefficiente originale c_{m+1}: occorrerà, come si è detto, tenere conto anche della perdita che si subisce per la diminuzione nella produzione restante. Questa perdita monetaria si può calcolare come segue. Per ogni unità che si produce di P_{m+1}:

- la produzione di P_1 diminuisce di $a_{1,m+1}$, e si perde un guadagno dato da $c_1 a_{1,m+1}$;
- la produzione di P_2 diminuisce di $a_{2,m+1}$, e si perde un guadagno dato da $c_2 a_{2,m+1}$; e così via.

Questo giustifica la formulazione del guadagno netto stesso come

$$\{ c_{m+1} - c_1 a_{1,m+1} - c_2 a_{2,m+1} - \ldots - c_m a_{m,m+1}\}$$

(che è proprio il coefficiente di x_{m+1} nell'espressione di z) e la sua denominazione. Ora può capitare che quest'espressione sia negativa: in tal caso il guadagno per la produzione del nuovo prodotto non è sufficiente a coprire i minori introiti causati dalla riduzione della produzione degli altri articoli. Solo se la stessa espressione è positiva converrà cambiare piano di produzione. Se l'espressione risulta 0 allora il cambio di produzione è indifferente.

I coefficienti di guadagno ridotto possono risultare positivi, negativi o nulli. Si possono a questo punto enunciare le seguenti proprietà, che consentono di decidere se il procedimento del simplesso è terminato con la determinazione della soluzione di base corrente **x'** o se necessitano ulteriori iterazioni.

Proprietà 1. Se il coefficiente di guadagno ridotto di (almeno) una delle variabili (fuori base) è positivo, la soluzione corrente x' non è ottima e può essere migliorata.

Tuttavia, non è detto che la variabile con coefficiente di guadagno ridotto positivo possa entrare in base: occorre che sulla colonna corrispondente vi sia (almeno) un elemento positivo che possa fungere da pivot. Se tale elemento non esiste, il procedimento è terminato. Vale infatti la

Proprietà 2. Se in corrispondenza di una variabile con coefficiente di guadagno ridotto positivo vi è un vettore colonna della matrice A dei coefficienti dei vincoli con elementi tutti ≤ 0, allora il PL non presenta ottimo finito, nel senso che la funzione oggetto, nella Regione Ammissibile, è illimitata superiormente.

In effetti, in tale situazione la variabile in oggetto può assumere valori arbitrariamente grandi e, parallelamente, le altre variabili già in base, anziché dover diminuire, assumeranno a loro volta valori sempre più elevati, in soluzioni (non di base) che consentono di far assumere alla funzione oggetto valori grandi a piacere. Con una certa improprietà di linguaggio, si può affermare che l'ottimo è 'infinito'. La proprietà 2 è quindi una regola d'arresto. Più frequentemente, nei problemi ben formulati, la regola d'arresto è data dalla seguente proprietà 3.

Proprietà 3. Se tutti i coefficienti di guadagno ridotto sono negativi, allora la soluzione di base ultima individuata x' è soluzione ottima.

In questo caso, infatti, non è conveniente cambiare base: l'introduzione di una nuova variabile in base in sostituzione di una qualsiasi delle attuali comporta una diminuzione della funzione oggetto.
Ultima eventualità d'arresto è la seguente:

Proprietà 4. Se i coefficienti di guadagno ridotto sono non positivi ed almeno uno di essi è nullo, la soluzione ottenuta x' è ottima, ma vi sono infinite altre soluzioni ottime. Tra queste ve n'è certamente (almeno) un'altra ancora di base x" se (e solo se) è possibile effettuare un'ulteriore operazione di pivot sulla colonna della variabile con coefficiente nullo: in tal caso, sono soluzioni ottime anche tutte le combinazioni lineari convesse delle soluzioni x' e x". Se invece non è possibile effettuare l'operazione di pivot, esistono altre infinite soluzioni ottime situate lungo una semiretta uscente da x'.

Dalle proprietà 1 e 2 considerate nel complesso si può ottenere la regola che stabilisce quando si tratta di proseguire con le iterazioni, individuando una nuova soluzione ammissibile.

Proprietà 5. Se vi sono coefficienti di guadagno ridotto positivi e per ciascuna delle colonne corrispondenti è possibile scegliere un elemento come pivot, si può cercare una nuova soluzione di base ammissibile che risulti migliore (o non peggiore) della soluzione corrente.

Il problema a questo punto consiste nella scelta della colonna da fare entrare in base (tra quelle, ovviamente, con coefficiente di guadagno ridotto positivo). A tale proposito, è evidente che si sarebbe tentati di scegliere quella colonna (e la corrispondente variabile) che comportano il maggior incremento per la funzione oggetto, ma determinare quale sia la scelta giusta in questa direzione comporta una certa mole di calcoli che finirebbe per appesantire tutto il procedimento. Una regola empirica è quella di scegliere la variabile con il coefficiente di guadagno ridotto positivo massimo, ma questo non garantisce che poi sia massimo l'incremento di z.

Nelle tabelle del simplesso la funzione oggetto compare in un'ulteriore riga pensando la funzione stessa come un'eguaglianza e cioè:

$$z = c_1 x_1 + c_2 x_2 + \ldots + c_n x_n \qquad (*)$$

che poi, portando tutto a primo membro, diviene:

$$z - c_1 x_1 - c_2 x_2 - \ldots - c_n x_n = 0. \qquad (**)$$

In realtà si trascura di indicare la variabile z, mentre nell'ultima riga delle tabelle compaiono, in corrispondenza delle varie variabili, i coefficienti $-c_j$, cioè i coefficienti di guadagno cambiati di segno. La sostituzione del valore delle variabili in base mediante la loro espressione in funzione di quelle fuori base è effettuata con le stesse regole usate nell'operazione di cardine.

L'unico elemento da tenere poi presente è il fatto che, a causa del passaggio da (*) a (**), anche i coefficienti di guadagno ridotto risultano cambiati di segno e quindi saranno i valori negativi dell'ultima riga quelli che indicano le potenziali variabili da fare entrare in base.

Nel paragrafo successivo quanto sopra detto sarà chiarito con alcuni esempi.

2.1.13 La funzione oggetto nel simplesso e le regole d'arresto.

2.1.13.1 Introduzione della funzione oggetto in una tabella di simplesso con il sistema in forma standard.

Supponiamo di avere una tabella di simplesso come la seguente:

* * *

1	0	0	3	-1	4
0	1	0	2	5	3
0	0	1	1	6	5

che corrisponde alla soluzione di base (ammissibile) (4 3 5 0 0). Supponiamo poi che la funzione oggetto sia data dall'espressione

$$z = 4\,x_1 - 2\,x_2 + x_3 + 10\,x_4 - 2\,x_5.$$

In corrispondenza della soluzione corrente il valore della funzione oggetto è

$$z = 4*4 - 2*3 + 5 = 15.$$

Apparentemente converrebbe far entrare in base la variabile x_4, perché ha coefficiente 10, e non x_5, ma prima di stabilirlo occorre individuare i coefficienti di guadagno ridotti.
A tale scopo si pone

$$z - 4\,x_1 + 2\,x_2 - x_3 - 10\,x_4 + 2\,x_5 = 0$$

e si procede a sostituire alle variabili in base x_1, x_2 e x_3 le loro espressioni ricavate dai tre vincoli. In pratica, si aggiunge un'ulteriore riga alla tabella come segue:

*	*	*			
1	0	0	3	-1	4
0	1	0	2	5	3
0	0	1	1	6	5
-4	2	-1	-10	2	0

e si azzerano i coefficienti dell'ultima riga nelle colonne in base (si somma all'ultima riga la prima, moltiplicata per 4; poi la seconda riga, moltiplicata per –2 e infine la terza):

*	*	*			
1	0	0	3	-1	4
0	1	0	2	5	3
0	0	1	1	6	5
0	0	0	-1	-6	15

In questa tabella si possono leggere i coefficienti di guadagno ridotto (cambiati di segno) ed anche il valore corrente della funzione oggetto, z = 15, nella posizione in basso a destra.

In effetti, si costata come la variabile x_4 abbia un coefficiente di guadagno ridotto molto più modesto di quanto non sia il coefficiente iniziale, mentre la

variabile x_5, che pure ha un guadagno unitario originario negativo, può portare ad un miglioramento della funzione oggetto (si osservi che x_5 è sinergica con x_1, perché anziché competere per l'utilizzo della prima risorsa ne aumenta la disponibilità per x_1 stessa!). Comunque la soluzione corrente non è ottima e occorre proseguire con le iterazioni del simplesso (**proprietà 1** e **proprietà 5**).

Sulla base della tabella si possono fare entrare in base sia x_4 sia x_5.
Nei due casi, rispettivamente, per individuare il pivot occorre calcolare:
- (per far entrare x_4) min {4/3, 3/2, 5/1 } = 4/3, perciò il pivot sarebbe l'elemento 3;
- (per far entrare x_5) min {3/5, 5/6 } = 3/5, per cui il pivot sarebbe l'elemento 6.

In ogni caso l'ultima riga va trasformata in modo tale che in corrispondenza delle colonne in base anche in questa compaiano tutti elementi eguali a zero.

2.1.13.2 Regole d'arresto.

Supponiamo di avere la seguente tabella:

*	*	*			
1	0	0	3	0	8
0	1	0	2	-5	3
0	0	1	1	-2	1
0	0	0	7	-2	18

La variabile x_5 ha un coefficiente di guadagno ridotto di +2, ma sulla colonna corrispondente vi sono solo elementi negativi o nulli: non è possibile effettuare un'operazione di pivot e non esiste una soluzione ottima in quanto la funzione oggetto nella Regione Ammissibile è superiormente illimitata ("z* = +∞": **proprietà 2**)..

Supponiamo ora, invece, di avere la seguente tabella:

*	*	*			
1	0	0	3	-3	4
0	1	0	-4	7	3
0	0	1	1	5	5
0	0	0	1/2	8	15

In questo caso la soluzione di base corrente è ottima (i coefficienti di guadagno ridotto sono tutti negativi: **proprietà 3**).

Infine si consideri la tabella:

*	*	*

1	0	0	3	-1	3
0	1	0	2	5	3
0	0	1	1	6	5
0	0	0	0	1/4	15

In questo caso vi sono infinite soluzioni ottime (**proprietà 4**). Infatti, è possibile effettuare un'ulteriore operazione di pivot sulla quarta colonna (il pivot è l'elemento 3) ma si può costatare che nella nuova tabella non cambia il valore della funzione oggetto perché l'ultima riga non richiede cambiamenti:

*	*	*			
1/3	0	0	1	-1/3	1
-2/3	1	0	0	17/3	1
-1/3	0	1	0	19/3	5
0	0	0	0	1/4	15

La soluzione di base ammissibile **x'** = (3 3 5 0 0) è ottima, ma lo è anche il vettore **x"** = (0 1 5 1 0) e sono ottime tutte le combinazioni lineari convesse del tipo $\mathbf{w} = \alpha\,\mathbf{x'} + (1-\alpha)\,\mathbf{x"}$.

Ad es., per α=0.5 si ha la soluzione (3/2 2 5 1/2 0).

2.1.14 Costruzione di una soluzione di base iniziale.

In un problema in forma standard, in generale, non è immediatamente leggibile una soluzione di base e nemmeno è chiaro **se** esistono soluzioni ammissibili. Per risolvere questo problema preliminare esistono (almeno) due tecniche che richiedono in ogni modo la conoscenza dell'operazione di cardine: il metodo delle due fasi ed il procedimento di penalizzazione. E' viceversa chiaramente non proponibile una procedura per tentativi di triangolarizzazione del sistema perché questa può dar luogo a molte soluzioni di base non ammissibili, prima di individuarne una di accettabile (ammesso che esista!).

a) Tecnica delle due fasi.

Prevede la risoluzione di un problema ausiliario. Dato il problema in forma standard

$$\begin{aligned} & \max\ \mathbf{c}\,\mathbf{x} \\ \text{s.t.}\quad & A\,\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0}, \end{aligned} \qquad \text{(PL)}$$

si introduce un m-vettore **y** di variabili **ausiliarie** (o **artificiali**) e si risolve il problema nel vettore incognito a n+m componenti (**x**, **y**):

$$\begin{aligned} &\min\ y_1 + y_2 + \ldots + y_m \\ \text{s.t.}\quad &A\,\mathbf{x} + \mathbf{y} = \mathbf{b}, \qquad\qquad (PL') \\ &\mathbf{x}, \mathbf{y}, \geq \mathbf{0}. \end{aligned}$$

Il problema PL' ha certamente Regione Ammissibile non vuota, perché una soluzione è data dal vettore (**0**, **b**) e presenta certamente un minimo finito perché le variabili y_j sono non negative. Tale minimo sarà quindi ≥ 0. E' agevole dimostrare che PL ha soluzioni ammissibili se e solo se la soluzione ottima di PL' è il vettore nullo (e quindi il valore ottimo della funzione oggetto in quest'ultimo è $z^* = 0$).

Infatti, se il PL ha soluzioni ammissibili (e allora, per il Teorema Fondamentale della PL, ha anche soluzioni **di base** ammissibili), esiste un vettore **x** tale che A**x**=**b**, e quindi il vettore (**x**, **0**) è soluzione ammissibile per il problema PL'. Anzi, si tratta di una soluzione ottima, poiché in corrispondenza di essa il valore della funzione oggetto è 0, e la stessa, come già detto, non può assumere valori < 0, poiché sono non negative le componenti del vettore **y**.

Viceversa, se non esistono soluzioni ammissibili per PL, il minimo della f. oggetto di PL' non può essere 0: se lo fosse, per assurdo, il vettore **y** avrebbe necessariamente tutte le componenti nulle, ma allora essendo A**x** + **y** = A**x** + **0** = **b** esisterebbe anche un vettore **x** tale che A**x** = **b**, contro l'ipotesi che PL sia non ammissibile.

Dal punto di vista operativo, una volta impostato il problema PL', (**prima fase**) lo si risolve con il metodo del simplesso a partire dalla soluzione di base ammissibile (**0**, **b**). Se la soluzione ottima risulta il vettore nullo, allora anche PL ha soluzioni ammissibili: salvo casi di degenerazione, una di queste è individuabile nella ultima tabella del simplesso del problema ausiliario PL'. Da questa soluzione si può partire con le iterazioni per risolvere il problema di partenza, una volta eliminate le colonne del vettore ausiliario e introdotta la funzione oggetto originaria (**seconda fase**) sino alla applicazione di una delle regole d'arresto.

Può presentarsi il caso in cui nella tabella finale della prima fase il vettore **y** sia il vettore nullo, ma non tutte le sue componenti sono fuori base: supponiamo, tanto per fissare le idee, che sia ancora in base y_i. In tal caso non si ha a disposizione una soluzione di base formata da sole componenti del vettore **x**. Occorre allora fare uscire dalla base y_i. Per ottenere questo, basta effettuare un'operazione di pivot su di un qualsiasi elemento della riga i-esima corrispondente ad una delle variabili non ausiliarie. Ciò può essere effettuato anche con un elemento $a_{ik} < 0$: poiché il termine noto della equazione corrispondente è zero, non si perde l'ammissibilità della soluzione.

Se poi tutti gli elementi della riga i-esima a_{ik}, con k=1,2,…,n fossero nulli, ci si troverebbe di fronte ad un'equazione che è combinazione lineare di altre e quindi la caratteristica della matrice dei coefficienti A del sistema dei vincoli sarebbe inferiore a m, contro le ipotesi che si accettano generalmente nella impostazione di un problema in forma standard. In ogni caso, tale equazione va eliminata ed il procedimento può continuare con la seconda fase.

Per illustrare la procedura delle due fasi e per dare un esempio di risoluzione con il metodo del simplesso, consideriamo il seguente problema in forma standard:

$$\max\ 4\,x_1 + 3\,x_2 - 2\,x_3 + 7\,x_4$$

$$\begin{aligned} \text{s.t.}\quad &2\,x_1 + x_2 + x_3 - 3\,x_4 = 10, \qquad\qquad PL \\ &-3\,x_1 + 2\,x_2 - x_3 + 5\,x_4 = 12, \end{aligned}$$

$$x_j \geq 0.$$

Poiché il sistema dei vincoli non è in forma canonica, si formula il problema ausiliario (da risolvere nella prima fase):

$$\min \; y_1 + y_2$$

$$\begin{aligned} \text{s.t.} \quad & 2x_1 + x_2 + x_3 - 3x_4 + y_1 = 10, \\ & -3x_1 + 2x_2 - x_3 + 5x_4 + y_2 = 12, \end{aligned} \qquad \text{PL'}$$

$$x_j, y_i \geq 0.$$

Nella forma standard utilizzata nell'esposizione (funzione oggetto di massimo) il problema assume l'aspetto:

$$\max - y_1 - y_2$$

$$\begin{aligned} \text{s.t.} \quad & 2x_1 + x_2 + x_3 - 3x_4 + y_1 = 10, \\ & -3x_1 + 2x_2 - x_3 + 5x_4 + y_2 = 12, \end{aligned} \qquad \text{PL'}$$

$$x_j, y_i \geq 0.$$

e la tabella iniziale del simplesso risulta:

					*	*	
2	1	1	-3		1	0	10
-3	2	-1	5		0	1	12
0	0	0	0		1	1	0

Per effettuare le iterazioni, occorre azzerare i coefficienti dell'ultima riga corrispondenti alle variabili in base: basta aggiungere all'ultima riga le prime due dopo averle moltiplicate per –1. Si ottiene la tabella che segue:

					*	*	
2	1	1	-3		1	0	10
-3	2	-1	5		0	1	12
1	-3	0	-2		0	0	-22

A questo punto scegliendo come colonna entrante in base la seconda e quindi come pivot l'elemento $a_{22} = 2$, si ha la seconda tabella:

	*				*		
7/2	0	3/2	-11/2		1	-1/2	4
-3/2	1	–1/2	5/2		0	1/2	6

-7/2 0 -3/2 11/2 0 3/2 -4

Infine, scegliendo come colonna entrante la prima e quindi come pivot l'elemento $a_{11} = 7/2$, si ottiene la tabella:

*	*					
1	0	3/7	-11/7	2/7	-1/7	8/7
0	1	1/7	1/7	3/7	2/7	54/7
0	0	0	0	1	1	0

Si noti come la funzione oggetto, che è stata trasformata per avere un obiettivo di massimo, sia aumentata dal valore –22 sino al valore 0: contemporaneamente le variabili y_1 e y_2 sono uscite dalla base e si è ottenuta una soluzione di base con componenti positive costituite da elementi del vettore **x**.

Poiché si è ottenuto un valore ottimo pari a zero, si può affermare che il problema di partenza ammette soluzioni ammissibili e di queste ne abbiamo individuato una di base. Si eliminano quindi le colonne corrispondenti a y_1 e y_2 e si introduce la funzione oggetto originaria:

*	*			
1	0	3/7	-11/7	8/7
0	1	1/7	1/7	54/7
-4	-3	2	-7	0

Occorre adesso, come visto in precedenza, eliminare dalla funzione oggetto le variabili in base azzerando i coefficienti –4 e –3:

*	*			
1	0	3/7	-11/7	8/7
0	1	1/7	1/7	54/7
0	0	29/7	-90/7	194/7

A questo punto si può risolvere il problema di partenza (seconda fase) introducendo in base la variabile x_4 (che ha coefficiente di guadagno ridotto 90/7) usando come pivot l'elemento di valore 1/7. La tabella successiva (che è anche la conclusiva, poiché i coefficienti di guadagno ridotto sono negativi) risulta la seguente:

*			*	
1	11	2	0	86
0	7	1	1	54
0	90	14	0	722

b) Metodo di penalizzazione.

In alternativa alla procedura appena vista, è possibile ottenere la soluzione ottima del problema originario PL (o stabilire che la sua regione ammissibile è vuota) in una fase unica, mediante una formulazione del tipo:

$$\max \quad \mathbf{c}\,\mathbf{x} - M\,y_1 - M\,y_2 - \ldots - M\,y_m$$

$$\text{s.t.} \quad A\,\mathbf{x} + \mathbf{y} = \mathbf{b}, \qquad \mathbf{x}, \mathbf{y}, \geq \mathbf{0}, \qquad \text{(PL")}$$

dove M è una costante >0, scelta in modo tale da rendere conveniente che nella soluzione ottima del problema PL" il vettore **y** di variabili ausiliarie risulti nullo (M deve essere quindi sufficientemente grande).

Il problema PL" ha certamente regione ammissibile non vuota, poiché una soluzione ammissibile è data dal vettore (**x**, **y**) = (**0**, **b**), e quindi le iterazioni del simplesso si concluderanno in base ad una delle regole di arresto viste in precedenza.

Se M è grande a sufficienza e il problema iniziale PL ha ottimo finito, al termine del procedimento le variabili y_j risulteranno fuori base (proprio perché penalizzano la funzione oggetto) e quindi nella soluzione finale saranno positive solo componenti del vettore **x**. Se invece nella soluzione finale qualche componente y_j è rimasta in base, ciò significa che il problema di partenza non era ammissibile.

Il metodo di penalizzazione, apparentemente più semplice del metodo delle due fasi, deve essere utilizzato individuando un compromesso adeguato tra due esigenze contrastanti. Per 'forzare' l'uscita dalla base delle componenti del vettore **y** la costante M deve essere, come si è detto, sufficientemente grande, e ciò può portare a scegliere per la stessa valori di parecchi ordini di grandezza superiori a quello del più alto dei valori assoluti dei coefficienti dei vincoli e del vettore **b**. D'altra parte, occorre tenere presente che, a parte gli esempi didattici risolubili anche manualmente, l'utilizzo dell'elaboratore per la risoluzione di un PL comporta inevitabilmente arrotondamenti ed approssimazioni: in questi casi, la presenza simultanea in una matrice di coefficienti di valori di ordini di grandezza molto diversi può innescare una serie di errori e invalidare tutta la procedura.

2.1.15 Casi particolari: i problemi in forma canonica.

Nel caso dei problemi in forma canonica, se il vettore **b** dei termini noti dei vincoli è non negativo, una soluzione iniziale di base può essere ottenuta più agevolmente che non nel caso più generale di un problema in forma standard.

In una forma canonica può risultare $b_i < 0$, per qualche i. Si vedrà più avanti come in alcuni sviluppi, ad esempio nella teoria della dualità, sia utile lavorare con le forme canoniche, anche se alcuni b_i sono negativi. Il problema non si pone con la forma standard poiché in quest'ultima, con opportuni cambiamenti di segno, si possono avere termini noti non negativi.

In tale ipotesi. di fronte ad un problema di massimo l'introduzione delle variabili slack porta automaticamente alla presenza di una matrice identica e quindi si ha una prima soluzione di base, senza dover ricorrere al metodo delle due fasi.

Nei problemi di minimo, l'introduzione delle variabili surplus non dà lo stesso risultato perché si ottiene, accanto alla matrice originaria dei coefficienti, una matrice identità cambiata di segno.

Una prima soluzione di base in cui sono in base le variabili surplus avrebbe tutte le componenti non positive. Occorre ricorrere allora a procedure particolari, come la tecnica delle due fasi o l'algoritmo del simplesso duale che sarà descritto più avanti. Si può però ridurre la complessità del procedimento delle due fasi come segue.

Si consideri il generico problema, dopo l'introduzione delle variabili surplus:

$$\begin{aligned} & \min \; c_1 x_1 + c_2 x_2 + \ldots + c_n x_n \\ \text{s.t.} \quad & a_{11} x_1 + a_{12} x_2 + \ldots + a_{1n} x_n - s_1 = b_1 \\ & a_{21} x_1 + a_{22} x_2 + \ldots + a_{2n} x_n - s_2 = b_2 \\ & \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ & a_{m-1,1} x_1 + a_{m-1,2} x_2 + \ldots + a_{m-1,n} x_n - s_{m-1} = b_{m-1} \\ & a_{m1} x_1 + a_{m2} x_2 + \ldots + a_{mn} x_n - s_m = b_m, \\ & x_1, x_2, \ldots, x_n \geq 0. \end{aligned}$$

Individuiamo il massimo tra i termini noti nel vettore **b** e sottraiamo l'equazione corrispondente, membro a membro, da quelle degli altri vincoli. Supponendo che il coefficiente più grande sia b_m, si ottiene una formulazione del tipo:

$$\begin{aligned} & \min \; c_1 x_1 + c_2 x_2 + \ldots + c_n x_n \\ \text{s.t.} \quad & a'_{11} x_1 + a'_{12} x_2 + \ldots + a'_{1n} x_n - s_1 + s_m = b_1 - b_m \\ & a'_{21} x_1 + a'_{22} x_2 + \ldots + a'_{2n} x_n - s_2 + s_m = b_2 - b_m \\ & \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ & a'_{m-1,1} x_1 + a'_{m-1,2} x_2 + \ldots + a'_{m-1,n} x_n - s_{m-1} + s_m = b_{m-1} - b_m \\ & a_{m1} x_1 + a_{m2} x_2 + \ldots + a_{mn} x_n - s_m = b_m, \\ & x_1, x_2, \ldots, x_n \geq 0. \end{aligned}$$

dove con a'_{ij} si sono indicati i nuovi coefficienti nei vincoli dopo l'operazione descritta.

A questo punto le componenti del vettore dei termini noti sono tutte ≤ 0, tranne l'ultima. Con un cambiamento di segno nelle prime m-1 equazioni si ottiene:

$$\begin{aligned} & \min \; c_1 x_1 + c_2 x_2 + \ldots + c_n x_n \\ \text{s.t.} \quad & -a'_{11} x_1 - a'_{12} x_2 - \ldots - a'_{1n} x_n + s_1 - s_m = b_m - b_1 \\ & -a'_{21} x_1 - a'_{22} x_2 - \ldots - a'_{2n} x_n + s_2 - s_m = b_m - b_2 \\ & \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ & -a'_{m-1,1} x_1 - a'_{m-1,2} x_2 - \ldots - a'_{m-1,n} x_n + s_{m-1} - s_m = b_m - b_{m-1} \\ & a_{m1} x_1 + a_{m2} x_2 + \ldots + a_{mn} x_n - s_m = b_m, \end{aligned}$$

$$x_1, x_2, \ldots, x_n \geq 0.$$

Nella matrice dei coefficienti adesso sono presenti m-1 colonne (quelle corrispondenti alle prime m-1 variabili scarto) che formano una matrice identica, a meno dell'ultima colonna, cioè quella di s_m. Si può allora ricorrere al metodo delle due fasi introducendo soltanto una variabile ausiliaria, in corrispondenza dell'ultimo vincolo, con notevole risparmio nei calcoli.

Più in generale, nella applicazione del metodo delle due fasi è sufficiente introdurre tante variabili artificiali quante ne occorrono (ed in corrispondenza ai vincoli adeguati) affinché la matrice dei coefficienti del problema ausiliario contenga, salvo permutazioni, una sottomatrice identica di dimensione m. Salvo casi particolari, in genere basta aggiungere una variabile artificiale per ogni vincolo di eguaglianza e per ogni diseguaglianza del tipo ≥.

16. La degenerazione.

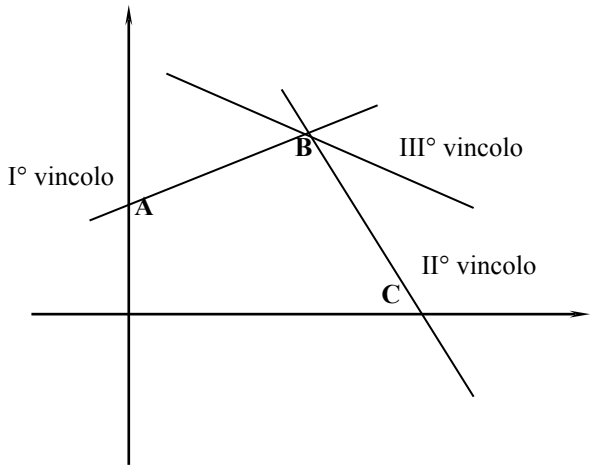
Nel caso in cui dopo un'iterazione dell'algoritmo del simplesso si ottenga una soluzione di base degenere, si può presentare il fenomeno del **ciclaggio** ('cycling' nella letteratura anglosassone): in base a determinate regole della scelta del pivot, in particolare ricorrendo al coefficiente di guadagno ridotto di valore massimo, può capitare che le iterazioni producano un insieme di soluzioni di base ammissibili, tutte degeneri, sempre con lo stesso valore della funzione oggetto, che ritornano ciclicamente.

Il fenomeno era considerato, nei primi studi sulla Programmazione Lineare, un fatto eccezionale, perciò si faceva affidamento molto su una 'improbabilità statistica' del suo verificarsi, ma le esperienze successive hanno dimostrato che un buon software di PL doveva essere in grado di superare tale ostacolo.

Un metodo per risolvere il problema è quello cosiddetto di perturbazione. In effetti, la natura di una soluzione di base degenere può essere facilmente illustrata con un esempio di problema con vincoli di diseguaglianza in due variabili (che quindi ammette una rappresentazione geometrica). Supponiamo che vi siano tre vincoli e che la regione ammissibile sia un poligono chiuso e limitato (individuato dagli assi coordinati e dai tre vincoli in questione). La figura con maggior numero di lati ottenibile è un pentagono (irregolare), ma anche altri poligoni sono possibili: se un vincolo è ridondante, si può avere un quadrilatero e così via. Il problema può essere descritto in forma generica come:

$$\begin{aligned} & \max \; c_1 x + c_2 y \\ \text{s.t.} \quad & a_{11} x + a_{12} y \leq b_1 \\ & a_{21} x + a_{22} y \leq b_2 \\ & a_{31} x + a_{32} y \leq b_3 \\ & x, y \geq 0. \end{aligned}$$

Supponiamo per semplificare l'esposizione che tutti i coefficienti nei vincoli (ed anche i termini noti) siano positivi e che la Regione Ammissibile sia il quadrilatero OABC della figura che segue:



Nella stessa si vede come il lato AB sia determinato dal primo vincolo; il lato BC dal secondo vincolo, mentre il terzo vincolo corrisponde ad una retta che passa per B (ed è all'atto pratico ridondante, perché la sua eliminazione non cambia nulla nella soluzione del problema).

Il punto B corrisponde ad una soluzione di base degenere. Infatti, la riduzione del problema a forma standard fa sì che si debbano introdurre tre variabili scarto (slack nel nostro caso), per cui il numero delle variabili stesse è 5 e quindi in una soluzione di base vi saranno due componenti certamente nulle ed altre tre componenti - in generale - diverse da zero.

Nel punto B si annullano viceversa tre variabili (tutte e tre le variabili scarto), per cui la soluzione è appunto degenere.

Si supponga adesso di variare il valore del coefficiente b_3 di una quantità $\varepsilon > 0$: per le ipotesi fatte sui segni dei coefficienti dei vincoli e sul vettore **b**, in corrispondenza di $b_3+\varepsilon$ il vincolo diventa del tutto superfluo (si creano due intersezioni distinte del vincolo stesso con gli altri due, corrispondenti a soluzioni di base non ammissibili); invece, in corrispondenza al valore $b_3-\varepsilon$ il vincolo (che si sposta verso l'origine degli assi) dà luogo ancora ad altre due intersezioni con i primi due e questa volta si tratta di due soluzioni di base ammissibili.

Se ne può dedurre che la soluzione degenere data dal punto B può essere considerata come una soluzione multipla (più soluzioni di base coincidenti). E' evidente poi come la perturbazione, sia che porti il terzo vincolo verso l'origine, sia che lo sposti nella direzione opposta, elimini la degenerazione e quindi certamente impedisce fenomeni di ciclaggio. Se ε è sufficientemente piccolo, si può determinare la soluzione ottima: con un errore limitato, se si tratta di B; in modo esatto, se si tratta di uno degli altri vertici del poligono.

In generale la tecnica di perturbazione suggerisce appunto di alterare di una quantità ε sufficientemente piccola i vincoli che danno luogo alla soluzione degenere (termine noto corrente nullo) e proseguire con le iterazioni.

Valgono a questo punto osservazioni analoghe a quelle sviluppate nell'uso dei coefficienti M del metodo di penalizzazione. La scelta di ε troppo piccolo, anche se sembra garantire a sua volta un errore piccolo nella soluzione finale, può creare problemi di carattere numerico nei calcoli dei coefficienti delle tabelle del simplesso, mentre se ε non è piccolo a sufficienza si può distorcere totalmente la regione ammissibile.

Il problema del ciclaggio è superabile ricorrendo ad una regola relativamente semplice formulata da Bland, all'Università di Lovanium nel 1977, regola che consiste in una scelta di tipo lessicografico sia della colonna che della riga del pivot.

La regola di Bland è la seguente:

Scegliere come colonna da fare entrare in base quella di indice più piccolo che presenta un coefficiente di guadagno ridotto positivo; se vi sono più elementi sulla stessa colonna che possono fungere da pivot, scegliere quello con indice più basso.

In pratica, la regola suggerisce, piuttosto che di cercare il coefficiente di guadagno ridotto più conveniente (quello di valore massimo), di utilizzare il primo coefficiente idoneo. Per quanto riguarda la riga, il problema di scelta nasce solo quando vi sono più elementi sulla colonna j-esima che entrerà in base per i quali contemporaneamente è minimo il rapporto b_i/a_{ij}: anche stavolta il suggerimento è di scegliere quello con indice minore. Bland dimostra effettivamente che così facendo si evita di cadere in un circuito di soluzioni di base.

17. Osservazioni conclusive.

Nella impostazione qui seguita si è adottata una formulazione del problema standard con funzione oggetto di massimo: in letteratura, viceversa, spesso si sviluppa il metodo del simplesso facendo riferimento ad una forma di minimo, cosa che comporta alcune variazioni (non essenziali) rispetto alle tabelle viste nei paragrafi precedenti.

Inoltre nell'ultima riga delle tabelle del simplesso si possono scrivere i coefficienti di guadagno o costo ridotto senza che gli stessi siano cambiati di segno.

Il problema è di 'lettura' delle condizioni di arresto e del valore della funzione oggetto. Se la tabella presenta coefficienti (di costo o guadagno) con il loro segno effettivo, è il valore della funzione oggetto che risulta invertito di segno e viceversa.

Infine è da osservare che la tecnica di simplesso finora sviluppata può essere definita **algoritmo del simplesso primale** in contrapposizione ad altre procedure, di cui sarà dato cenno più avanti, (in particolare, il **simplesso duale**), che permettono egualmente di ottenere la soluzione ottima con una successione di operazioni di cardine, in base ad altri principi che richiedono preliminarmente l'introduzione della teoria della dualità per i problemi di Programmazione Lineare.

2.2 La teoria della dualità.

2.2.1 Introduzione.

Per ogni problema di programmazione matematica, lineare o non lineare, sulla base di precise regole, se ne può scrivere un altro che è detto il suo **duale**. Tra problema originario e duale esistono legami molto stretti, che a prima vista possono apparire sorprendenti, ma che consentono di dedurre proprietà della soluzione di uno da quella dell'altro.

Come si è detto, si può parlare di dualità per un qualsiasi problema di programmazione matematica, ma nel caso dei problemi di Programmazione Lineare, l'impostazione è più semplice e risulta più immediata un'interpretazione. In generale, un software che risolve problemi di PL, fornisce informazioni anche sui relativi problemi duali, informazioni che risultano utili per condurre quella che in gergo tecnico è definita **analisi di sensitività**.

Per questo motivo, sarà qui introdotta la teoria della dualità per programmi lineari: prima di sviluppare gli aspetti teorici, il punto di partenza sarà costituito da un esempio didattico. Sarà ripreso in considerazione il caso della formulazione di un piano di produzione di guadagno massimo (biciclette da corsa e biciclette sportive) e si costruirà il problema duale corrispondente.

2.2.2 La dualità: un esempio di problema duale.

Nel problema della produzione delle biciclette sportive e delle biciclette da corsa, data una certa disponibilità di risorse, ed essendo l'azienda Ciclo attrezzata per produrre solo questi due tipi di articoli, i dirigenti si chiedevano quali fossero le quantità ottimali da produrre. Supponiamo, viceversa, che i titolari di un'altra azienda, la Meccan S.r.l., propongano alla Ciclo di prendere in affitto i tre reparti, con le ore macchina (potenziali) disponibili, per avviare una produzione diversa, ad es., di cofani per autovetture. La Meccan deve fare delle proposte in termini di canone di affitto orario mentre la Ciclo deve poi decidere se accettare o no ed eventualmente rinunciare alla produzione delle biciclette.

Il canone deve risultare attraente sia per la Ciclo, che prima produceva biciclette, sia per l'acquirente. Introdurremo pertanto tre variabili decisionali, che diremo a, b, c, che esprimono **i prezzi ai quali Meccan propone a Ciclo di prendere in affitto le ore dei tre reparti.**

La Meccan, da un lato cercherà di spendere il meno possibile per acquisire le risorse, d'altro canto deve prestare attenzione affinché anche per Ciclo la transazione sia più conveniente che non continuare con la vecchia produzione.

Pertanto, Meccan ha come obiettivo

$$\min 200\,a + 120\,b + 160\,c.$$

Per stabilire i vincoli, occorre partire dalla considerazione fatta sopra che Ciclo, se non ha introiti adeguati dalla cessione delle ore-reparto di cui è proprietaria, avrà convenienza a continuare la produzione di biciclette. Ora, mentre una bicicletta da corsa 'costa' 10 ore di stampaggio, 5 di montaggio e 4 di verniciatura, essa fa guadagnare 8: pertanto, se dalla vendita delle risorse necessarie per la fabbricazione di tale bicicletta non ricava almeno 8, Ciclo non accetterà la transazione. Pertanto il vincolo da rispettare è :

$$10\,a + 5\,b + 4\,c \geq 8.$$

Analogamente, considerando le biciclette sportive, dovrà essere

$$5\,a + 6\,b + 10\,c \geq 10.$$

Infine dovranno valere le consuete condizioni di non negatività.

$$a, b, c \geq 0.$$

In questo modo è stato formulato un altro problema di programmazione lineare (legato evidentemente a quello di programmazione della produzione) nel quale la funzione oggetto è di minimo, problema in forma canonica, come quello da cui si è partiti. Riassumendo, al problema (che diremo **primale**)

$$\max 8x + 10y$$

s.t.

$$\begin{aligned} 10x + 5y &\leq 200, \\ 5x + 6y &\leq 120, \\ 4x + 10y &\leq 160, \\ x, y &\geq 0, \end{aligned}$$

è stato associato il problema **duale**

$$\min 200a + 120b + 160c$$

s.t.

$$\begin{aligned} 10a + 5b + 4c &\geq 8, \\ 5a + 6b + 10c &\geq 10, \\ a, b, c &\geq 0. \end{aligned}$$

Più in generale, dato un qualsiasi problema di utilizzo ottimale di risorse per massimizzare il guadagno derivante dalla produzione di n articoli, introducendo m nuove variabili decisionali $y_1, y_2, ..., y_m$, per indicare il prezzo da fissare per le m risorse, la ditta che acquisisce le risorse stesse desidererà minimizzare la propria spesa, cioè minimizzerà l'espressione :

$$y_1 b_1 + y_2 b_2 + ... + y_m b_m.$$

Per l'azienda che cede le risorse, il calcolo è leggermente più complicato: occorre valutare l'equità dei prezzi delle risorse stesse pensando queste ultime incorporate negli oggetti (prodotti) che vanno a costituire. In particolare, in analogia con quanto prospettato nel caso delle biciclette da corsa e sportive, se l'azienda acquirente propone un sistema di prezzi **y**, le risorse che prima erano impegnate nella fabbricazione di un'unità del primo articolo risultano valutate dalla espressione:

$$a_{11} y_1 + a_{21} y_2 + ... + a_{m1} y_m$$

e questo importo deve essere non inferiore al guadagno c_1 che l'azienda produttrice otteneva impiegando le risorse nel proprio processo produttivo per fabbricare il primo articolo. Relazioni analoghe valgono per gli altri articoli.

In caso contrario, alla azienda che produce non conviene cedere le risorse, perché, se le usa lei direttamente, guadagna di più!

Infine i prezzi devono essere non negativi.

Il problema duale risulta quindi:

$$\min \quad y_1 b_1 + y_2 b_2 + \ldots + y_m b_m$$

(s. t.)

$$\begin{aligned} a_{11} y_1 + a_{21} y_2 + \ldots + a_{m1} y_m &\geq c_1, \\ a_{12} y_1 + a_{22} y_2 + \ldots + a_{m2} y_m &\geq c_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ a_{1n} y_1 + a_{2n} y_2 + \ldots + a_{mn} y_m &\geq c_n, \end{aligned}$$

con le condizioni di non negatività

$$y_1, y_2, \ldots, y_m \geq 0.$$

I valori $y_1, y_2, \ldots, y_m$ delle risorse prendono il nome di **prezzi ombra**, in quanto esprimono il valore delle risorse stesse una volta che siano incorporate nei vari prodotti.

Dall'esempio sopra riportato si possono ricavare alcune proprietà che si presentano sistematicamente nella teoria della dualità e che saranno riprese nei prossimi paragrafi. La **proprietà fondamentale** è la seguente:

il duale di un problema in forma canonica è a sua volta un problema in forma canonica.

Si osservi che è a questa proprietà che si fa riferimento ogniqualvolta si tratta di costruire il duale di un problema qualsiasi, nel senso che se un problema PL è espresso in una forma generale, con vincoli sia sotto forma di diseguaglianze, ≤ come ≥, sia di eguaglianze, e vi sono variabili negative o libere, il duale si ottiene passando attraverso la forma canonica del PL in questione.

Per la scrittura del duale valgono le seguenti **regole**:

1) se la funzione oggetto del problema di partenza (che prende come si è detto il nome di **primale**) **è di massimo, nel duale la funzione oggetto è di minimo e viceversa;**

2) nel passaggio dal primale al duale, i coefficienti della funzione oggetto e i termini noti dei vincoli si scambiano il ruolo;

3) la matrice A dei coefficienti del problema primale, nel duale è trasposta, cioè le righe diventano le colonne e viceversa; in questo modo ad ogni variabile del problema primale corrisponde un vincolo del problema duale e viceversa.

Sulla applicazione della proprietà fondamentale e delle regole 1), 2) e 3) torneremo nei prossimi paragrafi.

2.2.3 La dualità: scrittura del duale per un PL generale.

La scrittura del duale di un qualsiasi problema di Programmazione Lineare può essere effettuata attraverso i seguenti passi:

- **trasformazione** del problema primale in una delle forme canoniche (se occorre, anche utilizzando coefficienti $b_i < 0$);

- **scrittura** del duale nella **forma canonica** simmetrica (regole 1), 2) e 3));

- eventuale **conversione** del duale in un'altra forma (ad es., la standard).

Tuttavia per qualunque problema di programmazione lineare si può scrivere direttamente il duale, senza passare necessariamente attraverso la corrispondente forma canonica, applicando un insieme di regole: occorre però anche osservare che queste stesse regole si giustificano sempre attraverso la proprietà fondamentale e le regole 1), 2) e 3) enunciate alla fine del paragrafo precedente.

Innanzi tutto, va tenuta presente la **corrispondenza tra vincoli di un problema e variabili dell'altro**, per cui nella costruzione del problema duale occorre introdurre tante variabili quanti sono i vincoli del primale. Successivamente, il tipo di restrizione imposta ai singoli vincoli (=, ≥, ≤) determina il segno della variabile corrispondente nel duale e viceversa il tipo di restrizione su una variabile (≥0, ≤0, libera) determina la tipologia del vincolo corrispondente nell'altro problema.

Si dimostra che:

- **se nel primale vi è un vincolo di eguaglianza, la corrispondente variabile nel duale è libera; viceversa, se nel problema primale una variabile è libera, nel duale il vincolo corrispondente è sotto forma di eguaglianza.**

In particolare, tanto per fissare le idee, in un problema in forma standard (di massimo), ogni vincolo di eguaglianza

$$a_{i1}\, x_1 + a_{i2}\, x_2 + \ldots + a_{in}\, x_n = b_i$$

può essere sostituito dalla coppia di vincoli

$$a_{i1}\, x_1 + a_{i2}\, x_2 + \ldots + a_{in}\, x_n \le b_i$$
$$a_{i1}\, x_1 + a_{i2}\, x_2 + \ldots + a_{in}\, x_n \ge b_i$$

e per ottenere la forma canonica, il secondo di questi può essere scritto nuovamente come:

$$-a_{i1}\, x_1 - a_{i2}\, x_2 - \ldots - a_{in}\, x_n \le -b_i.$$

Il problema si presenta allora come segue:

$$\max\ c_1\, x_1 + c_2\, x_2 + \ldots + c_n\, x_n$$

$$\begin{aligned} \text{s. t.} \quad & a_{11}\, x_1 + a_{12}\, x_2 + \ldots + a_{1n}\, x_n \le b_1, \\ & a_{21}\, x_1 + a_{22}\, x_2 + \ldots + a_{2n}\, x_n \le b_2, \\ & \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ & a_{m1}\, x_1 + a_{m2}\, x_2 + \ldots + a_{mn}\, x_n \le b_m, \\ & -a_{11}\, x_1 - a_{12}\, x_2 - \ldots - a_{1n}\, x_n \le -b_1, \end{aligned}$$

$$
\begin{aligned}
- a_{21}\, x_1 - a_{22}\, x_2 - \ldots - a_{2n}\, x_n &\le - b_2, \\
&\ldots\ldots\ldots\ldots\ldots\ldots \\
- a_{m1}\, x_1 - a_{m2}\, x_2 - \ldots - a_{mn}\, x_n &\le - b_m, \\
x_1, x_2, \ldots, x_n &\ge 0.
\end{aligned}
$$

Per costruire il duale, diciamo u_i (i = 1, 2, …, m) le variabili associate ai primi m vincoli e v_i (i = 1, 2, …, m) quelle associate ai successivi m. Nel duale, la funzione oggetto è allora:

$$\min b_1 u_1 + b_2 u_2 + \ldots + b_m u_m - b_1 v_1 - b_2 v_2 - \ldots - b_m v_m$$

e il generico vincolo j-esimo risulta

$$a_{1j} u_1 + a_{2j} u_2 + \ldots + a_{mj} u_m - a_{1j} v_1 - a_{2j} v_2 - \ldots - a_{mj} v_m \ge c_j.$$

E' evidente che si può sostituire ad ogni coppia di variabili u_i e v_i una nuova variabile y_i con la posizione:

$$y_i = u_i - v_i$$

ed il problema è allora

$$\min b_1 y_1 + b_2 y_2 + \ldots + b_m y_m$$

$$
\begin{aligned}
\text{s. t.} \quad & a_{11} y_1 + a_{21} y_2 + \ldots + a_{m1} y_m \ge c_1, \\
& a_{12} y_1 + a_{22} y_2 + \ldots + a_{m2} y_m \ge c_2, \\
& \ldots\ldots\ldots\ldots\ldots\ldots \\
& a_{1n} y_1 + a_{2n} y_2 + \ldots + a_{mn} y_m \ge c_n.
\end{aligned}
$$

Le variabili y_i non hanno vincoli di segno (la differenza di due quantità positive può essere indifferentemente positiva, negativa o nulla), per cui si dice che le y_i sono 'libere' oppure si scrive anche: $y_i >< 0$.

Viceversa, se nel primale (che, tanto per fissare le idee, supponiamo di massimo con vincoli del tipo ≤) una variabile x_j è libera, passando alla forma canonica con la posizione

$$x_j = w_j - t_j$$

essendo w_j e t_j due variabili non negative, si ottiene un problema duale di minimo nel quale (in corrispondenza, rispettivamente di w_j e di t_j) figurano i due vincoli

$$
\begin{aligned}
a_{1j} y_1 + a_{2j} y_2 + \ldots + a_{mj} y_m &\ge c_j, \\
-a_{1j} y_1 - a_{2j} y_2 - \ldots - a_{mj} y_m &\ge -c_j,
\end{aligned}
$$

che, complessivamente, equivalgono all'unico vincolo (di eguaglianza)

$$a_{1j} y_1 + a_{2j} y_2 + \ldots + a_{mj} y_m = c_j.$$

Infine, si può evitare di cambiare verso ai vincoli di diseguaglianza nelle forme che non siano coerenti con la forma canonica (vincoli di ≥ con problemi di massimo e vincoli di ≤ con problemi di minimo) con la seguente regola:

- **ad ogni vincolo di tipo ≤ in un problema di minimo oppure di tipo ≥ in un problema di massimo corrisponde nel duale una variabile ≤ 0.**

Ovviamente utilizzando questa regola, la matrice dei coefficienti del problema duale è la trasposta della matrice del problema primale, senza dover procedere a cambiamenti nei segni. La regola stessa si giustifica ancora una volta con un passaggio alla forma canonica. Ad esempio, se in un problema di massimo

$$\max \; c_1 x_1 + c_2 x_2 + \ldots + c_n x_n$$

compare il vincolo

$$a_{i1} x_1 + a_{i2} x_2 + \ldots + a_{in} x_n \ge b_i$$

che non è coerente con la forma canonica (che prevederebbe tutti vincoli di tipo ≤), si può cambiare verso allo stesso, riscrivendolo:

$$- a_{i1} x_1 - a_{i2} x_2 - \ldots - a_{in} x_n \leq - b_i,$$

e, nel successivo passaggio al duale, la corrispondente variabile, che diremo u_i, risulterà non negativa, ma in ciascun vincolo avrà coefficiente $– a_{ij}$. Per tornare ai coefficienti a_{ij}, basterà la posizione

$$y_i = - u_i$$

e quindi y_i sarà una variabile ≤ 0.

Riassumendo, le regole da applicare nel passaggio da primale a duale sono riassumibili nella seguente tabella:

	problema primale		problema duale	
f. oggetto	max		min	
	min		max	
vincoli	$\leq$	con max	variabile	≥ 0
	$\geq$	con min		≥ 0
	$\geq$	con max	variabile	≤ 0
	$\leq$	con min		≤ 0
	eguaglianza		variabile	libera

variabile	≥ 0	vincolo	$\leq$ (max) o $\geq$ (min)
	≤ 0		$\geq$ (max) o $\leq$ (min)
	libera		eguaglianza

E’ il caso di ribadire il legame incrociato tra tipologia dei vincoli e segno delle corrispondenti variabili: come regola mnemonica, si può far ricorso al fatto che in un problema con vincoli di disuguaglianza la forma più naturale è quella canonica, nella quale i vincoli si presentano, come già osservato in precedenza, ‘**antagonisti**’ rispetto alla funzione oggetto (vincoli di $\geq$ con funzione obiettivo di minimo e di $\leq$ con funzione obiettivo di max), ed in tal caso la variabile corrispondente risulta nella forma consueta di ≥ 0; viceversa, con vincoli non antagonisti la variabile duale corrispondente risulta ≤ 0.

Da un punto di vista pratico, per scrivere un duale si procede come segue:

- si associa a ogni vincolo del primale una nuova variabile (occorreranno tante nuove variabili duali quanti sono i vincoli del primale);
- si scrive la funzione oggetto scambiando max con min e viceversa;
- si traspone la matrice dei coefficienti;
- si adottano le tipologie di vincoli corrispondenti alla tabella precedente;
- si assegnano alle variabili i vincoli di non negatività o non positività come nella tabella precedente.

Per illustrare la procedura, si consideri il seguente esempio.

Sia dato il problema:

$$\max x_1 + 2 x_2 + 3 x_3$$

s.t.

$$
\begin{aligned}
4\,x_1 + 5\,x_2 + 6\,x_3 &\leq 7,\\
8\,x_1 + 9\,x_2 - 3\,x_3 &\geq 10,\\
11\,x_1 - 2\,x_2 + 12\,x_3 &= -3,
\end{aligned}
$$

$$x_1 \geq 0;\ x_2 \leq 0;\ x_3 \text{ libera.}$$

Il duale dovrà prevedere tre variabili y_1, y_2 e y_3, associate rispettivamente al primo, al secondo ed al terzo vincolo, e conterrà tre vincoli, a loro volta associati rispettivamente alle variabili x_1, x_2 e x_3 del primale. Il duale risulta:

$$
\begin{aligned}
&\min\ 7\,y_1 + 10\,y_2 - 3\,y_3\\
\text{s.t.}\quad &\\
&4\,y_1 + 8\,y_2 + 11\,y_3 \geq 1\\
&5\,y_1 + 9\,y_2 - 2\,y_3 \leq 2\\
&6\,y_1 - 3\,y_2 + 12\,y_3 = 3
\end{aligned}
$$

$$y_1 \geq 0;\ y_2 \leq 0;\ y_3 \text{ libera.}$$

Si osservi infine che, in base alle regole sopra scritte, è facile verificare che il duale del duale è il problema primale di partenza.

2.2.4 L'interpretazione economica del problema duale.

Ad un problema di programmazione lineare nel quale la funzione oggetto abbia il carattere di un costo oppure di un guadagno corrisponde un problema duale al quale si può attribuire un significato in base al quale la sua funzione oggetto ha carattere simmetrico rispetto al primale (cioè, rispettivamente, di guadagno oppure di costo), ma mantiene sempre un significato di tipo monetario.

Più in generale, come si vedrà più avanti in altri ambiti applicativi, si possono interpretare primale e duale in modo tale che la dimensione (in senso fisico) del valore z della funzione oggetto rimane la stessa nei due problemi.

Ad esempio, in un problema ingegneristico di massimizzazione del flusso di un bene che può scorrere attraverso una serie di canalizzazioni (ad es., in una rete di tubature di acquedotto con relative portate) che si può impostare come problema di programmazione lineare, anche il duale è interpretabile come la ricerca di tratti della rete aventi particolari caratteristiche di portata.

In questo paragrafo si darà una giustificazione di tale proprietà per il problema di trasporto e per il problema della dieta, già introdotti nella sezione precedente.

Il **problema di trasporto**, sinteticamente, è quello di un'ipotetica ditta D che deve determinare un piano di trasporto di una particolare merce da più luoghi di produzione P_i (i = 1, 2, …, m) a più luoghi di consumo M_j (j = 1, 2, …, n) minimizzando i costi: questi sono proporzionali alla merce trasportata secondo dei coefficienti c_{ij} che variano a seconda delle varie coppie origine / destinazione.
Nel generico luogo P_i è disponibile la quantità a_i, mentre in M_j la richiesta è data da b_j.

Si tratta di un problema non speculativo, nel senso che si ipotizza che la ditta D debba soddisfare tutta la domanda e quindi questo va fatto nel modo meno costoso. A sua volta, ciò richiede che valga la diseguaglianza

$$\Sigma\, a_i \geq \Sigma\, b_j.$$

Analiticamente, il problema di trasporto si può scrivere in maniera diversa, a seconda che la relazione tra quantità disponibili e quantità richieste sia di diseguaglianza in senso stretto o sia invece una relazione di eguaglianza..

Se offerta e domanda complessive coincidono, caso al quale ci si può ricondurre con un semplice espediente già illustrato in precedenza (introduzione di un mercato fittizio), allora si può scrivere:

$$\min \ \Sigma_{ij}\, c_{ij}\, x_{ij}$$

$$\begin{aligned} \text{s.t.} \quad & x_{i1} + x_{i2} + \ldots + x_{in} = a_i, \qquad (i = 1, 2, \ldots, m), \\ & x_{1j} + x_{2j} + \ldots + x_{mj} = b_j, \qquad (j = 1, 2, \ldots, n), \end{aligned}$$

$$x_{ij} \geq 0.$$

Se invece più, in generale, la produzione complessiva è ≥ della domanda, i vincoli assumono l'aspetto

$$\begin{aligned} & x_{i1} + x_{i2} + \ldots + x_{in} \leq a_i \qquad (i = 1, 2, \ldots, m), \\ & x_{1j} + x_{2j} + \ldots + x_{mj} \geq b_j \qquad (j = 1, 2, \ldots, n). \end{aligned}$$

Questo sul duale ha solo conseguenze formali, non di sostanza.
Il problema duale richiede m+n variabili, che convenzionalmente si indicano:

- con i simboli u_i, (i = 1, 2, …, m), per le variabili associate ai primi m vincoli (vincoli sulle origini);
- con i simboli v_j, (j = 1, 2, …, n), per le variabili associate ai successivi n vincoli (vincoli sulle destinazioni).

Con riferimento alla prima formulazione data sopra, le variabili sono libere (perché i corrispondenti vincoli sono sotto forma di eguaglianza).

Il duale risulta allora:

$$\max \ \Sigma\, a_i\, u_i + \Sigma\, b_j\, v_j$$

s.t.

$$u_i + v_j \leq c_{ij} \qquad (i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n),$$

$$u_i,\ v_j \text{ libere.}$$

Per la costruzione del duale è utile fare riferimento alla matrice dei coefficienti del primale come è stata scritta per esteso nella prima sezione di questo capitolo:

$$
\begin{array}{cccccccccccccccccccc}
1 & 1 & 1 & \ldots & 1 & 0 & 0 & 0 & \ldots & 0 & \ldots\ldots\ldots & 0 & 0 & 0 & \ldots & 0 \\
0 & 0 & 0 & \ldots & 0 & 1 & 1 & 1 & \ldots & 1 & \ldots\ldots\ldots & 0 & 0 & 0 & \ldots & 0 \\
\multicolumn{20}{c}{\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots} \\
0 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 & \ldots & 0 & \ldots\ldots\ldots & 1 & 1 & 1 & \ldots & 1 \\
1 & 0 & 0 & \ldots & 0 & 1 & 0 & 0 & \ldots & 0 & \ldots\ldots\ldots & 1 & 0 & 0 & \ldots & 0 \\
0 & 1 & 0 & \ldots & 0 & 0 & 1 & 0 & \ldots & 0 & \ldots\ldots\ldots & 0 & 1 & 0 & \ldots & 0 \\
0 & 0 & 1 & \ldots & 0 & 0 & 0 & 1 & \ldots & 0 & \ldots\ldots\ldots & 0 & 0 & 1 & \ldots & 0 \\
\multicolumn{20}{c}{\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots} \\
0 & 0 & 0 & \ldots & 1 & 0 & 0 & 0 & \ldots & 1 & \ldots\ldots\ldots & 0 & 0 & 0 & \ldots & 1
\end{array}
$$

Tenendo presente che, come già osservato illustrando il problema (primale) di trasporto, in corrispondenza ad ogni colonna vi sono due soli elementi diversi da zero (corrispondenti a tutte le possibili coppie (i,j) di origini e destinazioni), ci si può facilmente rendere conto che in ogni vincolo del duale vi sono solo due variabili e le stesse sono messe in relazione con il coefficiente c_{ij} avente gli indici delle due variabili. In pratica, la struttura del duale è molto più semplice di quella del primale e si presta alla costruzione di tecniche risolutive alternative al metodo del simplesso.

Per dare un'interpretazione economica al problema duale, si pensi ad un'azienda di trasporto T che propone alla ditta D, che produce e commercializza l'articolo, di prendere in appalto l'operazione di trasporto.

Tecnicamente, si può pensare che la ditta T acquisti da D ai prezzi u_i la merce disponibile nei luoghi di produzione, e poi rivenda la merce stessa di nuovo alla ditta D, nei luoghi di consumo, ai prezzi v_j. T deve cercare un sistema di prezzi per i quali entrambe le ditte devono avere convenienza nello stipulare il contratto di appalto. Questo giustifica da un lato la funzione oggetto e dall'altro i vincoli:

- per quanto riguarda la funzione oggetto, i prezzi u_i e v_j devono essere tali da massimizzare il ricavo della ditta T;
- per quello che concerne i vincoli, la somma algebrica dei prezzi relativi ad un generico percorso dal luogo di produzione P_i al luogo di consumo M_j non deve superare il costo c_{ij} che l'azienda D sostiene per effettuare il trasporto in proprio.

Intuitivamente, i prezzi proposti da T, u_i e v_j, che nella formulazione del duale sono variabili libere, dovrebbero rispettare le diseguaglianze

$$u_i < 0, \qquad v_j > 0,$$

poiché l'azienda T **acquista** da D nei luoghi di produzione mentre **vende** a D nei luoghi di consumo, evidentemente ad un prezzo maggiorato che la ricompensi dei costi di trasporto. I prezzi da fissare sono indipendenti dai possibili percorsi, (sono fissi per ogni P_i e per ogni M_j), perché alla ditta D che dovrebbe accettare l'appalto non interessa più quali sono i viaggi effettivamente scelti da T: come organizzare il trasporto diventa solo un problema di quest'ultima azienda. Il segno 'ovvio' delle variabili emerge chiaramente se si scrive il duale a partire dal primale con i vincoli di diseguaglianza. Infatti, in tal caso ai vincoli sulle origini

$$x_{i1} + x_{i2} + \ldots + x_{in} \leq a_i \qquad (i = 1, 2, \ldots, m),$$

che sono antitetici rispetto alla forma canonica di minimo, corrispondono variabili $u_i \leq 0$, mentre ai vincoli sulle destinazioni corrispondono variabili ≥ 0.

Ci si può rendere conto però che in realtà a T non interessano tanto i valori assoluti delle variabili, $|u_i|$ e $|v_j|$, quanto le differenze $|v_j| - |u_i|$ oppure, se si vuole, le somme algebriche, $u_i + v_j$. Per esemplificare, se $u_i = -10$ e $v_j = 30$, la ditta T ricava per ogni unità di merce trasportata da P_i a M_j lo stesso importo (20) che se i prezzi fossero, ad esempio, $u_i = -15$ e $v_j = 35$ oppure $u_i = 0$ e $v_j = 20$. Più in generale, sono equivalenti tutti i sistemi di prezzi che differiscono tra loro per una costante k sommata ad ogni v_j e sottratta ad ogni u_i:

$$u_i - k,\ v_j + k.$$

Tale costante è irrilevante per la funzione oggetto (considerato che la somma delle quantità prodotte coincide con la somma delle quantità richieste).

Dal punto di vista di D, continuando l'esempio numerico precedente, un costo di trasporto $c_{ij} = 25$ renderebbe l'appalto comunque conveniente perché D spenderebbe di più ad organizzare il viaggio $P_i \rightarrow M_j$ con la propria organizzazione.

Si osservi infine che, al di là dei limiti evidenti nel modello, la convenienza per T ad assumere l'appalto dipende da elementi che nel modello non compaiono. Si può pensare che, plausibilmente, la ditta T, specializzata nel campo dei trasporti, abbia costi unitari per le varie coppie (origine, destinazione) più bassi dei costi per la ditta D. Tuttavia questo, ripetiamolo, esula dal modello così come è impostato.

Il duale del **problema della dieta** può essere interpretato come il problema che deve risolvere una ditta D che produce elementi nutritivi sintetici (ad esempio, confezionati in pillole) proposti come alternativi ai cibi naturali costituenti la dieta stessa. L'azienda D deve stabilire i prezzi y_i ($i = 1, 2, \ldots, m$) ai quali vendere le varie pillole, tenendo conto ovviamente dei prezzi di mercato dei cibi convenzionali con i quali si vuole porre in competizione. Pertanto, se da un lato D sa di poter vendere i quantitativi $b_1, b_2, \ldots, b_m$ dei vari fattori nutritivi essenziali alla dieta, ricavandone la somma

$$b_1 y_1 + b_2 y_2 + \ldots + b_m y_m,$$

che tenderà a rendere la più grande possibile, d'altro canto per restare competitiva deve far sì che gli stessi quantitativi dei vari fattori nutritivi che sono contenuti in un'unità del generico alimento j-esimo, tradotti in pillole e valutati ai prezzi y_i, non diano luogo ad una spesa superiore a quella occorrente per acquistare l'alimento stesso. Ciò dà luogo ai vincoli

$$a_{j1} y_1 + a_{j2} y_2 + \ldots + a_{jm} y_m \leq c_j \qquad (j = 1, 2, \ldots, n).$$

La teoria della dualità permette di giungere ad una conclusione che può apparire inaspettata: non vi è nei casi sopra riportati di problemi di natura economica una vera e propria competizione, perché il valore ottimo della funzione oggetto nel duale coincide con quello del problema primale e quindi è indifferente nel caso della ditta Ciclo accettare le proposte di Meccan o continuare a produrre le biciclette, sportive e da corsa, come pure è indifferente, per l'azienda D che produce e commercializza in luoghi diversi, ricorrere alla ditta di trasporto T.

A questi aspetti della dualità sono dedicati i prossimi paragrafi.

Occorre anche fare attenzione a non dedurre da queste ultime affermazioni conclusioni prive di fondamento. Restando sugli esempi precedenti, non è detto che Meccan non finisca con affittare i macchinari della Ciclo: in effetti tutto dipende dalla redditività del processo produttivo che Meccan intende mettere in atto, e di questo il modello non dice nulla. Se la redditività è sufficientemente elevata, Meccan potrà proporre dei prezzi di affitto leggermente più bassi dei valori (di equilibrio) che rendono per Ciclo indifferente accettare o no la transazione, come pure – nel duale del problema di trasporto – l'azienda T, se è bene organizzata, proporrà prezzi leggermente più convenienti per D che non la gestione in proprio da parte di D stessa dei viaggi e otterrà l'appalto.

Ad una conclusione analoga si può pervenire per il problema della dieta.

2.2.5 Teoremi sulla dualità: proprietà reciproche di primale e duale.

Il problema duale di un problema assegnato, pur con le giustificazioni che sono state date attraverso i vari esempi, rimarrebbe un elemento puramente formale se non vi fossero particolari proprietà che rendono effettivamente utile la sua introduzione. Queste proprietà si traducono in alcuni teoremi che saranno esposti in questo paragrafo.

Il primo risultato fa riferimento al caso in cui si abbiano informazioni su entrambi i problemi, primale e duale.

Se sia un problema primale sia il suo duale ammettono soluzioni ammissibili:

a) entrambi ammettono soluzione ottima e il valore ottimale della funzione oggetto nei due problemi coincide;

b) il valore della funzione oggetto nel problema di minimo in corrispondenza di una qualunque soluzione ammissibile è ≥ del valore della funzione oggetto nel problema di massimo.

c) se in corrispondenza di una coppia di tali soluzioni ammissibili il valore della funzione oggetto nei due problemi coincide, allora le due soluzioni sono ottime per i rispettivi problemi.

Con informazioni su di un problema solo, si dimostra quanto segue:

d) se un problema primale ha ottimo finito, anche il duale ha ottimo finito;

e) se il primale è ammissibile e non ha ottimo finito, allora il duale ha regione ammissibile vuota;

f) se il primale non è ammissibile, il duale o è a sua volta non ammissibile oppure è ammissibile ma non ha ottimo finito.

Le proprietà sopra enunciate sono valide a prescindere dalla particolare formulazione (standard, canonica o generale) del primale e quindi anche del duale.

Gli enunciati b) e c) sono pressoché immediati. Infatti, data una coppia di problemi in forma canonica,

$$\max \mathbf{cx} \quad \text{s.t.} \quad \mathbf{Ax} \leq \mathbf{b}, \ \mathbf{x} \geq \mathbf{0}; \qquad \min \mathbf{yb} \quad \text{s.t.} \quad \mathbf{yA} \geq \mathbf{c}, \ \mathbf{y} \geq \mathbf{0},$$

se si moltiplica a sinistra il sistema dei vincoli nel primale per **y** e si moltiplica a destra il sistema dei vincoli nel duale per **x**, (tenendo presente che **x** e **y** sono vettori non negativi) si ottengono le diseguaglianze (valide per ogni coppia di soluzioni ammissibili):

$$\mathbf{cx} \leq \mathbf{yAx} \leq \mathbf{yb}.$$

Pertanto, come si vede, il valore della funzione oggetto nel problema di massimo è superiormente limitato da qualsiasi valore della funzione oggetto nel problema di minimo e viceversa. E' poi evidente che se s'individua una coppia di soluzioni, $\mathbf{x}^\circ$ del primale e $\mathbf{y}^\circ$ del duale, per le quali il valore della funzione oggetto coincide, allora si tratta necessariamente delle soluzioni ottime dei rispettivi problemi, proprio a causa della limitazione reciproca appena stabilita.

La stessa relazione può essere utilizzata per dimostrare l'enunciato e). Infatti, se uno dei due problemi, poniamo per esemplificare si tratti del problema di massimo, è ammissibile ma non ha ottimo finito, allora il suo duale è necessariamente vuoto, perché se per assurdo ciò non fosse ed il problema di minimo avesse una soluzione ammissibile $\underline{\mathbf{y}}$, allora dovrebbe essere anche $\mathbf{cx} \leq \underline{\mathbf{y}}\mathbf{b}$ per ogni soluzione ammissibile del primale **x** contro l'ipotesi che l'ottimo del primale non esista finito.

L'enunciato f) si può dimostrare con opportuni esempi.

Le dimostrazioni precedenti rimangono sostanzialmente le stesse se il primale è in forma standard: in tal caso si perviene alla relazione

$$\mathbf{cx} \leq \mathbf{yAx} = \mathbf{yb},$$

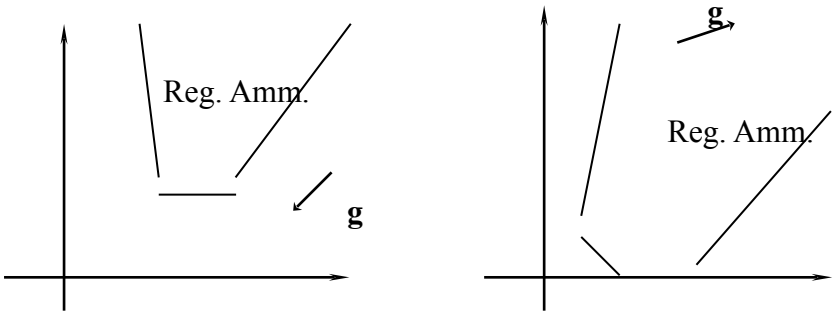
con le stesse conseguenze viste sopra.

Per quanto riguarda gli enunciati a) e d) occorre fare riferimento all'espressione della soluzione di un problema di programmazione lineare e questa è più facilmente descrivibile ricorrendo alla forma standard, che è quella richiesta per il metodo del simplesso ed alla rappresentazione matriciale della soluzione ottima e delle condizioni d'ottimalità sui coefficienti di guadagno ridotto, che saranno illustrati nel prossimo paragrafo.

A commento dei precedenti enunciati si osservi che, nel caso di ammissibilità sia del primale sia del duale, nei due problemi coincidono i due valori ottimali z* delle funzioni oggetto, **non** le due soluzioni **x*** e **y***: queste in generale sono vettori di spazi di dimensione diversa, quindi nemmeno confrontabili.

Inoltre, è il caso di porre attenzione al fatto che un ottimo 'non finito' richiede necessariamente regione ammissibile illimitata, mentre non è detto valga il viceversa, come si può vedere dai due esempi che seguono.

Nel primo caso, il gradiente, che indica la direzione di crescita della funzione oggetto, fa si che l'ottimo sia finito. Nel secondo caso l'ottimo non esiste.



2.2.6. La forma matriciale del metodo del simplesso.

Le tabelle del simplesso possono essere espresse in forma matriciale. Come passo iniziale si tratta di individuare nella matrice dei coefficienti A una sottomatrice di base B, e di conseguenza la restante sottomatrice delle colonne fuori base N, e quindi partizionando il vettore dei guadagni come $\mathbf{c} = (\mathbf{c}_B, \mathbf{c}_N)$.

Con il cambiamento di segno in questi ultimi vettori, si ottiene la tabella:

B	N	**b**
$-\mathbf{c}_B$	$-\mathbf{c}_N$	0

La leggibilità immediata della soluzione di base corrispondente richiede che la tabella, in seguito a opportune trasformazioni (quali ad es., il metodo delle due fasi), sia portata in una forma nella quale in luogo della matrice B compare una matrice identica I di dimensione m×m, i termini noti sono la soluzione del sistema $B\,\mathbf{x}_B = \mathbf{b}$, mentre nell'ultima riga compaiono

- un vettore nullo in corrispondenza di I;
- i coefficienti di guadagno ridotto cambiati di segno in corrispondenza delle colonne fuori base.

Per ottenere questo risultato è sufficiente moltiplicare a sinistra la matrice iniziale per la matrice seguente:

$$\begin{array}{c|c} B^{-1} & \mathbf{0} \\ \hline \mathbf{c}_B B^{-1} & 1 \end{array}$$

Si ottiene la tabella tipica del simplesso come segue:

I	$B^{-1} N$	$B^{-1} \mathbf{b}$
0	$- \mathbf{c}_N + \mathbf{c}_B B^{-1}\mathbf{N}$	$\mathbf{c}_B B^{-1} \mathbf{b}$

La stessa è la tabella finale se i coefficienti di guadagno ridotto sono negativi o nulli, cioè se risulta positivo (o non negativo) il vettore

$$- \mathbf{c}_N + \mathbf{c}_B B^{-1} \mathbf{N}.$$

In tal caso il vettore $\mathbf{x} = (\mathbf{x}_B = B^{-1}\mathbf{b}, \mathbf{0})$ è soluzione ottima e il valore della funzione oggetto corrispondente è $z^* = \mathbf{c}_B B^{-1} \mathbf{b}$.

Come si vedrà in seguito, la formulazione matriciale dell'algoritmo del simplesso consente di dare una giustificazione anche delle proprietà a) e d) enunciate nel paragrafo 2.2.5 relative ai legami tra problema primale e duale.

7. Il teorema di complementarità.

Per ottenere la soluzione di un problema duale, nota che sia quella del problema primale, si può utilizzare un importante risultato della teoria della dualità: il **Teorema di Complementarità**. Il teorema stesso ha interesse anche per gli aspetti interpretativi che se ne possono trarre sul rapporto tra la soluzione ottima di un problema e i vincoli dell'altro.

Condizione necessaria e sufficiente affinché una coppia di soluzioni x e y, ammissibili rispettivamente per un problema primale e per il suo duale, siano soluzioni ottime, è che valgano le condizioni, dette di complementarità

$$\mathbf{x_j (a_{1j} y_1 + a_{2j} y_2 + \ldots + a_{mj} y_m - c_j) = 0 \quad (j = 1, 2,\ldots, n),}$$

$$\mathbf{y_i (a_{i1} x_1 + a_{i2} x_2 + \ldots + a_{in} x_n - b_i) = 0 \quad (i = 1, 2, \ldots, m).}$$

Come si vede, le condizioni fanno riferimento ciascuna ad una variabile di uno dei due problemi ed alla espressione del corrispondente vincolo sull'altro.

Da questo punto di vista, il teorema può essere enunciato distinguendo i vincoli in **attivi** e **non attivi**: si dice attivo, in corrispondenza di una soluzione, un vincolo soddisfatto come eguaglianza; in caso contrario il vincolo è non attivo. Le condizioni di complementarità implicano allora che

- se in corrispondenza della soluzione ottima di un problema un vincolo è non attivo (cioè è soddisfatto come diseguaglianza stretta), allora nella soluzione ottima del duale la corrispondente variabile (duale) è nulla;

- se una variabile nella soluzione ottima di un problema ha valore positivo, allora il vincolo corrispondente nel problema duale, in corrispondenza della soluzione ottima, è attivo (soddisfatto come eguaglianza).

Infatti in tali circostanze i prodotti che costituiscono le condizioni sono nulli.

Osservazione. Il teorema di complementarità rimane valido a prescindere dalla particolare forma della coppia di problemi tra loro duali per i quali è enunciato. Occorre però notare che se uno dei due problemi è in forma standard, i vincoli relativi sono sempre sotto forma di eguaglianza e quindi non se ne possono trarre informazioni sulla soluzione del duale. D'altra parte, occorre porre attenzione al fatto che le condizioni di complementarità stesse non dicono nulla nemmeno quando si costata che in una soluzione ottima una componente è nulla (il corrispondente vincolo duale può avere comportamento qualunque) oppure quando un vincolo è attivo (la corrispondente variabile duale può essere indifferentemente eguale o diversa da zero).

Dimostrazione. Supponiamo che $\mathbf{x^*}$ e $\mathbf{y^*}$ siano soluzioni ottime, rispettivamente, di un problema primale di massimo e del suo duale di minimo, che potremo supporre senza nulla togliere alla generalità del ragionamento, in forma canonica. In tal caso, con i simboli sinora adottati, risulta:

$$\mathbf{cx^* = y^*Ax^* = y^*b}.$$

In particolare, dalla relazione di sinistra, si deduce l'eguaglianza:

$$\mathbf{(c - y^*A)\, x^*} = 0$$

che, scritta in forma estesa, fornisce:

$$\Sigma\, (c_j - a_{1j}\, y_1 - a_{2j}\, y_2 - \ldots - a_{mj}\, y_m)\, x_j = 0. \qquad (*)$$

Ora, essendo $\mathbf{x^*}$ un vettore ammissibile a componenti tutte ≥ 0 ed essendo $c_j \leq \Sigma\, a_{ij}\, y_i$, ogni addendo della sommatoria (*) risulta ≤ 0: affinché la sommatoria stessa sia nulla, tutti gli addendi devono essere nulli (si noti che ciascuno di essi costituisce una delle condizioni di complementarità del teorema). In maniera del tutto analoga si stabiliscono le relazioni di complementarità tra vincoli del problema di massimo e variabili y_i del problema di minimo.

Viceversa, se valgono le condizioni di complementarità per una coppia di vettori $\mathbf{x}$ e $\mathbf{y}$, sommando a membro a membro le prime m condizioni, si ottiene l'eguaglianza (in forma matriciale)

$$\mathbf{(c - yA)\ x} = 0$$

e, analogamente, dalle successive n condizioni si ricava

$$\mathbf{y}\,(A\mathbf{x} - \mathbf{b}) = 0.$$

Da queste due ultime relazioni, infine, si può dedurre che $\mathbf{c\,x = y\,b}$ e questo indica che i vettori $\mathbf{x}$ e $\mathbf{y}$ sono soluzioni ottime dei rispettivi problemi.

2.2.8 Interpretazione economica delle condizioni di complementarità.

E' interessante leggere in chiave economica le condizioni di complementarità, nei modelli in cui la funzione oggetto rappresenta un costo oppure un guadagno. Facendo riferimento ad un problema primale di programmazione della produzione (per massimizzare un guadagno) in presenza di risorse limitate e al suo duale, di affitto delle risorse, stabilendone i relativi prezzi ombra, le condizioni si possono esprimere affermando che:

- se nella politica ottima del primale una risorsa non è consumata totalmente, allora il suo prezzo ombra nel duale è 0;
- se uno dei possibili articoli nella politica ottima del primale è effettivamente prodotto, allora il suo valore, espresso mediante i prezzi ombra, eguaglia il guadagno (valore e guadagno unitari).

Ovviamente queste due affermazioni vanno a loro volta interpretate, a evitare significati distorti. Da un punto di vista della scienza economica i prezzi ombra vanno interpretati come valori **marginali**, vale a dire essi esprimono non tanto il valore unitario valido per qualsiasi quantitativo della risorsa, quanto il valore che di un'unità aggiuntiva rispetto alla quantità che è utilizzata o che comunque è già a disposizione, ipotizzando anche che l'unità di misura sia sufficientemente piccola. Sotto quest'aspetto è positivo solo il valore marginale delle risorse scarse, cioè quelle che in una politica produttiva sono pienamente utilizzate.

Una maggiore disponibilità di risorse, in generale, consente di aumentare la produzione e quindi anche, in ipotesi di linearità, il relativo guadagno. E' però evidente che sino a che non si accresce la disponibilità delle risorse che sono completamente esaurite non si può avere alcun vantaggio. Viceversa, aumentando quelle scarse, aumenta pure la produzione: solo però sino a quando a loro volta le risorse che prima erano abbondanti non diventano loro stesse dei colli di bottiglia.

Allo stesso modo, ragionando non più sugli incrementi ma su eventuali riduzioni di disponibilità, si può affermare che non costa nulla ridurre (di una quantità sufficientemente piccola) una risorsa il cui prezzo ombra è 0. Tuttavia, se la riduzione procede, oltre una certa soglia la risorsa diviene scarsa e il suo prezzo ombra diventa positivo. Banalizzando, si può pensare all'ossigeno dell'aria come risorsa per rendere operativo un processo di combustione. La sua disponibilità è praticamente illimitata: ma in un ambiente extraterrestre, come ad es. su di un pianeta privo d'atmosfera, occorrerebbe procurarsi tale risorsa ed il suo prezzo ombra risulterebbe presumibilmente positivo.

Analoghe interpretazioni si possono dare nel caso degli altri esempi classici dei problemi di programmazione lineare.

Nel problema di trasporto, ai percorsi (i, j) effettivamente utilizzati per il trasporto della merce ($x_{ij} > 0$) corrispondono relazioni del tipo $c_{ij} = u_i + v_j$, cioè la somma (algebrica!) dei prezzi di vendita ed acquisto della merce da parte della ditta di trasporto coincide con il costo c_{ij} che sarebbe sostenuto dall'azienda che produce e commercializza lo stesso, mentre se nel duale si ha $c_{ij} > u_i + v_j$, allora sul percorso (i, j) non si trasporta alcuna merce. Infine, nel problema della dieta, è positivo il prezzo ombra delle pillole relative ai fattori nutritivi scarsi, mentre un fattore 'abbondante' ha prezzo ombra nullo.

9. La soluzione del problema duale mediante le condizioni di complementarità.

Una volta risolto un problema (primale), le condizioni di complementarità consentono di risolvere anche il suo duale in quanto in corrispondenza della soluzione ottima del primale:

- per ogni vincolo non soddisfatto come eguaglianza, cioè non attivo, si può stabilire che è nulla la variabile duale associata;
- per ogni componente positiva della soluzione stessa, il vincolo corrispondente del duale vale come eguaglianza.

Eliminati dal sistema dei vincoli del duale quelli di diseguaglianza stretta, i restanti vincoli attivi formano un sistema di equazioni che, salvo casi particolari, è quadrato e ammette una sola soluzione, ottenibile con i consueti metodi.

Si consideri il seguente esempio :

$$\begin{aligned} \max \quad & 2x_1 + 3x_2 \\ \text{s.t.} \quad & x_1 + 4x_2 \le 24, \\ & 3x_1 + 2x_2 \le 22, \\ & x_1 - x_2 \le 4, \\ & x_1, x_2 \ge 0. \end{aligned}$$

Il problema è facilmente risolubile per via grafica e la soluzione ottima è data dal vettore $\mathbf{x^*} = (4, 5)$. Nella soluzione ottima entrambe le componenti sono positive e la soluzione stessa soddisfa i primi due vincoli come eguaglianza, mentre il terzo vincolo non è attivo. Si consideri ora il problema duale:

$$\begin{aligned} \min \quad & 24y_1 + 22y_2 + 4y_3 \\ \text{s.t.} \quad & y_1 + 3y_2 + y_3 \ge 2, \\ & 4y_1 + 2y_2 - y_3 \ge 3, \\ & y_1, y_2, y_3 \ge 0. \end{aligned}$$

Questo problema non si presta ad una risoluzione grafica. Sulla base del teorema di complementarità, tuttavia, si può dedurre che:

- entrambi i vincoli, in corrispondenza della soluzione ottima devono essere soddisfatti come eguaglianza (perché nella soluzione ottima del primale sia x_1 sia x_2 sono positive);
- la variabile y_3 è nulla, perché il terzo vincolo del primale non è attivo (è soddisfatto come diseguaglianza).

Pertanto il sistema dei vincoli del duale diventa:

$$y_1 + 3y_2 = 2,$$

$$4y_1 + 2y_2 = 3,$$

che risolto (e tenuto conto che $y_3 = 0$) fornisce la soluzione ottima del duale

$$\mathbf{y^*} = (1/2,\ 1/2,\ 0).$$

E' immediato verificare che il valore ottimo della funzione oggetto, sia per il primale sia per il duale è $z^* = 23$.

2.2.10 La forma matriciale e la soluzione del problema duale.

Dato il problema in forma standard

$$\max \mathbf{cx} \quad \text{s.t.} \quad \mathbf{Ax = b},\ \mathbf{x \geq 0},$$

consideriamo il relativo problema duale:

$$\min \mathbf{yb} \quad \text{s.t.} \quad \mathbf{yA \geq c},\ \mathbf{y} \text{ libero}.$$

Si può dimostrare che se in corrispondenza di una certa base B si ha la soluzione ottima del primale $\mathbf{x} = (\mathbf{x}_B = B^{-1}\mathbf{b}, \mathbf{0})$, allora il vettore

$$\mathbf{y°} = \mathbf{c}_B B^{-1}$$

è soluzione ottima del problema duale.

Infatti, se consideriamo la partizione $A = (B \mid N)$ della matrice (iniziale) dei coefficienti dei vincoli, dove B è la sottomatrice di base corrispondente alla soluzione ottima del primale, si può scrivere:

$$\mathbf{y°}A = \mathbf{y°}\ (B \mid N) = (\mathbf{c}_B B^{-1} B,\ \mathbf{c}_B B^{-1} N) = (\mathbf{c}_B,\ \mathbf{c}_B B^{-1} N).$$

Il sistema dei vincoli del duale richiede che sia

$$(\mathbf{c}_B,\ \mathbf{c}_B B^{-1} N) \geq (\mathbf{c}_B,\ \mathbf{c}_N).$$

E' evidente che quest'ultima condizione equivale a

$$\mathbf{c}_B B^{-1} N \geq \mathbf{c}_N,$$

e questa diseguaglianza è certamente verificata in corrispondenza della sottomatrice B perché non è altro che la condizione di negatività dei coefficienti di guadagno ridotto del primale corrispondenti alla soluzione ottima. Si può affermare quindi, per ora, che il vettore **y°** è soluzione ammissibile per il duale, in quanto ne soddisfa il sistema dei vincoli. Poiché in corrispondenza di **y°** la funzione oggetto del duale assume il valore

$$\mathbf{y^\circ b} = \mathbf{c}_B B^{-1} \mathbf{b},$$

che coincide con il valore z ottimo per il problema primale, si può affermare anche che $\mathbf{y^\circ}$ è soluzione ottima del duale.

Così facendo si dimostra anche l'enunciato d) delle proprietà che legano primale e duale in quanto partendo dalla conoscenza della soluzione del primale è stata costruita (e quindi, a maggior ragione, esiste) la soluzione ottima del duale.

Per quanto riguarda l'altro enunciato non ancora dimostrato a), basta aggiungere a quanto appena visto che essendo i valori della funzione oggetto di primale e duale limitati (rispettivamente, superiormente per il problema di massimo, inferiormente per quello di minimo) ambedue devono avere ottimo finito e, sulla base della dimostrazione del punto d), i due valori ottimali della funzione oggetto necessariamente devono coincidere.

Infine si può osservare che, a volte senza particolari accorgimenti, altre mediante l'aggiunta nella tabella iniziale di una matrice identica, il procedimento del simplesso (primale), descritto nella prima sezione del capitolo, consente di leggere direttamente sulla tabella finale anche la soluzione ottima del duale o di poterla ricavare con pochi calcoli.

Dai coefficienti finali di guadagno ridotto

$$- \mathbf{c}_N + \mathbf{c}_B B^{-1} N,$$

sommando il vettore $\mathbf{c}_N$, si può ovviamente ricavare il vettore $\mathbf{c}_B B^{-1} N$. Da quest'ultimo, però, non è detto si possa sempre ricavare il vettore $\mathbf{c}_B B^{-1}$: ciò dipende dalla matrice N. Il caso più favorevole è quello in cui N è una matrice identica (N è in genere rettangolare); si è poi ancora più agevolati se il vettore $\mathbf{c}_N$ è il vettore nullo. In tale situazione, nelle posizioni sottostanti N la soluzione del duale può essere letta immediatamente.

Se ciò non si verifica, si può ricorrere all'aggiunta di una matrice identica, analogamente a quanto capita nel procedimento delle due fasi, ponendo nell'ultima riga, in corrispondenza alla stessa, un vettore nullo. Nelle successive iterazioni le colonne di quest'ultima matrice non sono mai fatte entrare in base. Nella tabella finale, in corrispondenza della matrice aggiunta si ottiene l'inversa della sottomatrice di base B ottima mentre nell'ultima riga si ha la soluzione del duale.

E' interessante analizzare il caso dei problema di massimo in forma canonica e come opera in questi l'aggiunta delle variabili slack.

Dal problema

$$\max \mathbf{cx} \qquad \text{s.t.} \quad \mathbf{Ax} \le \mathbf{b}, \ \mathbf{x} \ge \mathbf{0},$$

si passa a

$$\max \mathbf{cx} + \mathbf{0s} \qquad \text{s.t.} \quad \mathbf{Ax} + \mathbf{Is} = \mathbf{b}, \ (\mathbf{x}, \mathbf{s}) \ge \mathbf{0}.$$

Il duale di quest'ultimo è

$$\min \mathbf{yb} \qquad \text{s.t.} \quad \mathbf{yA} \ge \mathbf{c}, \ \mathbf{yI} \ge \mathbf{0}, \quad \mathbf{y} \text{ libero}.$$

Si vede immediatamente come, a tutti gli effetti, il duale sia un problema di minimo in forma canonica, perché il secondo insieme di vincoli richiede che il vettore **y** sia non negativo, e questo prevale ovviamente sul fatto che le variabili siano libere.

In corrispondenza della soluzione ottima del primale potranno essere in base solo colonne di A oppure anche colonne della matrice corrispondente alle variabili slack, I.

Nel primo caso (con le variabili slack tutte fuori base e quindi nulle), in corrispondenza della matrice identica iniziale delle variabili slack, nella tabella finale comparirà l'inversa della sottomatrice di base ottima. Nelle corrispondenti posizioni sull'ultima riga compare la soluzione del duale.

Infatti dalla tabella iniziale:

B	N	I	**b**
$-\mathbf{c}_B$	$-\mathbf{c}_N$	**0**	0

moltiplicando per la matrice

B^{-1}	**0**

$\mathbf{c}_B B^{-1}$	1

si ottiene la tabella (finale)

I	$B^{-1} N$	B^{-1}	$B^{-1}\ \mathbf{b}$
0	$-\mathbf{c}_N + \mathbf{c}_B B^{-1} N$	$\mathbf{c}_B B^{-1}$	$\mathbf{c}_B B^{-1} \mathbf{b}$

In quest'ultima tabella la soluzione del duale quindi può essere letta subito.

Lo stesso si può dire se in corrispondenza della soluzione ottima sono in base sia colonne di variabili x_j sia colonne di variabili slack (cioè sia colonne che inizialmente erano in A che colonne di I, e verosimilmente non tutte le colonne di I!). Infatti, in tale secondo caso, supponendo per semplicità che non vi sia degenerazione, se una variabile slack s_i è in base (e assume valore positivo), il corrispondente vincolo nella formulazione canonica è sotto forma di diseguaglianza e quindi, per le condizioni di complementarità, la variabile corrispondente del duale è nulla, cioè $y_i = 0$, e in effetti è nullo il coefficiente di guadagno ridotto corrispondente a s_i. Se invece la slack è fuori base, il coefficiente di guadagno ridotto sarà in generale > 0 e risulta una delle componenti (non nulle) della soluzione del duale.

2.2.11 L'analisi di sensitività.

L'analisi di sensitività studia come varia la soluzione ottima di un problema di programmazione lineare al variare dei dati in ingresso. E' di particolare interesse, in pratica, individuare qual è l'entità delle variazioni in input che implicano dei cambiamenti nella soluzione (ottima) di un problema o nel valore ottimo della funzione oggetto. In maniera equivalente, si tratta di stabilire fino a che punto si può cambiare un dato senza che la soluzione ottima ne risenta.

L'analisi acquista particolare valore per il fatto che, generalmente, i parametri di un problema (i coefficienti della funzione oggetto, gli elementi della matrice dei vincoli, i termini noti degli stessi, ma anche la stessa presenza di certe variabili o la specificazione dei vincoli) non sono necessariamente certi o possono essere soggetti a rettifiche. Ben si comprende quindi l'importanza di precisare i margini di 'tenuta' di una soluzione.

Concretamente può esserci incertezza simultanea su più elementi, ma per caratterizzare i diversi casi, si ipotizza che cambi un solo tipo di dato per volta. Il caso più generale è lasciato all'analisi dell'istanza specifica.

Inoltre è più immediato visualizzare le situazioni facendo riferimento a problemi che si possono illustrare e risolvere graficamente, cioè con vincoli di diseguaglianza e due sole variabili. Questo però per alcune tipologie di variazioni è scarsamente significativo o addirittura fuorviante e sarà segnalato di volta in volta.

a) Variazione dei termini noti nel sistema dei vincoli:

$$\mathbf{b} \rightarrow \mathbf{b} + \Delta\mathbf{b}.$$

Nella interpretazione economica tipica di un problema di programmazione della produzione con risorse limitate, e quindi con un problema in forma canonica di massimo, si tratta di un cambiamento nella disponibilità delle risorse. E' allora immediato concludere che se i cambiamenti si riferiscono a risorse abbondanti, per incrementi Δb_i limitati o comunque fino a che una risorsa non diventi scarsa, non vi sono effetti, né sulla soluzione ottima né sul valore ottimo della funzione oggetto.

Se invece una delle risorse è scarsa, la politica di produzione generalmente cambia e i quantitativi prodotti aumentano o diminuiscono secondo il segno della variazione.

Da un punto di vista formale, facendo riferimento alla forma standard, tipica del metodo del simplesso, si può esprimere facilmente in termini matematici la variazione nella soluzione ottima e nel valore della funzione oggetto fino a che non cambia la base ottima. In tal caso, la soluzione

$$\mathbf{x} = (\mathbf{x}_B = B^{-1}\,\mathbf{b},\, \mathbf{0})$$

diventa

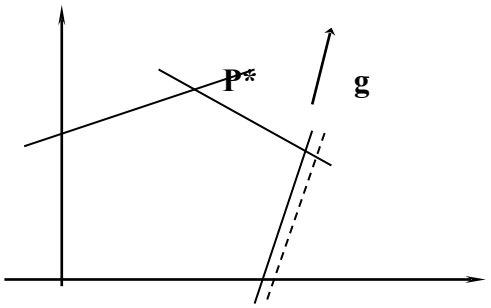
$$\mathbf{x} = (\mathbf{x}_B = B^{-1}\,\mathbf{b} + B^{-1}\,\Delta\mathbf{b},\, \mathbf{0}),$$

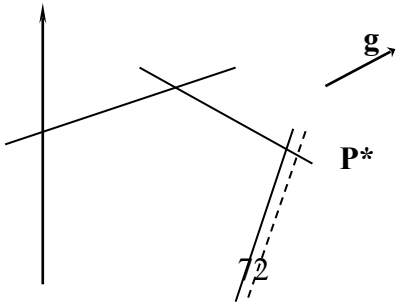
ed il valore della funzione oggetto cambia di conseguenza. Quest'ultima variazione può essere espressa anche facendo ricorso al problema duale. Infatti, fintantoché non cambia la base ottima, la soluzione del duale (che non dipende dal vettore **b**, ma solo dai coefficienti $\mathbf{c}_B$ della funzione oggetto del primale) rimane invariata. Il valore ottimo della funzione oggetto dopo la variazione $\mathbf{b} \rightarrow \mathbf{b} + \Delta\mathbf{b}$ risulta

$$z' = \mathbf{y}\,(\mathbf{b} + \Delta\mathbf{b}),$$

con una variazione quindi $\Delta z = \mathbf{y}\,\Delta\mathbf{b}$. Da questa impostazione si può dedurre ancora quanto stabilito all'inizio di questo paragrafo e cioè che se la variazione si riferisce al cambiamento di disponibilità di una risorsa abbondante (il cui prezzo ombra è nullo), fino a che la base ottima rimane la stessa non si hanno cambiamenti nella funzione oggetto.

Da un punto di vista grafico, in due variabili, il cambiamento in questione dà luogo ad uno spostamento di rette collegate ai vincoli parallelamente a se stesse, come si può vedere nella figura che segue, in due casi, nei quali si hanno oppure non si hanno rispettivamente conseguenze sulla soluzione.



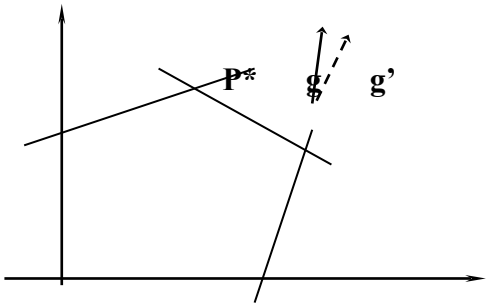


Si nota come lo spostamento di un vincolo comporta che cambino anche tutti i vertici situati nelle intersezioni di questo vincolo con i rimanenti. La cosa rimane senza effetto pratico se si tratta di un vincolo non attivo e lo spostamento è sufficientemente piccolo. Viceversa, se la variazione si verifica per un vincolo attivo (o anche, in un vincolo non attivo, ma in maniera consistente), cambia anche la soluzione ottima (ma, per piccole variazioni, non cambiano le variabili che nella soluzione ottima sono positive, cioè non cambia la base ottima).

b) Variazione dei coefficienti nella funzione oggetto :

$$\mathbf{c} \rightarrow \mathbf{c} + \Delta\mathbf{c}.$$

Questa variazione può essere considerata come la 'duale' della precedente: cambiano in effetti i termini noti dei vincoli nel duale, per cui si possono fare considerazioni analoghe a quelle del caso precedente riferite, appunto al problema duale. Da un punto di vista grafico, questa variazione comporta un cambiamento nella inclinazione delle rette (piani, iperpiani) parallele che rappresentano le curve di isolivello per i valori della funzione oggetto. Una diversa inclinazione può essere irrilevante in quanto il vertice corrispondente alla soluzione ottima non cambia, come si può vedere nella figura seguente:



Da un punto di vista formale, il cambiamento di un coefficiente c_j fa variare nella tabella finale del simplesso il vettore dei coefficienti di guadagno ridotto (cambiati di segno):

$$(\mathbf{0}, -\mathbf{c}_N + \mathbf{c}_B B^{-1} N).$$

Se tutte le componenti di questo vettore rimangono positive, la soluzione ottima rimane immutata. Rimane invariato poi anche il valore della funzione oggetto se l'indice j è quello di una variabile fuori base.

E' evidente però che se la variazione supera un determinato ammontare, il vertice ottimale può cambiare (ed allora vi sarà, in generale, anche una precisa variazione per la quale le soluzioni ottime sono in numero infinito).

Fino a che la soluzione ottima rimane inalterata, è facile individuare il cambiamento nel valore della funzione oggetto, che sarà dato da:

$$\Delta z = \Delta \mathbf{c}\, \mathbf{x}.$$

c) Variazione di coefficienti della matrice A

$$a_{ij} \rightarrow a_{ij} + \Delta a_{ij}$$

Terzo caso di variazione di un coefficiente è costituito dal cambiamento di **un coefficiente in un vincolo**: geometricamente ciò comporta una diversa inclinazione della retta (piano, iperpiano) rappresentativa del vincolo stesso.

L'effetto è intuibile : se il vincolo non è attivo, e la variazione è sufficientemente piccola, allora la soluzione ottima non ne risulta modificata, mentre, se il vincolo è attivo, o diventa tale in seguito alla variazione, si può avere un cambiamento nella soluzione: in questa possono cambiare solo i valori delle variabili in base oppure può cambiare la base stessa.

d) Variazioni nel numero dei vincoli e delle variabili.

Vi sono infine altri due tipi di variazioni nei dati di un problema di Programmazione Lineare che sono prese in considerazione nella analisi di sensitività. Si tratta dell'inserimento di un vincolo aggiuntivo, da un lato, e dell'inserimento di una nuova variabile decisionale, dall'altro.

Tali variazioni sono una duale dell'altra. L'aggiunta di un vincolo può non avere alcun effetto (e ciò succede se il vincolo è non attivo) come pure può portare ad un cambiamento sia nella soluzione ottima che nel valore ottimale della funzione oggetto. Se si aggiunge una variabile, si tratta di vedere se nella soluzione ottima tale variabile è o no in base: se rimane fuori base, l'effetto è nullo (sia per quanto riguarda tutti i vincoli che per la funzione oggetto). Se invece la variabile è in base, allora si può, con opportune rettifiche nell'ultima tabella del simplesso, individuare la soluzione ottima, senza dover riavviare il simplesso stesso a partire dall'inizio.

12. Il metodo del simplesso duale.

Il metodo del simplesso primale, che è stato sviluppato nella prima parte di questo capitolo, sostanzialmente si pone l'obiettivo di giungere ad una tabella finale nella quale il sistema dei vincoli abbia le seguenti proprietà:

- nella matrice dei coefficienti vi è una sottomatrice identica m×m;
- i termini noti sono tutti positivi (o non negativi);
- nella riga della funzione oggetto, i coefficienti sono tutti ≥ 0.

(a meno che non si giunga alla conclusione che non esiste un ottimo finito).

Considerando i requisiti appena elencati dal punto di vista soltanto formale, pragmaticamente si può affermare che, data una tabella iniziale, qualsiasi tecnica che consenta di soddisfare le tre regole viste sopra è idonea alla risoluzione di un problema di PL. In particolare, non è certamente tassativo che durante il procedimento i termini noti siano sempre tutti non negativi ($b_i \geq 0$): basta che questo si verifichi nella tabella finale. In altri termini, non è necessario che le soluzioni intermedie nelle varie iterazioni siano tutte ammissibili: basta che sia ammissibile la soluzione finale (e ottima).

Naturalmente occorre sempre che vi sia nella procedura una garanzia di 'finitezza', cioè la sicurezza che in un numero finito di passi si giunga alla soluzione ottima o si possa decidere che tale soluzione non esiste.

In alcune situazioni si è visto che è relativamente difficile avere una soluzione ammissibile di partenza per il problema (primale) che si sta studiando. Invece si può ricavare facilmente una soluzione non ammissibile per il primale, alla quale corrisponde nella tabella iniziale una soluzione ammissibile per il duale. E' il caso dei problemi canonici di minimo (con $\mathbf{b} \geq \mathbf{0}$) nei quali l'introduzione delle variabili surplus fa comparire una matrice identica cambiata di segno e, a sua volta, un cambiamento di segno in tutti i vincoli, dà luogo ad una soluzione di base iniziale a componenti tutte negative o nulle.

Si consideri il seguente esempio:

$$\begin{aligned} & \min \; 5x_1 + 3x_2 + 4x_3 \\ \text{s.t.} \quad & x_1 + 2x_2 + 2/3\,x_3 \geq 7, \\ & 2x_1 + 3x_2 + x_3 \geq 9, \\ & x_1, x_2, x_3 \geq 0, \end{aligned}$$

che, introducendo le variabili scarto (surplus), trasformando la funzione oggetto da min a max e cambiando di segno i vincoli, diventa:

$$\begin{aligned} & \max \; -5x_1 - 3x_2 - 4x_3 \\ \text{s.t.} \quad & -x_1 - 2x_2 - 2/3\,x_3 + s_1 = -7, \\ & -2x_1 - 3x_2 - x_3 + s_2 = -9, \\ & x_1, x_2, x_3, s_1, s_2 \geq 0. \end{aligned}$$

Nel sistema dei vincoli, che è in forma canonica, si può riconoscere la soluzione di base non ammissibile data dal vettore **x** = (0 0 0 -7 -9), mentre la corrispondente tabella del simplesso risulterebbe:

-1	-2	-2/3	1	0	-7
-2	-3	-1	0	1	-9
5	3	4	0	0	0

Questa tabella sarebbe addirittura quella conclusiva se la soluzione fosse ammissibile perché gli elementi nell'ultima riga sono positivi (i coefficienti di guadagno ridotto sono negativi). Si noti poi che in corrispondenza della matrice identica, gli elementi nell'ultima riga, entrambi nulli, corrispondono alla soluzione **y** = (0 0) ammissibile per il problema duale, con valore della funzione oggetto (sia per il duale che per la soluzione non ammissibile del primale) eguale a 0.

Si tenga presente che il duale si può scrivere come segue:

$$\max\ 7\,y_1 + 9\,y_2$$

s.t.

$$\begin{aligned} y_1 + 2\,y_2 &\le 5,\\ 2\,y_1 + 3\,y_2 &\le 3,\\ 2/3\,y_1 + y_2 &\le 4,\\ y_1,\ y_2 &\ge 0, \end{aligned}$$

e quindi, evidentemente, ammette come soluzione ammissibile di base il vettore nullo.

Ci si può chiedere a questo punto se non sia possibile, con un'adeguata scelta dei pivot, ottenere da questa tabella la soluzione ottima. In effetti, se un termine noto è negativo, la scelta di un pivot negativo sulla stessa riga porta nella nuova soluzione ad una componente positiva. Si tratta allora di vedere se e come tale scelta può anche essere effettuata in modo che i coefficienti nell'ultima riga rimangano positivi (o nulli). Per stabilire questo, notiamo come si modificano i coefficienti stessi dell'ultima riga, sulla base delle formule generali che danno i nuovi coefficienti dopo un'iterazione di simplesso:

$$a'_{ij} = a_{ij} - a_{iq} * a_{pj} / a_{pq}.$$

che, nel caso specifico, indicando convenzionalmente con h_i i coefficienti di guadagno ridotto cambiati di segno, scriveremo come:

$$h'_j = h_j - h_q * a_{pj}/a_{pq}.$$

Tenendo conto che $a_{pq} < 0$ e che si vuole che sia $h'_j \ge 0$, dovrà essere

$$h_j \ge h_q * a_{pj}/a_{pq}. \qquad (*)$$

Ora, se $a_{pj} \geq 0$ la diseguaglianza è sempre verificata, (si tenga presente che $h_k \geq 0$ per ogni k), mentre se risulta $a_{pj} < 0$, la (*) equivale alla relazione

$$h_j/a_{pj} \leq h_q/a_{pq}$$

oppure, passando ai valori assoluti:

$$\left| h_j/a_{pj} \right| \geq \left| h_q/a_{pq} \right| .$$

Se ne può ricavare la seguente regola, che sta alla base del procedimento del **simplesso duale**:

il pivot va scelto in una riga ove il termine noto è negativo e deve essere a sua volta un elemento negativo; in presenza di più elementi $a_{pq} < 0$ sulla stessa riga p-esima dove si intende scegliere il pivot, per mantenere la non negatività dei coefficienti della funzione oggetto – cambiati di segno – il pivot deve essere scelto in corrispondenza dell'elemento per cui risulta minimo il valore assoluto del rapporto h_j/a_{pj}.

Con riferimento all'esempio che si sta studiando, tenuta presente la tabella

-1	-2*	-2/3	1	0	-7
-2	-3	-1	0	1	-9
5	3	4	0	0	0

se si intende scegliere il pivot sulla prima riga, occorre calcolare i rapporti 5/1; 3/2; 4/(2/3), che danno, rispettivamente, 5; 1.5; 6 e pertanto il pivot deve essere l'elemento $a_{12} = -2$, in corrispondenza del quale il rapporto è minimo (in valore assoluto). La nuova tabella risulta:

1/2	1	1/3	-1/2	0	7/2
-1/2	0	0	-3/2	1	3/2
7/2	0	3	3/2	0	-21/2

che fornisce immediatamente la soluzione ottima del problema

$$x_1 = 0;\ x_2 = 7/2;\ x_3 = 0.$$

Per quanto riguarda la funzione oggetto, si può notare che effettuando sempre l'operazione di pivot in una riga dove il termine noto è negativo, e in una colonna ove è positivo il coefficiente ridotto h_i, il valore della stessa funzione oggetto diminuisce ad ogni iterazione, e quindi non occorrono particolari ulteriori accorgimenti per la scelta della riga del pivot (è sufficiente che il termine noto sia negativo).

Infatti, il valore z della funzione oggetto si aggiorna secondo la formula:

$$z' = z - h_q * b_p/a_{pq}$$

dove z è il valore della iterazione precedente, b_p (≤ 0) è il termine noto (non positivo) sulla riga del pivot ed a_{pq} (< 0) è il pivot stesso.

Si tenga presente che, nella impostazione che si è qui data alle tabelle del simplesso, il primale in forma standard che si risolve è un problema di massimo e quindi, essendo ad ogni iterazione, tranne che alla fine, in presenza di una soluzione ammissibile per il duale, che è di minimo, il valore della funzione oggetto non può che decrescere verso il valore (comune) ottimale del problema di massimo.

Infine è il caso di osservare come nell'esempio precedente si sarebbe potuta scegliere come riga del pivot la seconda riga (il pivot allora è l'elemento $a_{22} = -3$) dopodiché si può costatare facilmente che rimane un termine noto negativo: una seconda iterazione del simplesso conduce poi, comunque, alla stessa soluzione che si è trovata sopra.

3. Complementi

2.3.1 La programmazione lineare a più obiettivi.

E' raro che nella vita di tutti i giorni quello che si può esprimere come un problema di programmazione matematica (ottimizzazione di una funzione oggetto in presenza di vincoli sulle variabili di decisione) abbia effettivamente un obiettivo unico. La massimizzazione di un guadagno o la minimizzazione di un costo sono semplificazioni che non hanno corrispondenza nella realtà, nemmeno nella pratica aziendale. Generalmente con la scelta di una politica il decisore deve soddisfare più esigenze, a volte anche contrastanti, che possono dar luogo a obiettivi differenti. Inoltre, si possono presentare situazioni nelle quali il decisore non è unico: si tratta di un gruppo di persone o enti, che, in presenza dello stesso insieme di vincoli, hanno tuttavia differenti funzioni oggetto.

La **programmazione a più criteri** ha lo scopo di dettare regole razionali per situazioni di questo tipo. Un problema di programmazione multicriterio prevede più funzioni oggetto (che si possono scrivere, senza minor generalità, tutte come funzioni di massimo) e un insieme di vincoli.

Se tutte le funzioni oggetto e i vincoli sono lineari si ha a che fare con un problema di programmazione a più obiettivi lineare.

Formalmente scriveremo un problema di questo tipo come segue:

$$\begin{aligned} \max\ & f_1(\mathbf{x}), \quad \max f_2(\mathbf{x}), \quad \ldots, \quad \max f_k(\mathbf{x}), \\ \text{s.t.}\ & A\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned}$$

Per un problema di questo tipo valgono i consueti concetti di regione ammissibile e soluzioni ammissibili di base. Il fatto è che generalmente i criteri sono contrastanti e quindi un vettore che sia soluzione ottima sulla base di un criterio non lo è più se valutato con uno dei rimanenti. Si tratta allora di adottare una soluzione di compromesso, sulla cui scelta ultimo giudice non può che essere il decisore. Non è in ogni caso un buon compromesso una soluzione rispetto alla

quale ve ne siano altre che permettono di acquisire un risultato migliore per tutti gli obiettivi, cioè soluzioni che la dominano. Pertanto compito dell'analista è certamente quello di mettere a disposizione del decisore soluzioni non dominate da altre.

Per precisare la situazione ed alcuni concetti fondamentali, premettiamo alcune definizioni.

Dato un problema di programmazione lineare a più obiettivi si dice che una soluzione ammissibile **x** è dominata da un'altra soluzione ammissibile **x'** se vale la relazione

$$f_i(\mathbf{x}) \leq f_i(\mathbf{x'}) \qquad (i = 1, 2, \ldots, k),$$

ed esiste (almeno) un indice i per il quale la relazione vale in senso stretto.

E' evidente che data invece una soluzione non dominata, non è possibile rispetto a questa apportare delle variazioni per cui migliorano tutti gli obiettivi: se uno o più di essi migliora, ne esiste almeno uno per il quale il valore della funzione oggetto diminuisce.

Le **soluzioni non dominate** si dicono anche **soluzioni Pareto-ottime**.
E' interessante nei casi concreti evidenziare queste ultime: da un punto di vista geometrico, la cosa è agevole se il problema contempla due variabili di decisione e due obiettivi, altrimenti la cosa diventa problematica.

Con due variabili e due obiettivi, in effetti, si può dare una raffigurazione geometrica sia delle soluzioni ammissibili che dei valori conseguibili con le funzioni oggetto. Si parla rispettivamente di rappresentazione nello **spazio delle soluzioni** e nello **spazio degli obiettivi**. Considerato ad es. il seguente problema biobiettivo (che potremmo definire in forma 'canonica'):

$$\max \{f_1(\mathbf{x}) = c_1x_1 + c_2x_2\}, \quad \max \{f_2(\mathbf{x}) = c'_1x_1 + c'_2x_2\},$$

$$\text{s.t.} \qquad \mathbf{a}_1x_1 + \mathbf{a}_2x_2 \leq \mathbf{b},$$

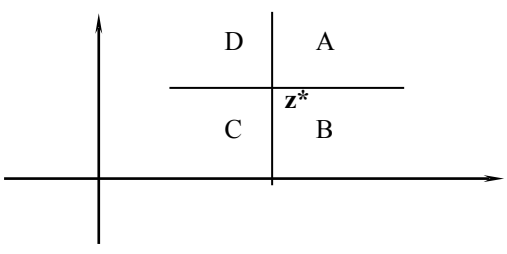
$$x_1, \; x_2 \geq 0,$$

si può innanzi tutto rappresentare su un primo piano cartesiano la Regione Ammissibile: questa, salvo casi particolari, è un poligono dotato di punti interni. Successivamente ad ogni soluzione ammissibile **x** si può fare corrispondere in un secondo piano cartesiano la coppia di valori delle funzioni oggetto, $(f_1(\mathbf{x}), f_2(\mathbf{x}))$. Le coppie così ottenute formeranno una figura geometrica che, se le due funzioni oggetto hanno gradienti con direzioni differenti, cioè se il determinante

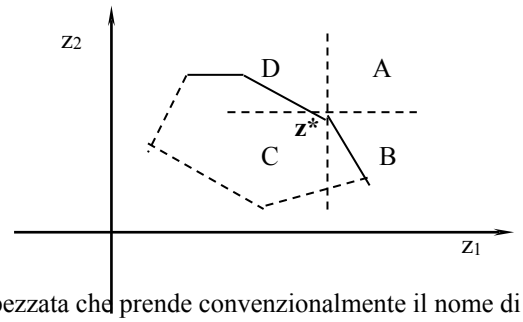
$$\begin{vmatrix} c_1 & c_2 \\ c'_1 & c'_2 \end{vmatrix}$$

è diverso da zero, è dotata di punti interni, così come la Regione Ammissibile di partenza ed è quindi un altro poligono P. Per di più, ai vertici della Regione Ammissibile corrispondono (biunivocamente) vertici di P.

Consideriamo un punto $\mathbf{z}^* \in P$ e la corrispondente soluzione $\mathbf{x}^*$. Le rette per $\mathbf{z}^*$ parallele agli assi dividono il piano cartesiano in quattro regioni:



Le (eventuali) soluzioni **x** alle quali corrispondono valori **z** degli obiettivi nell'angolo A sono soluzioni che dominano la soluzione **x***. L'angolo C corrisponde invece alle soluzioni dominate da **x***, mentre infine gli angoli B e D sono quelli che corrispondono a soluzioni che non dominano né sono dominate da **x***: si parla di soluzioni **non confrontabili** con **x***. Va rilevato che la relazione di dominanza tra soluzioni ammissibili è interpretabile geometricamente solo nello **spazio degli obiettivi**, per cui operativamente è necessario poi risalire da quest'ultimo allo spazio delle soluzioni. Le soluzioni Pareto-ottime, cioè le soluzioni non dominate, sono quelle che nello spazio degli obiettivi non presentano alcuna altra soluzione che si vada a collocare nell'angolo A (vedi la figura che segue).



Esse formano una spezzata che prende convenzionalmente il nome di **frontiera di nord-est**.

2.3.2 I software di PL: C-plex, LINDO ed Excell.

I problemi di programmazione lineare possono essere risolti con l'utilizzo di strumenti software relativamente diversi, più o meno sofisticati. Per problemi con dimensioni contenute sono da segnalare LINDO, che sarà qui presentato, come anche il software di ottimizzazione presente nelle versioni più recenti di Excell, del quale verranno accennate in questo paragrafo le istruzioni fondamentali; per dimensioni maggiori va segnalato C-plex.

E' il caso di osservare che nella pratica l'aspetto critico più rilevante da affrontare sta nelle dimensioni del programma lineare. Queste, in applicazioni economiche come le tabelle input-output di un sistema economico, possono arrivare comunemente a migliaia di vincoli e decine di migliaia di variabili. Si tratta allora di saper gestire ingresso e uscita dei dati, evitando di immagazzinarli nella memoria centrale tutti contemporaneamente.

Il software LINDO consente di impostare qualsiasi problema di programmazione lineare (compatibilmente con il numero delle variabili e dei vincoli che la particolare versione del software consente) introducendo vincoli e variabili in maniera molto semplice.

A titolo esemplificativo delle tabelle visualizzate da LINDO, consideriamo il problema di programmazione che segue:

$$\max \quad 2x + 3y - z + u$$

con i vincoli

$$\begin{aligned} x + y - z - u &\le 10, \\ 3x + 2y + z + 3u &\le 5, \\ x + 2z - 4u &\le 4, \\ y - z &\le 2, \\ 3x - y + z &\le 1.5, \end{aligned}$$

oltre ai vincoli di non negatività.

La risposta del programma è costituita dalla seguente tabella :

LP OPTIMUM FOUND AT STEP 3

	OBJECTIVE FUNCTION VALUE
1)	6.666667

VARIABLE	VALUE	REDUCED COST
X	0.000000	0.000000
Y	2.333333	0.000000
Z	0.333333	0.000000
U	0.000000	1.000000

ROW	SLACK OR SURPLUS	DUAL PRICES
2)	8.000000	0.000000
3)	0.000000	0.666667
4)	3.333333	0.000000
5)	0.000000	1.666667
6)	3.500000	0.000000

NO. ITERATIONS= 3

Dalla tabella si ricava la soluzione ottima (x=0, y=7/3, z=1/3, u=0) ed il corrispondente valore della funzione oggetto 20/3 = 6.666667: la soluzione stessa si ottiene dopo tre iterazioni dell'algoritmo del simplesso. Si vede anche come in corrispondenza alla soluzione ottima siano attivi il secondo ed il quarto vincolo (quelli corrispondenti alle righe 3 e 5) per i quali le variabili scarto (in questo caso slack) sono nulle, mentre le altre slack sono rispettivamente, 8 (primo vincolo) 10/3 (terzo vincolo) ed infine 7/2 (nel quinto vincolo).

Si può ricavare dalla tabella anche la soluzione del duale (DUAL PRICES): in esso la funzione oggetto - di minimo - è

$$\min \quad 10a + 5b + 4c + 2d + 3/2e$$

e la soluzione ottima è

$$a=0,\ b=2/3,\ c=0,\ d=5/3,\ e=0$$

in corrispondenza alla quale la funzione oggetto ha ancora valore 20/3, come nel primale.

Se si ha a disposizione una versione (relativamente) recente di **Excell**, è possibile risolvere problemi di PL di dimensione contenuta in maniera rapida, rinunciando tuttavia alle informazioni sulla soluzione del duale.

Aperto il foglio elettronico, si tratta innanzitutto di impostare alcune celle da riservare alle variabili in gioco; altre celle dedicate ai vincoli ed infine una cella per la funzione oggetto. Una maniera di procedere è la seguente:

- su tante celle quante sono le variabili scrivere le variabili stesse: ad es., x in A1, y in A2;
- su una cella successiva impostare '=' seguito dalla funzione oggetto, espressa tuttavia **non** in funzione delle variabili originarie (ad es., x e y), ma con le denominazioni delle relative celle impostate precedentemente (es.: **= 3*A1+5*A2**); sulla cella compare la scritta: '= VALORE!';
- su ogni cella dedicata ad un vincolo, scrivere il primo membro dello stesso preceduto da =; ad es., il vincolo $3x-2y \leq 6$, in questa fase viene inserito come '=3*A1-2*A2'; la specificazione '≤ 6' verrà fatta in seguito;
- dal **Menu** aprire **Strumenti** e quindi **Risolutore**;
- nella finestra che si apre, si impostano: il tipo di funzione oggetto (Max/Min); le variabili in gioco (**Modificando**: si indicano le celle delle variabili) infine il tipo ($\leq$ o $\geq$) dei vincoli ed i termini noti (aprendo un'altra finestra alla voce '**aggiungi**'); si imposta da **opzioni** la **non negatività** delle variabili e la **linearità del modello** (supponi: modello lineare); si dà poi avvio alla soluzione con '**risolvi**'.

Ricompare la pagina di foglio elettronico iniziale nella quale nelle celle 'variabili' compaiono le componenti della soluzione, mentre nella cella originariamente dedicata alla funzione oggetto compare il valore z*.

E' consigliabile, per non perdere di vista il significato delle celle, sin dalla fase di impostazione delle variabili e della funzione oggetto, scrivere nelle celle a lato (ad es., nella colonna B) il significato delle celle A (ripetendo ad esempio le denominazione delle variabili e una sigla identificativa della funzione obbiettivo.

3. LA PROGRAMMAZIONE SU RETI E LA TEORIA DEI GRAFI.

1. Introduzione.

In questo terzo capitolo introdurremo due strumenti fondamentali che saranno poi utilizzati nel seguito della esposizione per lo studio e la risoluzione di problemi specifici legati ai trasporti.

Si tratta innanzi tutto della **programmazione intera** ed in particolare della **programmazione su reti**, tipologia quest'ultima che fa riferimento ad una rappresentazione molto comune che è quella desunta dalla **teoria dei grafi**. Questa costituisce appunto l'altro strumento che sarà qui presentato.

I problemi che si vogliono affrontare, in effetti, dal punto di vista delle variabili di decisione richiedono in genere soluzioni a componenti intere (spesso addirittura vettori le cui componenti possono valere solo 0 oppure 1), mentre dal punto di vista di un'interpretazione geometrica fanno ricorso molto spesso al concetto di **grafo**, in quanto è proprio mediante un insieme di punti collegati tra loro da linee, (oggetti costitutivi di un grafo), che si rappresentano gli elementi di un sistema e le relazioni che li legano.

Alla programmazione intera in quanto tale saranno dedicati due soli paragrafi, che vogliono essere di rinvio ad una letteratura molto vasta. La programmazione su reti, viceversa, sarà brevemente delineata nel paragrafo successivo, ma emergerà sostanzialmente nei capitoli dal 5° in poi, attraverso la descrizione di varie problematiche dal forte contenuto applicativo.

Per quanto riguarda la teoria dei grafi, ad essa sarà dedicata la maggior parte di questo capitolo, con l'obiettivo di presentarne gli elementi di base, rinviando alle varie applicazioni le definizioni e le tipologie più direttamente correlate a casi concreti.

Nella teoria dei grafi si possono delineare due impostazioni relativamente divergenti: una può essere definita di tipo algebrico, l'altra viceversa è finalizzata alle applicazioni.

Ovviamente una distinzione netta non è possibile né utile e, anzi, per certi aspetti può essere pericolosa. Tuttavia non è un caso che di grafi si occupano matematici 'puri' (esistono insegnamenti di Teoria dei Grafi nei corsi di laurea scientifici), come pure studiosi di Ricerca Operativa, e questi ultimi vedono i grafi in maniera assai più strumentale.

Va anche detto però che proprio in Ricerca Operativa temi quali la Programmazione Matematica ed in particolare la Programmazione Intera sono studiati a volte in quanto tali, a

prescindere da applicazioni specifiche e proprio in tali circostanze acquistano importanza grafi di struttura molto particolare che con le loro proprietà sono strumenti per la risoluzione di problemi teorici, piuttosto che applicativi.

2. Programmazione intera e programmazione su reti.

3.2.1 La programmazione intera: generalità.

Nella **programmazione a numeri interi** si richiede che il valore delle componenti di un vettore soluzione siano quantità intere.

La condizione è ovvia in numerose situazioni pratiche.

Da un punto di vista teorico si può dimostrare che :

- qualunque problema in variabili intere, mediante opportune trasformazioni, può essere riformulato come problema in variabili **binarie** (o **booleane**), vale a dire variabili che possono assumere solo uno dei due valori, 0 e 1;
- a sua volta, qualunque problema in variabili 0-1 può essere scritto utilizzando polinomi nelle variabili interessate ed ogni monomio di questi polinomi contiene ciascuna variabile al grado 1.

Per giustificare la prima affermazione si può procedere come segue.
Sia x una variabile che può assumere i valori interi dell'intervallo [0, M], con $2^n \leq M < 2^{n+1}$. La variabile stessa può essere allora sostituita da un insieme di variabili $x_0, x_1, \ldots x_n$, ognuna delle quali vale 0 oppure 1, tali che:

$$x = x_0 + 2\,x_1 + 4\,x_2 + \ldots + 2^n\,x_n.$$

In sostanza, le x_i costituiscono, se sono lette in ordine inverso, le cifre della rappresentazione del numero intero x nella numerazione in base 2.

Per avere invece il secondo risultato, basta osservare che qualsiasi potenza intera di 0 oppure 1 conserva il valore (0 oppure 1, rispettivamente) e quindi non ha utilità pratica considerare potenze delle variabili che siano superiori alla prima.

Sulla programmazione intera (in particolare, sulla programmazione lineare a numeri interi) vi sono numerosi studi che hanno dato origine a svariate tecniche risolutive. Un elemento che è opportuno evidenziare immediatamente riguarda la difficoltà di questi problemi: in genere, un problema di programmazione lineare intera è più 'difficoltoso' da risolvere rispetto al corrispondente problema nel continuo. Ovviamente, al termine 'difficoltoso', a questo punto, non può che

essere dato un significato intuitivo: ma esiste una teoria, la teoria della complessità computazionale, che dà un significato preciso a questo concetto. Essa sarà sviluppata nel capitolo 4.

2. Esempi di problemi di programmazione intera.

Uno dei più noti esempi di problema di programmazione intera è il cosiddetto **problema dello zaino** (o **knapsack problem**)..

Per illustrarlo, si supponga di avere un contenitore di capacità K e siano dati n tipi di oggetti, A_1, A_2, …, A_n. Un generico esemplare del tipo di oggetti A_j comporta un'utilità (o un valore) u_j ed un peso (o volume, secondo com'è espressa la capacità del contenitore) p_j. Potremo supporre per semplicità, ma senza nulla togliere alla validità generale del modello, che sia la capacità del contenitore sia i pesi degli oggetti siano quantità intere.

Occorre caricare il contenitore rispettandone la capacità, con gli oggetti che danno la maggiore utilità complessiva.

E' evidente che la denominazione del problema fa riferimento ad un ipotetico escursionista che, prima di partire, deve caricare il suo zaino con gli oggetti che gli sono più utili per il viaggio, compatibilmente con la capienza dello zaino stesso o con il peso che egli riesce a sopportare (o con tutti e due!).

Vi sono parecchie varianti del problema dello zaino (a seconda, ad esempio, del numero di oggetti dello stesso tipo che è consentito inserire), tra le quali la formulazione 0-1 (quando di ogni tipo di oggetto si può inserire al più solo un esemplare), quella a numeri interi (con limitazioni o meno nel numero di oggetti dello stesso tipo), oppure la formulazione 'a scelta multipla' nella quale gli oggetti sono divisi in classi e all'interno di ogni classe è possibile scegliere (al più) solo un elemento.

Il caso della possibile scelta di più oggetti dello stesso tipo non richiede esemplificazioni. E' solo da osservare che non è detto che in tal caso vi sia proporzionalità tra numero di oggetti e utilità che se ne ricava (dopotutto, la prima lattina di una bibita dissetante è più utile per l'escursionista della seconda!).

Il caso del problema 'a scelta multipla' può essere illustrato con la formazione di un team o squadra di intervento compositi, in cui sono richieste alcune competenze specifiche, ma, una volta individuato tra tutti quelli disponibili uno specialista di un settore (ad es., il tecnico del suono per una ripresa televisiva), non ne servono altri. Il vincolo di capacità potrebbe essere dato dal budget complessivamente a disposizione per gli stipendi, mentre l'obiettivo è la massimizzazione del 'valore' complessivo della squadra.

Infine, proprio l'interpretazione del problema come la decisione di caricare uno zaino vero e proprio può dare origine ad un'altra versione nella quale vi sono due vincoli, uno di capacità ed uno di volume. Nel caso delle problematiche di trasporto, in effetti, un automezzo deve rispettare condizioni di entrambi i tipi, ed è da vedere caso per caso quale dei due vincoli è il più stringente.

Nella versione booleana, il problema si può formalizzare ricorrendo a n variabili di decisione x_j, ciascuna delle quali indica la politica che si adotta relativamente all'oggetto di tipo j:

$$\max \sum u_j x_j$$

s.t.

$$\sum p_j x_j \leq K,$$
$$x_j \in \{0, 1\}.$$

Perché il problema abbia senso occorre che vi siano oggetti con un peso (volume) inferiore alla capacità dello zaino e che, d'altra parte, la somma dei pesi di tutti gli oggetti sia superiore alla capacità stessa.

Per la risoluzione del problema dello zaino vi sono sia tecniche esatte sia tecniche approssimate. Tra queste ultime va segnalata la procedura di inserire nello zaino gli oggetti in ordine di **utilità specifica** decrescente, ricercando cioè innanzi tutto l'oggetto per il quale è massimo il rapporto u_j/p_j e inserendolo nello zaino (a patto che ciò sia possibile), quindi nello spazio residuo inserire (sempre se possibile) l'oggetto che ha utilità specifica immediatamente più piccola e ciò fino a esaurire la capacità. Si può dimostrare che in questo modo non si ottiene necessariamente la soluzione ottima, ma comunque si trova una soluzione che generalmente è una buona base di partenza per individuare, con tecniche successive, la soluzione ottima stessa.

Un altro problema, che può essere inteso anche come caso particolare del problema dello zaino, è il **problema del subset-sum**: dato un insieme I di n elementi di peso assegnato ed un valore K, si tratta di individuare un sottoinsieme degli elementi di I di peso complessivo $K' \leq K$ in modo tale che la differenza K-K' sia minima. Si tratta, in effetti, di un problema di zaino in cui tutti gli oggetti hanno la stessa utilità.

Sia sul problema dello zaino che sul subset sum problem esiste un'ampia letteratura: tali problemi sono interessanti anche perché una loro risoluzione efficiente è richiesta nell'ambito di modelli più complessi.

Molti problemi di programmazione intera sorgono nell'ambito dello **scheduling**, termine con il quale si designa l'organizzazione di attività nel tempo. Problemi di scheduling si hanno nel settore produttivo di un'azienda quando si organizzano le fasi tattiche e operative per ottenere i prodotti e far fronte agli ordini dei clienti. In un altro contesto, è di scheduling il problema di costruire i turni di lavoro del personale in un'azienda di trasporto pubblico o quello di stabilire le corse che devono essere effettuate da uno stesso veicolo.

Nello scheduling della produzione, la disponibilità di più macchinari in grado di eseguire ciascuno differenti operazioni e la necessità di sottoporre dei semilavorati a determinate lavorazioni, prima di ottenere il prodotto finito, pongono in genere il problema della scelta del miglior avvicendamento dei lavori sulle varie macchine, rispettando i vincoli di capacità produttiva e cercando di perseguire un criterio di efficienza (quale, ad esempio, il completamento di tutte le operazioni nel più breve tempo possibile).

Il problema si può impostare ricorrendo a variabili booleane x_{ij}, ognuna delle quali vale 1 se e solo se l'operazione i è effettuata presso la macchina j, vale 0 in caso contrario.

In aggiunta a quelli appena descritti, molti dei problemi portati come esempi introduttivi alla Programmazione Lineare possono essere formulati a loro volta come problemi di Programmazione Intera. E' il caso, evidentemente, della determinazione di un piano di produzione che massimizzi il guadagno, quando gli articoli da fabbricare devono essere un numero intero o quello in cui occorre

stabilire il piano di trasporto di costo minimo, se le quantità da trasportare hanno natura discreta, anziché continua. In tali circostanze basta aggiungere alle formulazioni già esposte gli ulteriori vincoli, rispettivamente:

$$x_j \text{ intere}, \ \forall\, j, \qquad \text{oppure} \qquad x_{ij} \text{ intere}, \ \forall\, i, j.$$

Occorre però osservare subito che, secondo il problema che si studia, si possono avere situazioni profondamente diverse. Mentre il problema (intero) del piano di produzione ottimale diventa nettamente più oneroso rispetto al caso continuo, nel problema di trasporto si può dimostrare che se le quantità disponibili nei vari luoghi di produzione, a_i, e quelle richieste nei vari luoghi di consumo, b_j, sono intere, le soluzioni di base del sistema dei vincoli hanno coordinate intere, per cui con il metodo del simplesso si ottengono soluzioni ottime che esse stesse hanno come componenti numeri interi. In quest'ultimo caso, problema intero e problema continuo coincidono.

Diventa allora interessante individuare quali sono i problemi che, anche se impostati nel continuo, hanno soluzioni intere: in effetti è quello che accade in molti problemi su reti e questo è, evidentemente, motivo di facilitazione nella soluzione di parecchi problemi reali. Tali casi saranno messi in evidenza nel seguito della esposizione.

Si osservi anche come i problemi dello zaino ed il subset-sum problem se affrontati nel continuo non danno soluzioni intere. D'altra parte, è il caso di ribadire che l'arrotondamento agli interi più vicini delle componenti di una soluzione che sia ottima nel continuo non sempre dà buoni risultati, nel senso che si possono portare esempi con i quali si dimostra che con tale procedura ci si può allontanare sensibilmente dalla soluzione intera ottima.

3. La programmazione su reti.

Una trattazione della programmazione su reti non può prescindere dalla rappresentazione dei problemi mediante i grafi e quindi da un'introduzione formale di questi ultimi. Tuttavia, ricorrendo ad esempi tratti dalla vita di tutti i giorni e attribuendo ai termini adoperati il loro significato corrente, non è difficile esemplificare situazioni che danno luogo a problemi di ottimizzazione su rete.

Le categorie principali di problematiche su rete si possono riassumere in:

- problemi di **percorso**;
- problemi di **flusso**;
- problemi di **localizzazione**.

Alla prima classe appartengono tipicamente quei casi in cui un decisore deve individuare il migliore itinerario in una rete stradale assegnata, nella quale ogni tratto ha determinate caratteristiche fisiche, quali lunghezza, compatibilità con i vari tipi di veicoli, tempi di percorrenza nelle varie situazioni di traffico. Al percorso si richiede, a seconda dei casi, di unire due località prestabilite, oppure di comprendere certi tratti, pure indicati a priori, di visitare un insieme di clienti sparsi

sul territorio, minimizzando il costo ed eventualmente rispettando requisiti temporali, quali intervalli di visita ai clienti ecc.

Alla seconda categoria appartengono quei problemi in cui si tratta di stabilire il quantitativo di un fluido che deve attraversare determinate condutture (quali ad es. le tubature di un acquedotto, variamente collegate tra di loro). Può essere che si debba cercare di inviare la maggior quantità possibile di fluido da un'origine ad una destinazione (e allora saranno vincolanti le capacità dei vari tratti di tubatura) oppure di scegliere l'instradamento migliore per la spedizione di un quantitativo di fluido prefissato. Ma vi sono anche svariati altri problemi che saranno proposti nel seguito della esposizione e, in ogni caso, il 'fluido' può rappresentare una merce o un insieme di veicoli o altro ancora.

Nella terza categoria vanno compresi quei problemi che, in una rete di comunicazioni assegnata, richiedono di individuare uno o più punti o località ove posizionare dei servizi (la cui vicinanza può essere gradita o no alla popolazione residente, ma che risultano comunque necessari). La casistica va dal posizionamento di punti vendita, magazzini intermedi, servizi pubblici come scuole e istituti di istruzione, alla localizzazione di servizi di emergenza, come ospedali, ambulanze, stazioni di polizia, vigili del fuoco ecc.

Ma tutti i problemi rappresentativi sinora esposti sono molto spesso utili in quanto la loro formulazione matematica coincide con quella che emerge da altre problematiche: ad esempio, può succedere che un problema di scheduling abbia la stessa formulazione di un problema di percorso ottimo ed allora è evidente che le tecniche usate per risolvere quest'ultimo si trasferiscono automaticamente all'altro problema. Pertanto le classificazioni sopra esposte sono puramente indicative. Sta poi all'utente individuare problemi e modelli simili al caso che si sta studiando e adattarli con il minor sforzo possibile al caso stesso.

3.3 Elementi di teoria dei grafi.

3.3.1 I grafi come metodologia di rappresentazione di sistemi.

La rappresentazione di un sistema mediante un grafo è assai comune ed è a volte utilizzata in maniera inconsapevole. Gli esempi stessi che sono portati per dimostrare come nella storia si sia fatto ricorso a tale rappresentazione lo mettono in evidenza. Dal classico caso di Eulero, che rappresentò con un grafo i quartieri di Koenigsberg e i ponti che li collegavano, alla rappresentazione delle molecole con le formule di struttura o alla descrizione dell'organigramma di un'azienda, si vede come il ricorso a 'caselle' o 'nodi' o 'punti' collegati tra di loro da 'linee' o 'frecce' riguardi ambiti disciplinari molto diversi e tra loro relativamente indipendenti.

Gli esempi sopra riportati possono essere anche utilizzati per introdurre problematiche che facendo ricorso a risultati della teoria dei grafi hanno avuto interessanti sviluppi. Ciò per dire che le rappresentazioni mediante grafi descritte sopra non sono fini a se stesse, ma sono servite o possono servire per risolvere problemi o mettere in evidenza proprietà nell'ambito della disciplina interessata.

In particolare, Eulero con la sua rappresentazione dimostrò che non era possibile individuare un percorso che partendo da uno dei quartieri della città vi ritornasse, transitando una ed una sola volta per ognuno dei sette ponti che collegavano i quartieri tra di loro. L'interesse attuale sta nel fatto che vari problemi di igiene urbana (ad es., la raccolta rifiuti, lo spazzamento delle strade ecc.) si possono risolvere sviluppando i ragionamenti di Eulero.

Nel caso delle formule chimiche, note proprietà dei grafi portano a escludere l'esistenza di sostanze con molecole formate da determinati numeri di atomi.

Infine, la codifica mediante un grafo dell'organigramma di un'azienda mette in evidenza aspetti relativi alla tipologia della organizzazione aziendale, alle modalità del flusso di informazioni e/o degli ordini all'interno della struttura ecc.

Un grafo viene di solito visualizzato mediante un insieme (finito) di **punti** (in pratica, cerchi o quadrati) collegati tra di loro in vari modi mediante **linee** o **frecce**. Questa concezione, tuttavia, fa perdere di vista il fatto che la rappresentazione mediante una figura (un disegno), in quanto tale, non è il grafo: lo stesso grafo può essere rappresentato in maniere diverse, anche se equivalenti.

Da un punto di vista matematico, un **grafo G** consiste in una **coppia di insiemi**, l'insieme N dei **nodi** (o **vertici**) e l'insieme A degli **archi**.

In questa esposizione, per semplicità, useremo i termini 'nodi' e 'vertici' come sinonimi; così pure intenderemo come sinonimi i termini 'archi', 'spigoli' e 'linee'. In letteratura, viceversa, a volte si utilizzano tali concetti con riferimento a contesti diversi (ad es. si parla di archi solo per grafi orientati ecc.), ma il vantaggio che se ne ricava è dubbio. In effetti la teoria dei grafi è relativamente recente e quindi manca ancora di definizioni universalmente accettate. Spesso un articolo o un contributo scientifico che utilizzi i grafi come strumento inizia specificando le definizioni adottate, e ciò è indispensabile elemento di chiarezza nella lettura.

Si scrive simbolicamente

$$G = (N, A).$$

Spesso si utilizzano anche i simboli V (per vertici) ed E (per gli archi). Inoltre il grafo stesso viene a volte denominato **rete**.

L'insieme N è non vuoto: esso è rappresentativo di entità tra le quali si intendono evidenziare dei legami a coppie. Gli elementi di N potranno essere indicati con le lettere a, b, c, ... u, v, w, z oppure con v_1, v_2, ..., v_n.

L'insieme A (che, viceversa, può essere anche vuoto) è formato da coppie di elementi di N: se si tratta di **coppie ordinate**, il grafo si dice **orientato** o digrafo (directed graph); se le **coppie non** sono **ordinate**, il grafo si dice **non orientato**.

Vi sono poi casi in cui alcune coppie sono ordinate mentre altre non lo sono (ad esempio, raffigurando una rete stradale, le strade a senso unico sono rappresentabili con archi orientati, le altre non richiedono orientamento): si ha allora un **grafo misto**.

Tornando alla osservazione precedente, in base alla quale un grafo è una coppia di insiemi, e non tanto la figura che può servire a rappresentarli, si dice che due raffigurazioni sono **isomorfe** se rappresentano lo stesso grafo, ma risultano graficamente differenti: evidentemente tra le due si può stabilire una corrispondenza biunivoca (ad ogni nodo della prima figura corrisponde un nodo della seconda e viceversa in modo tale che l'arco che collega due nodi esiste nella prima rappresentazione se e solo se figura anche nella seconda).

Gli elementi di A rappresentano, come si è detto, legami tra elementi di N.

Sulla natura degli elementi di N non si fanno particolari ipotesi: in pratica, in problematiche di trasporto è frequente associare ai nodi gli incroci di una rete stradale e quindi gli archi corrispondono ai tratti stradali che connettono gli incroci stessi, e in questo caso ogni arco rappresenta un'entità fisica. Ma, sempre in relazione a problemi connessi ai trasporti, può

succedere che i nodi rappresentino le corse che un'azienda di trasporto pubblico deve effettuare, mentre un arco orientato tra due corse esprime il fatto che uno stesso autobus, dopo aver effettuato la prima corsa può effettuare anche la seconda.

In altri casi, i nodi possono rappresentare persone (e gli archi rapporti di parentela) oppure compiti da eseguire (e gli archi, precedenze tra compiti) e così via.

Un arco che collega i nodi u e v si può rappresentare come coppia (u, v) o anche soltanto con la scrittura uv (a meno che non gli si dia una denominazione, in genere mediante una lettera minuscola: arco 'a'). Se il grafo è non orientato è evidente che lo stesso arco si potrà rappresentare anche con (v, u) oppure vu.

Dato l'arco a = (u, v), si dice che

- u e v sono i due **estremi** dell'arco;
- a **congiunge** u con v:
- u e v sono **adiacenti**;
- l'arco è **incidente** in u ed in v;
- u e v sono **incidenti** nell'arco a.

Se il grafo è orientato, e quindi necessariamente la coppia (u, v) è ordinata, u è il primo estremo o **origine** dell'arco, mentre v è il secondo estremo o **nodo finale**.

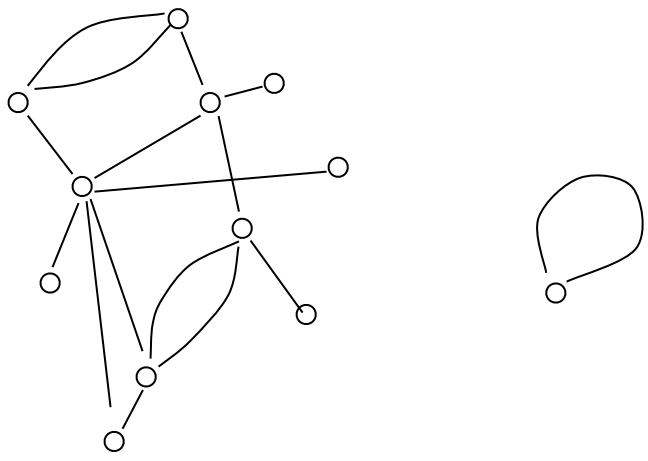
Due archi come h = (u, v) e g = (v, w), aventi un estremo in comune, sono **adiacenti**.

Oltre a ciò, se il grafo è orientato, h e g, considerati nell'ordine sono **consecutivi** e si dice anche che **h precede g** e che u è un **antenato** di w, mentre w si dice **discendente** di u..

In situazioni particolari occorrono rappresentazioni in cui vi sono più archi paralleli che collegano la stessa coppia di nodi: si parla allora di **multigrafo**.

Un multigrafo potrebbe essere rappresentato indicando per ogni arco (coppia di nodi) la sua molteplicità.

A volte occorre considerare anche archi per i quali i due estremi coincidono: si hanno i **loop** o **cappi**, archi del tipo (u, u).



un multigrafo un cappio

In un grafo orientato, due archi del tipo (u, v) e (v, u) si dicono **opposti**.
Se in un grafo orientato per ogni arco vi è anche l'opposto, il grafo si dice **simmetrico**, mentre se ciò non accade mai il grafo si dice **antisimmetrico**..

Un **grafo non orientato** nel quale esiste un arco per ogni coppia di nodi si dice **completo**.

Un **grafo orientato** è **completo** se per ogni coppia di nodi u e v esistono sia l'arco (u, v) che l'arco (v, u) (e quindi si tratta anche di un grafo simmetrico).

Non vi è denominazione particolare per indicare la presenza o meno in un grafo, altrimenti completo, di loop: ciò è detto esplicitamente caso per caso.

Si osservi infine che in un grafo completo orientato comprensivo di loop avente n nodi vi sono n^2 archi, mentre un grafo non orientato completo privo di cappi contiene n(n-1)/2 archi.

A volte, poi, serve dare un'indicazione del fatto che un grafo ha un numero relativamente piccolo di archi oppure, al contrario, che ne contiene molti: si parla rispettivamente di **grafo sparso** e **grafo denso**. I due concetti hanno carattere qualitativo (sono termini 'sfumati' o fuzzy'!): è eccessivo pensare ad un grafo sparso come ad uno che ha meno della metà degli archi di un grafo completo!

2. Sottografi, sottografi indotti e sottografi di supporto.

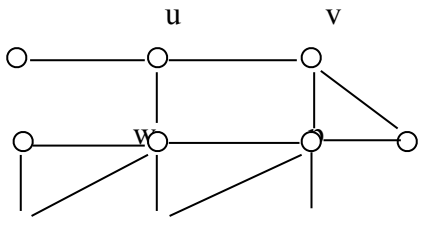
Dato un grafo G = (N, A), siano N' ⊆ N e A' ⊆ A sottoinsiemi di nodi e archi, rispettivamente, e supponiamo che A' contenga solo archi aventi estremi in N'. Diremo allora che G' = (N', A') è un **sottografo** di G.

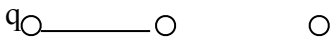
In pratica, per costruire un sottografo di un grafo assegnato G, è sufficiente cancellare da G alcuni nodi e alcuni archi, con l'accortezza di eliminare tra questi ultimi anche tutti gli archi che hanno uno o entrambi gli estremi tra i nodi cancellati. Come caso limite, è possibile cancellare solo alcuni archi. D'altra parte, si può ritenere ogni grafo come sottografo (improprio) di se stesso.

Si parla di **sottografo indotto** dai nodi di N' quando il sottografo G' contiene tutti gli archi di A che collegano nodi in N'; si parla invece di **sottografo di supporto** quando N' coincide con N (mentre, in generale, A' è un sottoinsieme proprio di A).

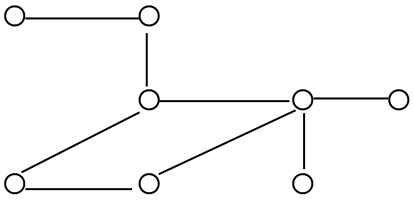
Infine un concetto che è utile in alcune applicazioni pratiche è quello di **grafo di sostegno non orientato di un grafo orientato** o di un **grafo misto**.
Con tale definizione si intende il grafo (non orientato) che si ottiene eliminando l'orientamento degli archi (in un grafo misto, di quelli che lo sono!), eventualmente eliminando le duplicazioni dovute ad archi opposti.

Quelli delle figure che seguono sono esempi, rispettivamente, di un grafo non orientato G, di un sottografo e di un sottografo indotto di G stesso.

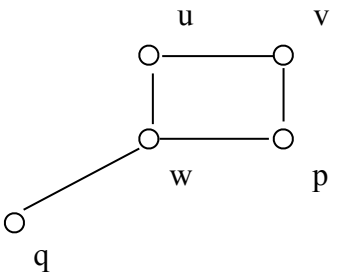




grafo G



un sottografo di G



il sottografo di G indotto da u, v, w, p, q.

3.3.3 Percorsi in un grafo; connessione.

In un **grafo non orientato** si dice **cammino** o **catena** una successione di archi, $e_1, e_2, \ldots, e_k$, che si possano scrivere nella forma

$$e_1 = (u_1, u_2),\ e_2 = (u_2, u_3),\ \ldots,\ e_k = (u_k, u_{k+1}).$$

Per comodità diremo che **sia gli archi e_j che i loro estremi appartengono al cammino**.

Poiché il grafo è stato supposto non orientato, si ha lo stesso cammino della definizione se in alcuni degli archi si scambiano tra loro i due estremi. Pertanto, ad esempio, si ha lo stesso cammino della definizione scrivendo gli archi come $e_1 = (u_1, u_2)$, $e_2 = (u_3, u_2)$, $e_3 = (u_4, u_3)$, ecc. Tuttavia non è corretto definire un cammino come 'successione di archi che hanno un estremo in comune ciascuno con quello che segue nell'insieme stesso', perché in tale definizione rientrerebbero altre strutture come la seguente:

$$e_1 = (u_1, u_2),\ e_2 = (u_1, u_3),\ e_3 = (u_1, u_4),\ \ldots,\ e_k = (u_1, u_{k+1}),$$

che è un grafo particolare (una stella) avente centro u_1 e k raggi.

Tornando alla definizione data sopra, partendo dal nodo u_1, è spontaneo pensare al cammino come ad un'entità che è possibile percorrere, attraversando via via tutti gli archi e_1, e_2, ecc., sino ad e_k, incontrando successivamente i nodi u_2, u_3, ecc., e giungendo in u_{k+1}. Da questo punto di vista viene spontaneo anche definire u_1 l'origine del cammino ed u_{k+1} come il termine dello stesso, ma, essendo il grafo non orientato, si possono capovolgere i concetti e definire u_{k+1} come origine e u_1 come nodo finale.

Un cammino è **semplice** se gli archi che contiene sono a due a due diversi, mentre è **elementare** se sono a due a due differenti i nodi $u_1, u_2, \ldots, u_k, u_{k+1}$.

Un cammino prende il nome di **circuito** o **ciclo** se i vertici u_1 ed u_{k+1} coincidono. Anche un circuito può essere semplice o elementare.

Un cammino (circuito) elementare è anche semplice, ma non necessariamente viceversa.

Si osservi poi che in un grafo non orientato i due termini 'cammino' e 'catena' sono sinonimi, così come sono considerati sinonimi 'circuito' e 'ciclo'. Si vedrà tra poco come invece nei grafi orientati la cosa non sia più valida.

In un grafo orientato si definisce **cammino** o **sentiero** una successione

$$e_1 = (u_1, u_2),\ e_2 = (u_2, u_3),\ \ldots,\ e_k = (u_k, u_{k+1}),$$

nella quale u_1 è nodo iniziale del primo arco, u_{k+1} è nodo finale dell'ultimo, mentre ogni altro nodo u_j, per $j = 2, 3, \ldots, k$, è estremo finale di un arco e iniziale per l'arco successivo.

Pertanto, in un cammino di un grafo orientato l'orientamento degli archi è concorde.

Il concetto di catena è più generale, non richiedendo che l'orientamento degli archi sia sempre concorde: una **catena** è una successione di archi che inducono un cammino nel corrispondente grafo non orientato di sostegno.

In pratica, una catena può essere vista come una successione di archi, analoga a quella che individua un cammino, ma nella quale certi archi siano del tipo $e_j = (u_{j+1}, u_j)$, cioè orientati in senso opposto a quello 'naturale' che va da u_1 a u_{k+1}. Si osservi, tuttavia, che il concetto di catena comprende come caso particolare quello di cammino, quindi se in una catena si può 'andare contromano' ciò non è strettamente obbligatorio.

Un cammino nel quale u_1 e u_{k+1} coincidono si dice **circuito**. Una successione di archi alla quale nel grafo non orientato di sostegno corrisponde un circuito si dice **ciclo**.

Analogamente a quanto evidenziato sopra, in un ciclo è possibile percorrere qualche arco 'contromano', fermo restando che i circuiti sono casi particolari di cicli.

I concetti precedenti consentono di classificare i grafi in conformità a una proprietà, detta di **connessione**, come segue:

- un grafo orientato per il quale, data una qualunque coppia ordinata di nodi, esiste sempre un cammino che li unisce, si dice **fortemente connesso**;
- un grafo orientato per il quale tra ogni coppia orientata di vertici esiste sempre almeno una catena, ma non un cammino, si dice **semplicemente connesso.**

In un grafo non orientato, non si distingue tra connessione semplice e forte:

- un grafo non orientato è **connesso** se esiste un cammino tra qualsiasi coppia di nodi.

Se un grafo non orientato G non è connesso, si possono ripartire i nodi in classi di equivalenza sulla base di una proprietà detta di **raggiungibilità** e individuando dei sottografi di G detti **componenti connesse** di G stesso.

Operativamente, dato un nodo u_1, si tratta di determinare tutti gli altri nodi, $u_2, u_3, \ldots, u_h$, collegabili mediante un cammino con u_1, cioè **raggiungibili** da u_1: il

sottografo indotto da tali nodi costituisce una prima componente connessa G_1 di G. Successivamente, considerato uno dei nodi non compresi in G_1, che diciamo u_{h+1}, si possono individuare i nodi u_{h+2}, …, u_k raggiungibili da u_{h+1} che formeranno una seconda componente connessa G_2. Se vi sono nodi non ancora compresi nelle componenti connesse sinora individuate, se ne potrà determinare una terza e così via.

Se il grafo G è orientato, si dicono sue **componenti connesse** i sottografi corrispondenti alle componenti connesse del sostegno non orientato di G.

Sulla rilevanza del concetto di connessione si tornerà in un paragrafo successivo, insieme con alcune considerazioni legate ai grafi misti.

3.3.4 Gradi.

In un grafo non orientato G, si dice **grado** di un vertice v il numero di archi incidenti in v; se G è orientato si distinguono il **grado interno** (numero di archi che terminano in v) e **grado esterno** (numero di archi uscenti da v).

Il concetto di grado è importante per alcune problematiche - ad esempio quella relativa alla organizzazione dello sgombero dalla neve di una rete di strade - perché in relazione al grado dei singoli nodi si può stabilire se è possibile attraversare con un circuito tutti gli archi del grafo ciascuno una ed una sola volta, senza ripassare quindi più volte per lo stesso arco, cosa che garantisce il costo minimo.

Un nodo di grado 1 si dice **pendente**; se ha grado 0 si dice **isolato**.

In qualunque grafo non orientato, **la somma dei gradi di tutti i nodi è un numero pari** ed è eguale al doppio del numero degli archi.

Infatti, ogni arco contribuisce al grado di due nodi, i suoi due estremi.

Di conseguenza **il numero dei nodi aventi grado dispari è un numero pari**.

Infatti, poiché la somma S di tutti i gradi è pari, detratti da S i gradi dei nodi a grado pari, anche la somma dei gradi dei soli nodi a grado dispari è un numero pari e questo richiede necessariamente che i nodi a grado dispari siano in numero pari.

In un grafo orientato, **la somma dei gradi interni è eguale alla somma dei gradi esterni** e tale somma coincide con il numero degli archi del grafo stesso.

3.3.5 Alberi e loro proprietà.

Un **albero** è un grafo (non orientato) **connesso e privo di cicli**; un grafo privo di cicli ma non connesso si dice invece una **foresta** (e sono alberi le sue componenti connesse).

Per gli alberi si può dimostrare la seguente proprietà caratteristica:

Proprietà P: un albero con n nodi ha n-1 archi.

Tale proprietà è condizione necessaria e sufficiente affinché un grafo connesso sia un albero oppure perché sia un albero un grafo privo di circuiti e può anche essere usata per dare la stessa definizione di albero, nel senso che un albero può anche essere definito come:

- un grafo connesso che ha un arco in meno del numero di nodi;
- un grafo privo di circuiti con un arco in meno del numero di nodi.

Partendo comunque dalla definizione data in apertura, ci si può rendere conto della validità della proprietà P considerando il fatto che un albero può essere disegnato progressivamente tracciando dapprima un nodo e poi aggiungendo alla struttura esistente, uno per volta, un arco ed un altro nodo terminale dell'arco stesso, sino ad esaurire tutti gli archi. I nodi utilizzati sono tanti quanti gli archi, oltre al nodo da cui il procedimento è partito.

Dimostrazioni più formali sono quelle che seguono.

a) (condizione necessaria: se un grafo è un albero, vale la proprietà P).

Si procede per induzione completa sul numero di nodi. Se l'albero ha 1 nodo, evidentemente non può avere archi, mentre se ha due nodi, essendo connesso e privo di circuiti non può che avere un solo arco. Pertanto la proprietà P è vera per n=1 e n=2. Supponiamo allora di avere un albero T con $k > 2$ nodi e supponiamo che la proprietà P sia vera per tutti gli alberi aventi un numero di nodi $\leq k-1$. Dimostriamo che la proprietà vale anche per T.

Infatti, eliminando un arco da T si rompe la connessione (se ciò non succedesse vi sarebbero circuiti, contro la definizione di albero) e si ottengono due sottografi a loro volta connessi e privi di circuiti T_1 e T_2, che sono quindi anch'essi due alberi: sia j il numero di nodi del primo e k-j quello del secondo. Sia j sia k-j saranno compresi tra 1 e k-1 e quindi, per l'ipotesi induttiva, T_1 ha j-1 archi mentre T_2 ne contiene k-j-1. Complessivamente l'albero T contiene un numero di archi dato da $(j-1) + (k-j-1) + 1 = k-1$ (perché avrà tutti gli archi dei due sottoalberi oltre all'arco che era stato eliminato).

b) (condizione sufficiente s_1: se vale la proprietà P in un grafo connesso G, il grafo G è un albero).

Occorre dimostrare che G è privo di circuiti. In effetti, se per assurdo vi fossero circuiti, se ne potrebbe considerare uno qualunque, C_1, e 'romperlo' eliminando un arco (u, v). Con questo il grafo rimarrebbe connesso, perché tra ogni coppia di nodi rimane almeno un cammino, che eventualmente utilizza la parte di C che non comprendeva (u,v) e conterebbe n nodi ma n-2 archi. D'altra parte, il grafo che resta dopo l'eliminazione di (u, v) non può essere un albero (perché allora, per la dimostrazione precedente dovrebbe avere n-1 archi), e quindi dovrebbe contenere almeno un altro circuito, C_2. Ciò comporterebbe che è possibile eliminare un altro arco, stavolta da C_2, ancora senza rompere la connessione e iterando il procedimento si giungerebbe, sempre mantenendo la connessione, ad eliminare tutti gli archi. Ma ciò è assurdo perché alla fine del procedimento resterebbero solo n nodi isolati. E' impossibile quindi supporre che nel grafo vi siano dei circuiti.

c) (condizione sufficiente s_2: se vale la proprietà P in un grafo G privo di circuiti, G è un albero).

Stavolta occorre dimostrare che G è connesso. Infatti, se non lo fosse, avrebbe più componenti connesse (almeno due) e tutte per ipotesi prive a loro volta di circuiti: per definizione si tratterebbe di alberi. D'altra parte a ciascuna di esse si potrebbe applicare la dimostrazione a), per cui varrebbe la proprietà P e se h è il numero di nodi della componente, gli archi sono h-1. Ma allora, se diciamo n il numero complessivo di nodi di G, il numero di archi sarebbe inferiore a n-1 (con k componenti connesse, gli archi di G sarebbero, precisamente, n-k) contro l'ipotesi. Pertanto è assurdo ipotizzare che il grafo non sia connesso.

Si noti infine come, oltre alla minore 'linearità' delle dimostrazioni b) e c) - entrambe condotte per assurdo - le stesse sono complicate dalla necessità di collegare la proprietà P ad una delle altre due caratteristiche degli alberi, la connessione e la aciclicità.

Anche se a rigore il concetto di albero si introduce per grafi non orientati, a volte si indica con lo stesso termine un grafo orientato (o anche misto) privo di cicli e (semplicemente) connesso.

Per i grafi orientati si definisce il concetto di **arborescenza**: è tale una struttura nella quale esiste un nodo **r,** detto **radice**, che è origine (comune) di un certo numero di cammini che raggiungono tutti gli altri nodi e non vi sono cicli.

In un albero non esiste, a rigore, una radice. Tuttavia, in varie applicazioni, può tornare utile attribuire il ruolo di radice ad un particolare nodo a partire dal quale poi individuare dei cammini che congiungano la radice stessa a tutti o parte dei nodi restanti.

Caratteristica comune di alberi ed arborescenze è la presenza di un unico cammino (o di una unica catena) tra ogni coppia fissata di nodi, e quindi il problema di individuare il percorso più breve tra due nodi in un albero è banale.

Si tenga presente che, in alcune applicazioni (ad esempio, nella tecnica di programmazione detta PERT, che ha lo scopo di individuare il tempo minimo necessario per completare una serie di operazioni, parte in serie e parte in parallelo), si presentano grafi orientati privi di cicli (grafi aciclici) che tuttavia non sono arborescenze (in effetti hanno per sostegno non orientato grafi contenenti cicli).

Sistemi di trasporto che hanno struttura ad albero (o ad arborescenza) sono piuttosto rari: una rete fluviale può essere di questo tipo, ma in genere la presenza di canali navigabili o diramazioni nel delta introducono circuiti.

Ha in genere struttura ad albero l'insieme dei tratti autostradali che è possibile percorrere a partire da un dato casello di entrata sino ad incontrare un altro casello o una barriera : qui acquista importanza proprio il requisito della unicità del percorso tra due nodi, in modo che si abbia con certezza un unico pedaggio da pagare.

Gli alberi sono uno strumento utile in algoritmi risolutivi di alcuni problemi (in particolare, per gli scopi di questa esposizione, nel Problema del Commesso Viaggiatore).

Un albero di k+1 nodi (con k>1) dei quali uno ha grado k mentre tutti gli altri sono pendente (di grado 1) si dice **stella**. In un grafo orientato, l'insieme degli archi uscenti da un vertice v si dice **stella uscente** (forward star) da v, mentre l'insieme degli archi entranti in v si dice **stella entrante** (backward star) in v.

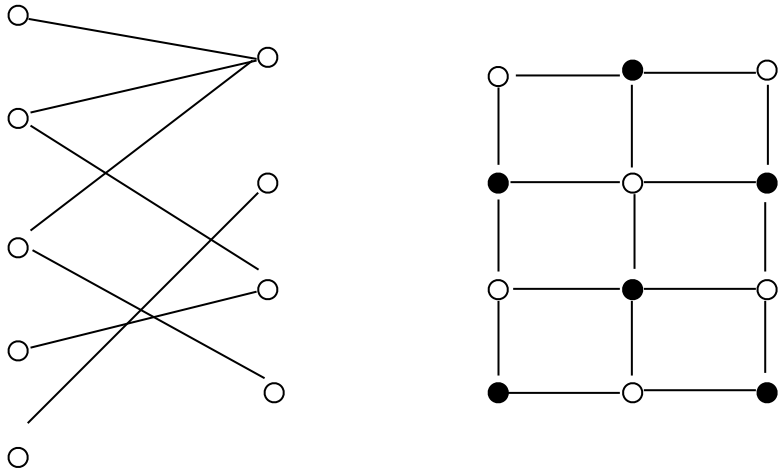
3.3.6 Grafi bipartiti.

Si dicono **bipartiti** i grafi in cui è possibile ripartire l'insieme N dei nodi in due sottoinsiemi (disgiunti), U e V, in modo tale che gli archi del grafo collegano solo nodi del primo sottoinsieme U con nodi del secondo sottoinsieme V. Un grafo bipartito, di conseguenza, si indica anche con una scrittura del tipo G = (U, V, A).

Il concetto è formulato per grafi non orientati.

Spesso si rappresenta un grafo bipartito evidenziando i nodi dei due insiemi su due colonne e gli archi collegano solo nodi della colonna di sinistra con nodi della colonna di destra. Ciò, comunque, non deve trarre in inganno perché può essere bipartito anche un grafo con una rappresentazione differente: ad es., sono

bipartiti entrambi i grafi della figura che segue (nella figura di destra, come sottoinsiemi U e V si possono assumere, rispettivamente, i nodi bianchi e neri).



Se ogni nodo di U è adiacente ad ogni nodo di V allora il grafo bipartito si dice **completo** (e se $|U| = m$, mentre $|V| = n$, allora $|A| = m \times n$).

E' immediato constatare che è bipartito ogni albero. Nel caso di grafi non connessi, essi sono bipartiti se e solo se lo sono tutte le loro componenti connesse.

In genere, in un grafo bipartito i due insiemi di nodi rappresentano due gruppi di elementi di natura diversa, tra i quali vi sono dei legami di natura logica. Ad esempio, il primo insieme può rappresentare dei compiti che devono essere svolti, mentre il secondo delle persone che devono svolgere tali compiti. Un arco che collega un nodo $u \in U$ con un altro nodo $v \in V$ esprime allora il fatto che il compito u può essere svolto dalla persona v.

Un altro caso tipico è costituito dal problema di trasporto, già visto nell'ambito della Programmazione Lineare. Si possono riconoscere immediatamente due insiemi di nodi, U, costituito dai luoghi di produzione di un bene, e V, formato dai luoghi dove questo bene è consumato per cui il grafo è bipartito ed in genere anche completo.

Il concetto di grafo bipartito sarà utilizzato nel problema di assegnazione, che a sua volta è un modello utile per la rappresentazione del problema (di scheduling) della attribuzione di un insieme di corse di un servizio di trasporto pubblico ai vari autobus che costituiscono il parco mezzi della azienda che gestisce il servizio stesso.

Si può dimostrare che:

i circuiti di un grafo bipartito hanno un numero pari di archi.

Si tratta di una condizione necessaria e sufficiente. Nelle dimostrazioni supporremo per semplicità che il grafo in questione sia connesso: l'estensione ai grafi non connessi è comunque immediata.

a) (Se un grafo G ha solo circuiti con un numero pari di archi è bipartito).

Si consideri un albero di supporto T di G e si scelga una radice r. Si ponga $r \in U$ e poi si attribuiscano di conseguenza tutti i restanti nodi dell'albero T ai due insiemi U e V. Si considerino gli archi di G che non figurano nell'albero. Ognuno di questi, se aggiunto a T, deve creare un circuito (se non fosse così, T non sarebbe di supporto). D'altra parte i circuiti sono solo di cardinalità pari e quindi qualunque arco aggiunto può collegare solo un nodo di U con uno in V (in caso contrario si avrebbe un circuito di cardinalità dispari). Pertanto, anche dopo la aggiunta degli archi che creano i circuiti il grafo rimane bipartito.

b) (Se un grafo è bipartito ha solo circuiti con un numero pari di archi).

Infatti, ogni circuito deve iniziare e terminare in uno stesso nodo, e supponiamo che come riferimento si tratti di un nodo di U. Poiché gli archi collegano solo nodi di U con nodi di V, con un numero dispari di archi si hanno solo cammini che iniziano in U e finiscono in V, ma non si possono avere mai circuiti.

3.3.7 Grafi misti e connessione.

Si è visto in precedenza come un grafo misto sia caratterizzato dalla presenza simultanea di archi orientati e archi non orientati ed è la rappresentazione più ovvia di una struttura quale una rete stradale in cui alcuni tratti sono non orientati (strade a doppio senso di circolazione) mentre altre vie sono dei sensi unici.

In alcuni problemi è possibile sostituire a un grafo misto G un grafo orientato G' sostituendo ad ogni arco non orientato di G una coppia di archi opposti in G'. Diremo **grafo orientato di sostegno del grafo misto** il grafo così ottenuto.

Va prestata attenzione al fatto che questa operazione non deve cambiare la natura del problema che si sta studiando.

Ad esempio, nel caso della raccolta dei rifiuti, una strada a doppio senso di circolazione in cui vi sono contenitori di rifiuti deve essere percorsa ma non necessariamente più di una volta (un unico passaggio può bastare per raccogliere i cassonetti posti su entrambi i lati): solo se vi è uno spartitraffico o se il mezzo effettua caricamento laterale sul lato destro, si rende necessario un passaggio in entrambi i sensi di marcia. Questo doppio passaggio diverrebbe comunque un obbligo se nella rappresentazione mediante grafo all'arco non orientato si sostituisse una coppia di archi orientati opposti.

Un discorso analogo vale per la distribuzione della posta: vi sono casi nei quali l'agente può ripetutamente attraversare la strada e percorrerla in definitiva solo in una delle due direzioni, mentre in altri casi occorre seguire il flusso consentito del traffico.

In problemi di organizzazione del traffico occorre a volte decidere quali tratti stradali rendere a senso unico per agevolare alcuni itinerari o, viceversa, per scoraggiarne altri. L'operazione in ogni caso deve mantenere la completa raggiungibilità reciproca tra tutte le coppie di nodi della rete stessa. Questo richiede che rimanga fortemente connesso il grafo di sostegno orientato corrispondente al grafo misto alla fine della operazione.

In effetti, gli uffici del traffico considerano anche un problema di raggiungibilità più ristretta, nel senso che spesso si cerca di evitare che, con la introduzione di sensi unici e svolte obbligate, determinati punti della rete (vicini fisicamente) siano troppo distanti tra di loro per un qualsiasi veicolo. In pratica ci si avvale di vari accorgimenti, quali ad es. l'inserimento di raccordi o bretelle per consentire inversioni di marcia. Da un punto di vista teorico, si possono studiare problemi di raggiungibilità entro un fissato numero di archi e la cosa può essere risolta utilizzando le matrici di adiacenza tra nodi (o matrici nodi-nodi) che saranno introdotte nel prossimo paragrafo.

Infine va segnalato il fatto che, per modellizzare situazioni di emergenza, anziché sostituire ad un grafo misto il corrispondente sostegno orientato, si può effettuare l'operazione inversa di cancellazione dell'orientamento degli archi che lo presentano, nella misura in cui dei mezzi di soccorso li possono attraversare anche contromano (ovviamente con tutte le cautele del caso!).

3.3.8 Grafi pesati: lunghezze e distanze.

Nei problemi di ottimizzazione su rete, il grafo rappresentativo di un sistema è generalmente **pesato** o **valutato**, nel senso che ai suoi elementi, nodi e/o archi, sono associati dei coefficienti o dei pesi: il significato di questi (utilità, penalità o importanza) dipende dal contesto in cui il problema stesso sorge.

Indicheremo con w_u il peso associato al (generico) nodo u e con c_{ij} o l_{ij} il peso associato all'arco (i, j).

Il peso associato ad un arco, in problemi di minimo, è detto **costo** dell'arco stesso oppure, anche senza particolari riferimenti a problemi di percorso, **lunghezza**. La lunghezza di un arco può anche essere un numero ≤ 0.

Il peso associato ad un nodo v può rappresentare l'importanza del nodo stesso: la popolazione residente in una località; la frequenza con cui il nodo esprime una certa domanda di beni o di servizi, ecc. Il peso associato ad un arco, può rappresentare la lunghezza di un tratto stradale che gli corrisponde, il numero di utenze distribuite lungo lo stesso tratto, il tempo necessario per percorrerlo o il guadagno che se ne può ricavare. Non è escluso, infine, che ad uno stesso arco siano associati più pesi.

In un grafo si introducono due differenti concetti di lunghezza di un cammino e di conseguenza di distanza tra due nodi. I due concetti fanno riferimento rispettivamente solo alla struttura di adiacenze del grafo oppure ad un sistema di pesi associati agli archi.

In un grafo (non pesato) si dice **lunghezza di un cammino** il numero di archi che lo compongono. Si dice poi **distanza tra due nodi** la lunghezza del cammino che li collega avente lunghezza minore.

Si vede come con queste definizioni si prende in considerazione solo la struttura algebrica del grafo: ciò è utile specialmente nei casi in cui gli archi indicano solo dei legami che non sono altrimenti quantificabili (non corrispondono a realtà passibili di una misurazione fisica).

In un grafo ad archi pesati si dice **lunghezza di un cammino** la somma delle lunghezze degli archi che lo compongono; **distanza tra due nodi u e v** è la lunghezza del cammino più breve (se esiste) tra i nodi stessi. Indicheremo la distanza dal nodo u al nodo v con una delle due scritture d(u, v) oppure d_{uv}.

Ovviamente il concetto di lunghezza è trasferibile anche ai cammini chiusi, cioè ai circuiti. Tuttavia è il caso di osservare che, con archi di lunghezza negativa, può risultare negativa la somma delle lunghezze degli archi che compongono il circuito stesso e ciò può avere forti ripercussioni pratiche, come sarà evidenziato tra poco.

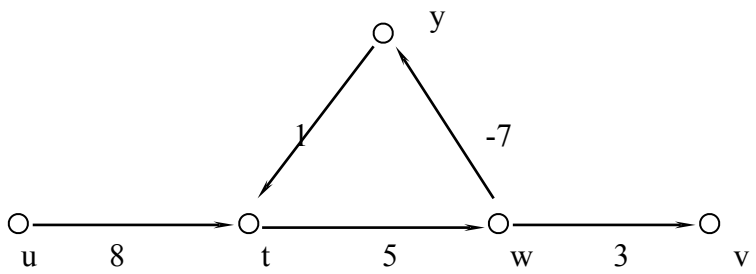
In presenza di archi con lunghezza negativa, **il cammino più breve tra due nodi u e v può non essere definito**. Infatti, se vi è un circuito di lunghezza negativa e questo, in un percorso da u verso v, può essere attraversato almeno una volta, ripetendo il circuito stesso un numero adeguato di volte, si ottiene un cammino di lunghezza negativa, arbitrariamente grande in valore assoluto

Ovviamente, in problemi di massimo, sono i circuiti di lunghezza positiva che rendono la soluzione ottima non definita.

Situazioni concrete che portano a rappresentazioni con queste proprietà (cammini di lunghezza infinita) sono rare e comunque, in genere, risultato di situazioni anomale e instabili. Un esempio è dato dall'arbitraggio spaziale nei cambi tra valute. Rappresentando con nodi un insieme

di valute e associando agli archi i logaritmi dei tassi di cambio tra due monete, non è difficile che in un determinato istante, per le inevitabili approssimazioni nelle cifre decimali dei tassi stessi, un cambio 'a circuito' di una moneta (ad es., con lire acquistare dollari, con questi marchi e poi riconvertire il ricavato in lire) dia luogo a un guadagno rispetto alla cifra iniziale. Allora immettendo nel circuito una somma di denaro che lo percorra un numero adeguato di volte, si può pervenire a un risultato grande a piacere. (Si noti che il ricorso a pesi logaritmici è l'accorgimento che consente di mantenere come definizione di lunghezza del circuito quella di somma delle lunghezze degli archi che lo compongono). In realtà, a parte la difficoltà di individuare il circuito e la presenza dei costi di transazione, non appena l'operazione diventa rilevante provoca essa stessa le inevitabili reazioni per cui il circuito di 'guadagno illimitato' è distrutto.

Quello che segue è un esempio in cui il cammino di lunghezza minima tra la coppia di nodi u e v non esiste.



Infatti, nel grafo in figura, il circuito formato dagli archi (t,w), (w,y), (y,t) ha lunghezza negativa e quindi non esiste il cammino più breve da u sino a v.

Nei grafi in cui sono stati introdotti pesi sia per i nodi che per gli archi si definisce anche il concetto di **distanza pesata**.

Supponiamo che le lunghezze degli archi siano numeri non negativi.
La distanza pesata da un nodo u ad un altro nodo v è definita come il prodotto della distanza da u a v moltiplicata per il peso w(v) del nodo v. Se indichiamo la distanza pesata con $d_w(u, v)$, si pone:

$$d_w (u, v) = d(u, v)\, w(v).$$

Il concetto di distanza pesata è utilizzato quando un percorso da un nodo ad un altro deve essere valutato un numero di volte proporzionale ad un coefficiente (peso) del nodo di arrivo. Ad esempio, se da una centrale telefonica si vuole collegare via cavo una località (ad es., una frazione di un comune) oltre che calcolare la distanza chilometrica occorre considerare il numero di utenti da raggiungere, perché ciascuno di questi richiede una propria coppia di cavi: nella determinazione della lunghezza complessiva dei cavi necessari per i collegamenti si deve allora moltiplicare la distanza fisica tra la frazione e la centrale per il numero di utenti della frazione stessa.

Analogamente, se si desidera collocare un centro da cui erogare un servizio che serve varie località, non basta individuare le distanze dai vari utenti, ma occorre anche considerare la frequenza con la quale il servizio è richiesto.

Lunghezze degli archi e distanze tra nodi di un grafo si possono riassumere in opportune matrici, come sarà illustrato alla fine del paragrafo che segue.

In particolare, le distanze si possono riassumere in tabelle triangolari, come tipicamente si descrivono le distanze tra varie località negli atlanti stradali. Occorre però prestare attenzione al fatto che operativamente è indispensabile indicare non soltanto il valore delle distanze stesse, ma anche i cammini (più brevi) che le implicano.

Da questo punto di vista generalmente le pubblicazioni geografiche sono carenti perché in genere i cammini minimi (che definiscono le distanze) non sono indicati esplicitamente e sono lasciati alla intuizione del lettore il quale li può rintracciare con esattezza solo … risolvendo nuovamente il problema di percorso minimo.

3.3.9 Grafi hamiltoniani e grafi euleriani.

Spesso le aziende di trasporto devono individuare dei percorsi di costo minimo con caratteristiche opportune.

Di frequente uno stesso veicolo deve raggiungere quanti più clienti può o passare per più strade con uno stesso viaggio che è generalmente un circuito, con arrivo e partenza presso un **deposito** o presso la residenza del conducente (e passaggio obbligato per il deposito).

Due casi limite sono dati dall'obbligo di transitare per tutti i nodi di una rete oppure per tutti gli archi.

Se il grafo rappresentativo è fortemente connesso, è evidente che esiste almeno un circuito (in generale non semplice e tanto meno elementare) che consente di visitare tutti i nodi come pure esistono circuiti che consentono di attraversare tutti gli archi. Il problema è che in questi circuiti alcuni nodi e, rispettivamente, alcuni archi dovranno essere visitati più volte.

Distingueremo innanzi tutto due categorie di problemi. Se si tratta di determinare uno o più circuiti che raggiungono tutti i **nodi** di un grafo si parla di problemi di tipo **hamiltoniano**.

Vengono in genere impostati come problemi hamiltoniani quelli legati alla distribuzione di merci a clienti.

Invece, se occorre garantire il servizio lungo tutti gli archi di un grafo (o per un sottoinsieme di archi ben individuato a priori), si parla di problemi di tipo **euleriano**.

Sono rappresentati come problemi euleriani, di solito, quelli che riguardano la pulizia delle strade (dalla neve, ad esempio), la raccolta dei rifiuti solidi urbani, la distribuzione della posta o di altri beni a utenti dislocati lungo le strade o le vie di una città.

E' evidente come sia la situazione specifica a suggerire la rappresentazione più adatta per la risoluzione di un problema particolare.

I problemi hamiltoniani saranno approfonditi in occasione della discussione dei modelli che risolvono casi di logistica distributiva, mentre quelli euleriani trovano il loro principale campo di applicazione nell'ambito dei problemi di organizzazione della raccolta di rifiuti.

Si danno le seguenti definizioni:

- si dice **hamiltoniano** un grafo nel quale è possibile individuare un circuito che passa per tutti i nodi esattamente una volta;
- si dice **euleriano** un grafo nel quale esiste un circuito che passa per tutti gli archi esattamente una volta.

Sia nei grafi hamiltoniani che in quelli euleriani, il circuito della definizione può non essere unico. Ma se il costo dei percorsi è misurato dalla lunghezza degli archi, nel caso dei grafi hamiltoniani si pone anche il problema di trovare il circuito hamiltoniano di costo minimo, mentre nei grafi euleriani ogni circuito euleriano ha lo stesso costo.

3.3.10 Rappresentazione di un grafo mediante matrice.

Grafi di dimensioni ridotte possono essere facilmente disegnati e sulla rappresentazione stessa si possono condurre semplici elaborazioni ma, sia a scopi di calcolo relativamente sofisticato che per la memorizzazione di grafi di dimensioni non banali occorre ricorrere a strumenti rappresentativi consistenti tipicamente in opportune matrici o in vettori che consentano l'uso del mezzo elettronico per la impostazione e risoluzione dei problemi.

Un grafo può essere codificato in vari modi tra i quali segnaliamo: la **matrice di adiacenza** o matrice **nodi-nodi**; la **matrice di incidenza** o matrice **nodi-archi** e le **liste di adiacenza**.

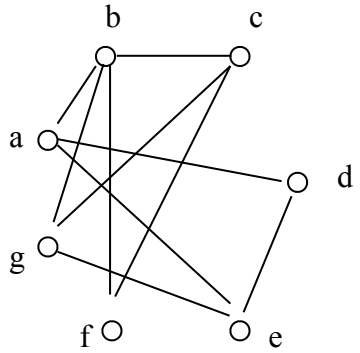
Un discorso a parte meritano alberi ed arborescenze, per i quali è possibile una rappresentazione mediante un unico vettore.

La scelta della rappresentazione è legata sia al problema che alle caratteristiche del grafo: come possono incidere queste ultime sarà messo in evidenza volta per volta.

La matrice di **adiacenza** per un grafo G = (N, A) di n nodi è una matrice quadrata nxn, avente una riga ed una colonna per ciascun nodo. L'elemento generico a_{ij} vale 1 se e solo se esiste l'arco dal nodo i al nodo j: in caso contrario vale 0.

La matrice di adiacenza può essere usata sia per grafi orientati che non orientati, (anche se in quest'ultimo caso risulta ridondante essendo necessariamente simmetrica) ed è uno strumento più comodo con **grafi densi**, mentre con grafi sparsi è preferibile ricorrere alla matrice di incidenza.

L'esempio che segue illustra la matrice di adiacenza di un particolare grafo.



grafo G

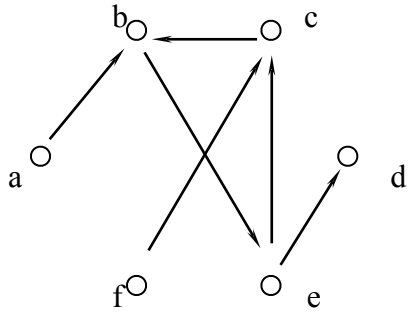
	a	b	c	d	e	f	g
a	0	1	0	1	1	0	0
b	1	0	1	0	0	1	1
c	0	1	0	0	0	1	1
d	1	0	0	0	1	0	0
e	1	0	0	1	0	0	1
f	0	1	1	0	0	0	0
g	0	1	1	0	1	0	0

matrice di adiacenza di G

La matrice di **incidenza**, invece, è formata da tante righe quanti sono i nodi e tante colonne quanti sono gli archi. In un grafo non orientato, per ogni arco sono indicati con 1 i due nodi estremi; gli altri elementi sono nulli. Invece se il grafo è orientato si indica con 1 il nodo origine e con -1 il nodo destinazione.

La attribuzione del valore negativo ai nodi destinazione è convenzionale, ma risulta più comoda per la impostazione di particolari problemi. Da un punto di vista mnemonico, si può pensare al nodo destinazione come ad un elemento che 'consuma' o 'divora' un determinato bene, da cui il segno negativo.

Nella figura che segue sono rappresentati un grafo orientato (costituito da 6 nodi ed altrettanti archi) e la sua matrice di incidenza.



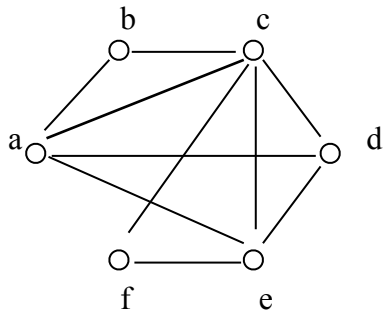
	u	v	w	p	q	r
a	1	0	0	0	0	0
b	-1	-1	0	1	0	0
c	0	1	-1	0	-1	0
d	0	0	0	0	0	-1
e	0	0	0	-1	1	1
f	0	0	1	0	0	0

La matrice di incidenza, come si è detto, è utile, evidentemente, quando sono poco numerosi gli archi, ma anch'essa utilizza una notevole quantità di memoria rispetto alle informazioni che fornisce.

La matrice nodi-archi ha particolari proprietà in relazione alla struttura del grafo. Ad esempio, se vi sono circuiti o catene le colonne corrispondenti agli archi che li compongono sono linearmente dipendenti.

Le **liste di adiacenza**, infine, indicano quali sono i nodi adiacenti ad un nodo fissato, e questo per tutti i nodi del grafo.

Per illustrarne l'uso, si consideri il caso seguente.



a b c d e
b a c
c a b d e f
d a c e
e a c d f
f c e

grafo G lista di adiacenza di G

Una lista di adiacenza presenta lo svantaggio che, in genere, il numero di nodi adiacenti ad un nodo fissato varia al variare di quest'ultimo e quindi, se si desidera trasformare la lista stessa in un vettore, occorre individuare in qualche modo (ad esempio, mediante un insieme di puntatori) dove, nel vettore complessivo, hanno inizio e termine i nodi adiacenti a ciascuno dei nodi del grafo.

Nel caso dell'esempio precedente, è chiaro che un vettore come

(b c d e a c a b d e f a c e a c d f c e)

presenta una certa ambiguità. L'inserimento di barre (|) può essere sufficiente a chiarire la situazione:

(b c d e | a c |a b d e f | a c e | a c d f | c e)

anche se nel caso specifico l'ordine (alfabetico) con cui sono stati scritti i nodi adiacenti a ciascuno dei 6 nodi può essere d'aiuto. Senza elementi separatori, tuttavia, non c'è comunque possibilità di esprimere l'esistenza di eventuali nodi isolati.

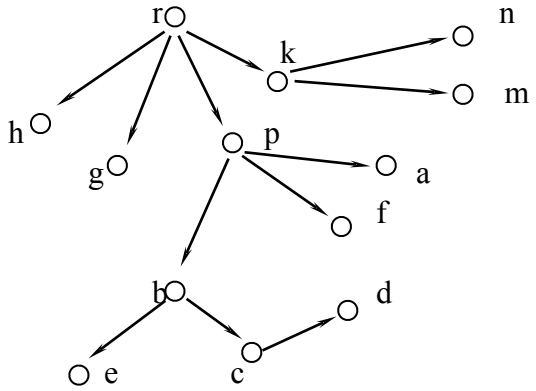
Per rappresentare un albero o una arborescenza si può ricorrere ad un unico vettore che sintetizza le informazioni necessarie, evitando l'uso di matrici.

Se il grafo è una **arborescenza**, il vettore dovrà contenere n-1 componenti, una per ciascun nodo eccettuata la radice. La generica componente corrispondente ad un nodo u contiene il predecessore di u, cioè l'origine dell'(unico) arco che ha u come nodo finale.

Banalmente, si tratta del principio in base al quale, in coda per una visita dal medico di famiglia, è sufficiente sapere che ognuno sappia chi gli sta immediatamente davanti perché sia rispettato l'ordine di priorità complessivo. In tal caso, tuttavia, la arborescenza è addirittura un cammino.

Si può ovviamente, se torna più agevole, utilizzare un vettore di n componenti associando alla radice un simbolo convenzionale, come ad es. ∅).

La figura che segue rappresenta una arborescenza e il relativo vettore.



r	***a***	***b***	***c***	***d***	***e***	***f***	***g***	***h***	***k***	***m***	***n***	***p***
∅	p	p	b	c	b	p	r	r	r	k	k	r

Come si è già detto, è possibile definire altre due matrici associate ad un grafo: la matrice delle lunghezze degli archi e la matrice delle distanze tra nodi.

La **matrice delle lunghezze** degli archi è sostanzialmente la matrice di adiacenza (o matrice nodi-nodi) nella quale al posto del valore 1, che indica la esistenza di un arco tra due nodi, è indicata la lunghezza dell'arco in questione. Se tra due nodi non esiste l'arco che li collega, non si potrà però più utilizzare il valore 0, che sarebbe interpretato come 'lunghezza nulla'.

Occorrerà piuttosto un valore, adeguato alle circostanze, che di fatto faccia sì che l'arco sia riconosciuto inesistente o comunque non utilizzato: in un problema di ottimizzazione con funzione oggetto da minimizzare, è opportuno usare valori numerici molto elevati, mentre occorrerà una quantità negativa grande in valore assoluto nel caso di problemi di massimo.

L'uso di valori elevati in valore assoluto e di segno appropriato è necessario ogni volta che si ricorre all'uso dell'elaboratore in un problema che è più facile impostare su un grafo completo, ma che si deve talvolta risolvere mancando qualche arco. Lo stesso vale se si deve forzare o , al contrario, impedire l'inserimento nella soluzione di un particolare arco. In letteratura questi valori 'elevati' vengono a volte indicati con una lettera, ad es., M.

Il valore da usare effettivamente va scelto con un certo equilibrio. Da un lato deve essere sufficientemente grande da garantire che l'arco in questione sia effettivamente escluso o scelto, a seconda di quello che si desidera.

D'altra parte, va evitato l'uso di grandezze esageratamente grandi (in valore assoluto) perché questo potrebbe condurre a errori nella soluzione finale dovuti alle approssimazioni effettuate dall'elaboratore alle prese, simultaneamente, con numeri molto grandi e numeri molto piccoli.

La **matrice delle distanze** riporta per ogni elemento d_{ij} la distanza tra i nodi i e j.

Mentre dalla matrice delle lunghezze è possibile risalire al grafo corrispondente, (proprio attraverso la interpretazione dei valori indicanti 'arco non esistente') la matrice delle distanze non lo consente.

L'ultima affermazione vale anche se alla matrice delle distanze si aggrega una matrice dei cammini minimi a partire da ciascun nodo, cioè una matrice avente per ogni riga il vettore che rappresenta la arborescenza con radice nel nodo stesso. Infatti, può succedere che qualche arco (u,v) non sia il cammino più breve tra i due estremi u e v e quindi esso non compare in nessuna delle arborescenze in questione. Ciò è legato ad una proprietà, sulla quale si tornerà in seguito, e cioè la proprietà triangolare: essa vale se e solo se per ogni terna di nodi i, j, k vale la relazione:

$$c_{ij} \leq c_{ik} + c_{kj}$$

cioè se il cammino più breve tra due nodi è comunque l'arco che li collega. La proprietà triangolare rende banali i problemi di percorso minimo non vincolato (in effetti, di solito, questi sono risolti in casi in cui non esistono tutti gli archi del grafo). La stessa proprietà, se non è verificata, può essere 'forzata' se si sostituisce all'arco (i,j) di costo c_{ij} un nuovo arco (fittizio) avente costo

$$\min_k \; c_{ik} + c_{kj}.$$

11. Grafi e programmazione lineare: i problemi di flusso.

Alcuni problemi di ottimizzazione su grafi si possono impostare come problemi di programmazione lineare (a variabili continue): in particolare rientrano in questa categoria i problemi di flusso. In questa sede saranno impostati due problemi che ritorneranno in successive sezioni a carattere applicativo: il problema del flusso di valore massimo e il problema del flusso (di valore assegnato) di costo minimo.

Premettiamo alcune definizioni:

- si dice **funzione flusso**, o semplicemente **flusso** su di una rete (grafo orientato) G una qualsiasi funzione che associa ad ogni arco del grafo un numero reale, che diremo **flusso sull'arco**; per gli scopi che interessano, ci limiteremo a considerare flussi non negativi;
- data una funzione flusso, ogni nodo u per il quale la somma dei flussi sugli archi entranti in u è inferiore alla somma dei flussi sugli archi uscenti prende il nome di **sorgente**; viceversa, se il flusso in ingresso supera quello in uscita si ha un **pozzo**.

Per ogni nodo, quindi, si può calcolare il bilancio tra flusso entrante ed uscente: se questo è in pareggio, allora si dice che per il nodo in questione vale il **principio di conservazione del flusso**.

La interpretazione più immediata dei concetti di flusso è ovviamente legata all'idea di un fluido che scorre per una rete di tubazioni: ogni nodo può essere fonte di nuovo flusso, un punto di consumo oppure semplicemente un punto di smistamento. In realtà si vedrà come si possono reimpostare come problemi di flusso questioni che nulla hanno a che vedere con questa interpretazione.

Si consideri ora un grafo orientato e supponiamo che ad ogni arco sia associato un numero positivo, **capacità dell'arco**, che esprime un confine superiore sul flusso sull'arco stesso. Nel grafo evidenziamo due nodi, che generalmente sono indicati con le lettere **s** e **t**, rispettivamente pozzo e sorgente. Ci si chiede allora quale possa essere la quantità più grande di flusso uscente da s che può pervenire in t rispettando i vincoli di capacità degli archi.

Da un punto di vista formale, se diciamo q_{ij} la capacità del generico arco (i, j), con x_{ij} il flusso che lo attraversa, si tratta di risolvere il problema:

$$\text{Max } z$$

s.t.

$$\Sigma\, x_{si} - \Sigma\, x_{is} = z$$
$$\Sigma\, x_{ij} - \Sigma\, x_{ji} = 0$$
$$\Sigma\, x_{it} - \Sigma\, x_{ti} = z$$
$$0 \le x_{ij} \le q_{ij}.$$

Nella formulazione, la variabile z, bilancio positivo del flusso uscente da s, prende il nome di **valore del flusso** ed il problema posto si dice del **flusso massimo** da **s** a **t**.

Il secondo problema preannunciato si può impostare supponendo siano dati, per ogni arco, due parametri, un parametro c(i, j) che rappresenta il costo di attraversamento di una unità di flusso sull'arco in questione, ed un secondo parametro q(i, j) che come sopra esprime la capacità dell'arco. E' dato anche il valore z del flusso (e si ipotizza che sia possibile inviare il flusso z da s a t in maniere alternative). Il problema del **flusso di costo minimo** si può scrivere allora come segue:

$$\text{Min } \Sigma\, c_{ij}\, x_{ij}$$

s.t.

$$\Sigma\, x_{si} - \Sigma\, x_{is} = z$$
$$\Sigma\, x_{ij} - \Sigma\, x_{ji} = 0$$
$$\Sigma\, x_{it} - \Sigma\, x_{ti} = z$$
$$0 \le x_{ij} \le q_{ij}.$$

I problemi visti godono della proprietà che se i coefficienti in gioco sono numeri interi, anche le soluzioni di base (e quindi la soluzione ottima) sono costituite da numeri interi. Tuttavia si tratta intrinsecamente di problemi nel continuo, a differenza di molti altri problemi su grafi, dove le variabili in gioco sono di tipo discreto, assai spesso 0 e 1.

I problemi di flusso sono problemi risolubili sia con la tecnica del simplesso, sia con tecniche apposite che utilizzano la particolare struttura dei vincoli.

Osserviamo infine che sulle reti di flusso si possono porre ancora altri problemi, per i quali si rinvia comunque alla letteratura.