Indice

Introduzione

La Ricerca Operativa è una disciplina relativamente giovane. Il termine *Ricerca Operativa* è stato coniato in ambito militare verso la fine degli anni '30 e deriva dal termine inglese *"Operational Research"*, ovvero la "ricerca sulle operazioni (militari)".

La Ricerca Operativa (di seguito indicata con l'acronimo *RO*) si occupa dello sviluppo e dell'applicazione di metodi quantitativi per la soluzione di problemi di decisione che si presentano nella gestione di imprese e organizzazioni.

Quando la complessità dei sistemi era relativamente contenuta, e la quantità di dati disponibili estremamente limitata, il personale esperto era sufficiente a prendere le decisioni necessarie alla conduzione dell'impresa.

La crescente complessità dei sistemi aziendali congiuntamente all'enorme quantità di dati messa a disposizione dall'informatizzazione capillare ha reso indispensabile l'utilizzo di strumenti automatici di decisione che attraverso la modellazione matematica permettano la soluzione di problemi di grandi dimensioni.

La RO, quindi, è caratterizzata dall'uso di modelli matematici definiti e risolti al fine di fornire indicazioni ai "decisori" nell'atto della scelta. Non a caso, la RO è anche nota come *management science*, e cioè la Scienza della Gestione, definizione che ne sintetizza finalità e ambizioni.

Breve storia della Ricerca Operativa

Il termine Ricerca Operativa, si è detto, è legato alle prime applicazioni della RO per aumentare l'efficienza di operazioni militari della Seconda Guerra Mondiale. Tuttavia esistono esempi importanti di anticipazione dei metodi della RO in anni più lontani; il più famoso risale a F. Taylor che nel 1885 elaborò uno studio sui metodi di produzione; prima ancora, nel 1776, G. Monge aveva studiato un problema di trasporti. Tuttavia la nascita della RO è legata agli studi che negli anni immediatamente precedenti alla Seconda Guerra Mondiale vennero condotti in Gran Bretagna per risolvere problemi strategici e tattici in operazioni militari. Più in particolare questi studi erano legati all'uso efficiente di un nuovo strumento di difesa: il radar. Infatti nel 1937 la Royal Air Force iniziò degli esperimenti di un sistema di controllo della difesa aerea basato sull'uso di una stazione radar situata a Bawdsey Research Station, nella costa est; già dai primi esperimenti si resero conto che era molto difficile gestire efficientemente le informazioni provenienti dal radar. Nel luglio 1938 furono compiuti altri esperimenti con l'aggiunta di quattro stazioni radar lungo la costa nella speranza che il sistema di controllo migliorasse sia in copertura sia in efficienza; invece non fu così ; dai nuovi esperimenti emersero seri problemi: c'era la necessità di coordinare e correlare le tante informazioni, spesso anche in conflitto tra di loro, che venivano ricevute dalle stazioni radar aggiunte. Nell'imminenza della Guerra si rese necessario tentare qualche nuovo approccio; perciò il sovrintendente della Bawdsey Research Station propose di sviluppare un programma di ricerca che riguardasse gli aspetti *operativi* del sistema e non più solamente quelli prettamente tecnici che erano da considerare soddisfacenti. Il termine *"Operational Research"* – Ricerca nelle operazioni (militari) – fu coniato per descrivere questa nuova branca delle scienze applicate. Fu quindi selezionato un gruppo di scienziati di vari discipline per costituire un "OR team"; il progetto fu diretto dal comandante in capo della Royal Air Force, Air Chief Marshal Sir Hugh Dowding. Nell'estate del 1939 la Gran Bretagna effettuò l'ultima esercitazione pre-bellica dove si evidenziò un notevole miglioramento nelle operazioni di difesa aerea grazie al contributo del gruppo di scienziati. Nacque quindi una vera e propria sezione che più tardi, nel 1941, prese il nome formale di "Operational Research Section". Durante il conflitto mondiale furono molteplici e importanti i contributi strategici di questa sezione permettendo di salvare piloti e aerei.

Al termine della guerra, alcuni degli scienziati coinvolti nel progetto formarono nuclei di ricercatori per lo sviluppo post bellico e la loro attività si estese a campi diversi da quello militare; in particolare, con l'espandersi delle iniziative industriali e con l'avvento dei computer che sono uno strumento essenziale per la risoluzione dei problemi, c'è stata un'espansione dell'utilizzo della RO all'interno di diverse realtà applicative.

Dagli anni '60 in poi le applicazioni della RO hanno avuto diffusione crescente, inizialmente nell'ambito di grandi gruppi industriali e succesivamente, grazie anche alla disponibilità di grandi potenze di calcolo a costi contenuti, in quasi ogni settore industriale, nei servizi e nelle amministrazioni pubbliche.

La Ricerca Operativa oggi

Ai nostri giorni la rilevanza applicativa delle tecniche della RO è riconosciuta e apprezzata in ambito industriale. Negli ultimi cinque anni il numero di addetti del settore è infatti cresciuto di un fattore 100. Contestualmente, si è allargata la richiesta di esperti di RO nelle imprese manifatturiere e di servizi: un laureato, esperto di tecniche della RO può ragionevolmente aspirare, per esempio, a ricoprire incarichi di responsabilità nelle industrie manifatturiere, nella assicurazioni, nel marketing, nelle società di consulenza aziendale, nella pianificazione e, sempre di più, nelle telecomunicazioni.

Alcuni esempi di problemi possono essere risolti per mezzo delle tecniche della RO sono i seguenti:

- *Finanza e Investimenti*;
  si vuole rispondere a domande del tipo: quanto dobbiamo investire, e come? Dove rimediare i capitali necessari? Quanto ci costerà? Alcuni esempi sono:
  - *Selezione degli investimenti*:
    si tratta di scegliere, fra un vasto insieme di alternative di investimento, quali attivare e quali no in presenza di vincoli di budget e con l'obiettivo di massimizzare i ricavi.
  - *Scelta del portafoglio*;
    consiste nel decidere in quali titoli e con quali quote investire i nostri capitali in modo da massimizzare il ricavo atteso, oppure minimizzare il rischio, etc.
  - *Determinazione del prezzo di derivati finanziari*;
    si vuole determinare il prezzo di un prodotto derivato finanziario (per esempio di un'*opzione* o di un *future*) in funzione del tempo e dell'andamento del titolo sottostaste.
- *pianificazione della produzione*;
  come assegnare la forza lavoro alle varie attività della nostra impresa? Su quali macchine e per quanto tempo ci conviene effettuare i nostri processi?

  Si tratta di pianificare i livelli di produzione e/o l'utilizzazione di risorse; si hanno spesso problemi di *allocazione ottima di risorse* cioè problemi riguardanti la distribuzione di risorse limitate tra alternative concorrenti in modo da minimizzare il costo complessivo o massimizzare il guadagno totale; tali risorse possono essere materie prime, manodopera, tempi di lavoro su macchine, capitali investiti.
- *gestione ottima delle scorte*;
  si tratta di determinare i livelli di produzione e di scorte nella gestione di materiali grezzi, prodotti in

lavorazione etc.; quando e quanto conviene riordinare materiali o beni in modo da ottenere il miglior compromesso fra costi di riordino e di produzione/acquisto e costi di immagazzinamento. Conviene, cioè, ordinare o produrre più spesso minori quantità per far fronte alla domanda corrente, oppure ordinare/produrre maggiori quantità e lasciarle in magazzino per soddisfare anche la domanda futura?

- *localizzazione e dimensionamento di impianti*;
  Quanti depositi di un'impresa di distribuzione alimentare costruire e dove localizzarli per servire i negozi a dettaglio in un'area d'interesse? Dove costruire degli ospedali (o scuole o stazioni dei vigili del fuoco) in modo da ottimizzare il servizio fornito? Dove conviene costruire le stazioni di base di una rete GSM/UMTS per coprire soddisfacentemente territorio e traffico, e con che potenza dovranno trasmettere? In senso lato, si tratta di problemi in cui si deve decidere dove istallare "impianti di produzione" in modo da "rifornire" in modo ottimale aree distribuite su un territorio.

- *progetto di reti di comunicazione / telecomunicazione*;
  si tratta di definire i collegamenti e dimensionare le capacità di una rete stradale, di telecomunicazione, di trasmissione dati, etc., in modo da garantire il traffico tra le varie origini e destinazioni e minimizzare il costo complessivo; ad esempio, per instradare le comunicazioni telefoniche e dati fra Roma e Venezia, conviene costruire una nuova linea ad alta velocità in fibra ottica fra Firenze e Bologna oppure istallare un ponte radio a larga banda?

- *determinazione di flussi ottimi*; si devono inviare merci (informazioni, telecomunicazioni, corrente elettrica, etc.) da alcune sorgenti (origini) a un certo numero di destinazioni utilizzando una rete di strade (fibre ottiche, doppini telefonici, cavi coassiali, emettitori) in modo da soddisfare le richieste minimizzando i costi di trasporto.

- *assegnazione di frequenze di trasmissione*;
  quali frequenze (prese da una banda limitata) devo assegnare a una rete di trasmettitori radiotelevisivi in modo da minimizzare le interferenze reciproche o massimizzare la copertura del territorio?

- *sequenziamento*;
  quali processo o operazione effettuare prima e quali dopo? Per esempio, come sequenziare i treni sulla rete in modo da evitare conflitti sulle tratte e minimizzare i tempi morti, le attese alle stazioni, etc.?

- *project planning*;
  come sequenziare le molteplici attività di un progetto? Quanto durerà il progetto? Come devono essere gestite le risorse?

- *allocazione ottima di componenti elettronici* (VLSI design);
  come disegnare una piastra madre in modo da minimizzare le lunghezze dei percorsi seguiti dai segnali elettrici?

- *determinazione dei turni del personale*;
  si tratta, ad esempio, di assegnare ai convogli il personale viaggiante sui treni (conducenti, bigliettai, etc.) in modo da minimizzare il numero di viaggi "a vuoto" (necessari per riportare il personale alla loro sede). Un problema analogo si presenta nell'assegnazione di equipaggi (piloti, hostess, stewart) a voli.

- *manutenzione di beni*;
  cioè il problema di decidere quando e se effettuare la manutenzione di alcuni beni soggetti ad usura, in modo da minimizzare il costo complessivo.

- *istradamento di veicoli*;
  quali percorsi devono seguire i veicoli di una flotta di automezzi per, ad esempio, raccogliere

l'immondizia, o rifornire una rete di negozi, in modo da minimizzare le distanze complessive percorse?

- *studi sulla struttura del DNA*;
  come assegnare sequenze a geni minimizzando la probabilità dell'errore sperimentale? Come determinare un albero filogenetico massimizzando la verosimiglianza?
- *progettazione di forme ottime*;
  che forma deve avere una macchina in modo da presentare meno resistenza possibile all'aria? Che profilo deve evere l'ala di un aereo in modo da massimizzare la portanza?
- *calcolo delle traiettorie ottime*;
  qual è la traiettoria che permette ad un veicolo spaziale di arrivare sulla luna e tornare usando la quantità minima di carburante?
- *ricostruzione di immagini*;
  come si possono visualizzare le informazioni fornite, per esempio, da una TAC in modo da renderle più leggibili possibili per il medico?
- *progettazione strutturale* ;
  qual è il progetto di un ponte o di un grattacielo che resiste meglio a venti molto forti o alle sollecitazioni derivanti da un terremoto?
- *yield management* ;
  Letteralmente traducibile come "Gestione del ritorno economico". In una azienda caratterizzata da varietà di servizi e di prezzi, domanda variabile nel tempo, stabilire quanti e quali servizi vendere avendo incertezza sulla domanda futura, allo scopo di massimizzare il profitto globale. Si tratta di un problema diffuso tra le compagnie di trasporto aereo, ferroviario, marittimo, ma anche per catene alberghiere e di noleggio auto.

Questa lista, lungi dall'essere esaustiva, serve a mettere in evidenza le potenzialità degli strumenti della RO nella risoluzione di problemi applicativi complessi e disparati.

In Italia la penetrazione della RO è stata piuttosto lenta. La situazione è rovesciata negli Stati Uniti e nell'Europa Centro-Settentrionale ove la crescita del settore è stata formidabile. Le ragioni del ritardo sono in primo luogo culturali: mancanza di conoscenze approfondite da parte delle aziende, insuffcente disseminazione dei risultati da parte dell'accademia. Lentamente, questa situazione va modificandosi anche in Italia, e la sensibilità delle aziende è fortemente cresciuta negli ultimi due-tre anni. In particolare ci si è resi conto che l'informatizzazione capillare e l'accresciuta potenza di calcolo non sono sufficienti a risolvere i problemi dell'organizzazione aziendale in modo ottimale.

A confermare questo asserto si consideri il seguente, illuminante esempio (dovuto a G. B. Dantzig): si supponga di essere a capo di un'azienda che impiega 70 dipendenti e deve assegnare ciascuno di essi a 70 differenti mansioni; poiché le capacità lavorative di ogni singolo dipendente sono diverse, non è indifferente per l'azienda come effettuare l'assegnamento. Naturalmente si deve fare in modo che ciascun dipendente sia assegnato ad una sola mansione e che ciascuna mansione sia svolta esattamente da un dipendente. Il problema consiste nel confrontare le 70! possibilità che ci sono per selezionare quella migliore nel senso che permetta di ottenere il maggiore utile per l'azienda. Le possibilità sono un numero molto grande, più grande di 10^{100}. Ora si supponga di disporre di un calcolatore capace di effettuare un milione di calcoli al secondo e che sia in funzione dal tempo del big bang; avrebbe questo calcolatore oggi nell'anno 2000 esaminato tutte le 70! combinazioni possibili ? La risposta è no. Supponiamo allora di disporre di un calcolatore che possa effettuare un bilione di assegnamenti per ogni nano secondo; la risposta sarebbe ancora no. Supponiamo allora di riempire la superficie terrestre di calcolatori di questo tipo che lavorano in parallelo; la risposta sarebbe ancora no. Si dovrebbe disporre in verità di 10^{40} terre ciascuna ricoperta di calcolatori di questo tipo, in funzione dal tempo del big bang fino a quando il sole si raffredderà.

Da questo esempio facile da enunciare si deduce come in certe situazioni sia assolutamente impossibile esaminare tutti i casi possibili per determinare qual è il migliore. Per questo, prima dell'avvento della

RO, l'unica possibilità era affidarsi al buon senso di persone guidate dall'esperienza che stabilivano regole "ad hoc" di base che dovevano essere seguite per risolvere i problemi (*"ad hoc" ground-rule approach*).

A questo tipo di approccio si contrappone la RO, il cui contributo centrale consiste nell'introduzione del cosiddetto *approccio modellistico-ottimizzatorio* per la soluzione di un problema di decisione. In questo approccio si organizza l'analisi di un problema reale in due fasi:

- la rappresentazione del problema attraverso un *modello matematico* che ne astragga gli aspetti essenziali e che schematizzi le interrelazioni esistenti tra i diversi aspetti del fenomeno che si sta studiando;
- lo sviluppo di *metodi matematici efficienti* (algoritmi di soluzione) per determinare una soluzione ottima del problema o una sua buona approssimazione.

Naturalmente, per costruire correttamente un modello matematico-ottimizzatorio che rappresenti un particolare fenomeno, si devono individuare i parametri di controllo significativi e un criterio per la valutazione della qualità della soluzione. La determinazione del modello è un'attività complessa e non completamente formalizzabile, che deve far ricorso da una parte a una conoscenza approfondita delle caratteristiche del problema in esame e dall'altra a strumenti che provengono da diverse branche della matematica. Una volta determinato il modello corretto, la RO si occupa di fornire una procedura esplicita per determinare una soluzione di un problema; tale procedura può essere rappresentata da metodi matematici analitici o, come più spesso accade, da metodi numerici che determinano la soluzione del problema mediante specifici algoritmi di calcolo. Da quanto detto si può capire come la RO sia una metodologia tipicamente interdisciplinare, applicabile nei più svariati contesti e come proprio dagli stimoli provenienti da campi anche molto distanti tra di loro tragga una delle principlai ragioni della sua attuale vitalità.

Capitolo 1

I Modelli della Ricerca Operativa

1.1 L'approccio modellistico

Il termine *modello* è di solito usato per indicare una costruzione artificiale realizzata per evidenziare proprietà specifiche di oggetti reali. Esistono modelli concreti (come ad esempio i prototipi di aerei o automobili), ma più spesso, come nella Ricerca Operativa, si considerano *modelli astratti* cioè *modelli matematici* che usano il simbolismo dell'algebra per mettere in evidenza le relazioni principali dell'oggetto che deve essere modellato. I modelli di cui si tratterà in seguito sono quindi modelli matematici, e sono costituiti da un insieme di relazioni che descrivono in modo semplificato, ma rigoroso, uno o più fenomeni del mondo reale. La nozione di modello matematico per rappresentare il mondo reale non è certo nuova: già Pitagora nel IV secolo a.C. tentava di costruire un modello matematico dell'Universo. L'interesse per la modellistica matematica è notevolmente cresciuto e attualmente si confida che attraverso modelli matematici sia possibile rappresentare molteplici aspetti del mondo reale e studiarne le proprietà. Ciò ha portato ad un enorme sviluppo delle applicazioni della modellistica matematica anche al di fuori delle tradizionali applicazioni alle scienze fisiche. Si è così avuta di fatto una vasta utilizzazione di modelli matematici in settori lontani dagli ambiti più tradizionali come, ad esempio, le scienze sociali, la biologia, le scienze ambientali, la psicologia. Come esempi concreti, si pensi agli studi sulla dinamica della popolazione, sulla diffusione delle epidemie, sul risanamento ambientale. Questa notevole diffusione della modellistica matematica è anche dovuta al fatto che l'evoluzione di un modello matematico può essere rapidamente studiata grazie all'uso di moderni calcolatori elettronici.

È evidente come in molti casi le situazioni rappresentate da un modello sono molto complesse e alcune volte influenzate da fenomeni di natura aleatoria; per questa ragione, sono state definite diverse classi di modelli matematici: *modelli stocastici* che considerano grandezze che possono essere influenzate da fenomeni aleatori e *modelli deterministici* che considerano grandezze esatte; inoltre a seconda che le interazioni tra le grandezze sono immediate o distribuite nel tempo, si parla di *modelli statici* e di *modelli dinamici.*

L'approccio modellistico per risolvere un problema di decisione o, più in generale, l'impiego di metodi matematici per la soluzione di problemi applicativi, viene di solito realizzato attraverso diverse fasi. Tali fasi possono essere schematizzate nel seguente modo:

- **Descrizione e Analisi del problema**
- **Costruzione del modello**
- **Analisi del modello**
- **Selezione di "buone" soluzioni (simulazione e/o ottimizzazione)**
- **Validazione del modello**

Descrizione e Analisi del problema
La prima fase consiste nell'*analisi della struttura del problema* e nell'*individuazione dei dati necessari* per una descrizione per una corretta definizione del problema. Si tratta cioè di individuare i parametri di controllo e di individuare i legami logico-funzionali che definiscono il problema e lo/gli obiettivi.

Costruzione di un modello matematico
Nella fase di costruzione del modello matematico si deve fornire una descrizione formalizzata del problema di decisione facendo uso del linguaggio della matematica. Si dovrà cercare, quindi, una corrispondenza tra relazioni del mondo reale (relazioni tecnologiche, leggi fisiche, vincoli di mercato, etc.) e relazioni matematiche (equazioni, disequazioni, dipendenze logiche, etc.).

$$\boxed{\textit{relazioni del mondo reale}} \quad \longleftrightarrow \quad \boxed{\textit{relazioni matematiche}}$$

La costruzione di un modello richiede valutazioni e scelte non facilmente codificabili in un procedimento standard. In particolare, per la costruzione di modelli soddisfacente è necessaria una conoscenza approfondita dell'applicazione d'interesse e dei metodi matematici di soluzione. La conoscenza dell'applicazione assicura che il modello sia soddisfacente e risponda alle domande concrete che l'utilizzatore gli porrà. La conoscenza dei metodi permette la definizione di modelli "risolvibili", cioé per i quali è possibile (al termine del processo di modellazione) la determinazione di soluzioni di buona "qualità".

È importante ribadire che un modello è definito per mezzo delle relazioni che lo costituiscono ed è quindi necessario che tali relazioni siano il più possibile indipendenti dai dati introdotti nel modello; questo perché uno stesso modello deve poter essere usato in differenti occasioni con dati (cioè costi, disponibilità di risorse, limiti tecnologici, etc.) diversi. Lo studio di questo aspetto, come già detto, rientra nella fase di analisi del modello sotto il nome di analisi della stabilità del modello rispetto ai dati introdotti.

In generale, la costruzione formale di un modello di Programmazione Matematica si può sintetizzare come segue:

1. Associare opportune *variabili di decisione* alle grandezze reali. Tali variabili costituiscono le incognite del problema.
2. Esprimere quantitativamente i *legami* esistenti tra le variabili e le *limitazioni* derivanti da considerazioni di carattere fisico, economico, etc. Tali legami e limitazioni definiscono i *vincoli*. L'insieme dei valori delle variabili per cui i vincoli sono soddisfatti costituisce l'*insieme ammissibile*.
3. Esprimere formalmente l'*obiettivo* che si intende minimizzare o massimizzare.

Analisi del modello matematico
Segue l'*analisi del modello* che prevede la deduzione per via analitica, in riferimento a determinate classi di problemi, di alcune importanti proprietà; le principali sono:

– *esistenza* della soluzione ottima;

– *condizioni di ottimalità*, cioè una caratterizzazione analitica della soluzione ottima;

– *stabilità* delle soluzioni al variare dei dati o di eventuali parametri presenti.

Lo studio delle condizioni di ottimalità ha sia motivazioni di natura teorica, sia motivazioni di natura algoritmica. Dal punto di vista teorico, una condizione di ottimalità può servire a caratterizzare analiticamente le soluzioni di un problema di ottimo e quindi consentire di svolgere analisi *qualitative*, anche in assenza di soluzioni numeriche esplicite; un esempio è l'analisi della sensibilità delle soluzioni di un problema di ottimo rispetto a variazioni parametriche.

Selezione di "buona" soluzione
La successiva fase di *selezione di "buona" soluzione* corrisponde alla possibilità di determinare tra tutte le scelte possibili costituite dalle soluzioni ammissibili, quella ottima o una sua buona approssimazione.

I problemi di ottimizzazione che si presentano nella pratica sono di solito così complessi che non è possibile determinarne una soluzione per via analitica. La complessità è determinata innanzi tutto dal numero di variabili e di vincoli, che definiscono la *dimensione* del problema; e poi dalla eventuale presenza di funzioni non lineari tra le funzioni che definiscono l'obiettivo e/o i vincoli.

Se il modello è molto semplice può essere possibile risolvere le relazioni e utilizzare i dati a disposizione per determinare una soluzione analitica. La soluzione analitica è possibile solo nel caso di poche variabili e di funzioni estremamente semplici, e cioè solo nei casi che si utilizzano come esempi ed esercizi nei testi e sulla lavagna. Molto spesso, anche se esiste una soluzione analitica è estremamente complessa e la sua determinazione richiede molte risorse di calcolo; ad esempio invertire una matrice è un banale esempio per il quale esiste una formula analitica, ma che dal punto di vista numerico può per certe istanze non essere affatto banale. Quindi, nella pratica, per determinare una "buona" soluzione di un problema di ottimizzazione occorre fare ricorso all'uso del calcolatore.
I due aspetti più importanti e in qualche modo complementari nella ricerca Operativa dell'uso del calcolatore per la soluzione di un modello matematico sono la *simulazione* e l'*ottimizzazione*.

Il **processo di simulazione** utilizza un modello matematico che consente di visualizzare l'effetto di alcune decisioni sul sistema in esame senza che queste debbano essere realizzate effettivamente sul processo reale. La simulazione consiste quindi nella valutazione numerica delle funzioni che definiscono il modello per alcuni valori delle variabili di interesse allo scopo di verificare come influenzino alcune misure di performance dell'uscita. Molto spesso, i modelli per i quali si utilizza la simulazione includono qualche aspetto di natura stocastica e anche dinamica (di evoluzione nel tempo).
La simulazione può essere usata come strumento per l'ottimizzazione nel senso che colui che prende le decisioni può *procedere per tentativi* e scegliere la "migliore" tra varie alternative possibile. Il modello matematico implementato nel simulatore consente di valutare l'effetto delle sue decisioni sul sistema nel suo complesso. Questo tipo di approccio è detto *analisi di scenari*. L'analisi di scenari è particolarmente utile nel caso di sistemi estremamente complessi e per i quali una rappresentazione analitica di tutti i legami logico-funzionali può non essere possibile. La soluzione determinata tramite l'analisi di scenari possibili non ha però alcuna garanzia di essere quella ottima o una sua approssimazione. L'uso della simulazione come strumento di ottimizzazione può essere poco significativo sebbene estremamente flessibile; viceversa può avere un ruolo molto significativo nella successiva fase di *validazione* del modello in quanto consente di individuare imprecisioni e/o errori nel modello stesso.
Non si deve confondere quindi il ruolo di ottimizzazione e simulazione. Dato un modello, il **processo di ottimizzazione** consiste nella determinazione della soluzione ottima, se esiste, o almeno di una sua buona approssimazione. Nella pratica, per risolvere un problema di ottimizzazione occorre fare ricorso ad un *algoritmo iterativo*, cioè ad un programma di calcolo che, a partire da una approssimazione iniziale x^0 della soluzione, determina, con una appropriata sequenza di operazioni che definiscono una successione di valori $\{x^k\}$, una nuova approssimazione x^*. La possibilità di realizzazione di algoritmi è fortemente legata alla capacità di definire condizioni di ottimalità che caratterizzano la soluzione ottima di un certo modello.

L'analisi del modello matematico e la definizione di un algoritmo per la sua soluzione sono aspetti fortemente legati tra di loro.
Molto spesso software di ottimizzazione e di simulazione sono integrati con software statistico o con fogli elettronici (spreadsheets). La combinazione di simulazione e/o ottimizzazione con la "visualizzazione" dei

risultati è una combinazione vincente. La visualizzazione rende l'uscita del processo di simulazione e/o di ottimizzazione molto più comprensibile e aggiunge spesso maggior credibilità al modello, soprattutto nei confronti di un pubblico non tecnico.
L'uso di fogli elettronici per la costruzione di modelli per l'uso di simulazione e di ottimizzazione sarà discusso nel Capitolo 9.

Validazione del modello
La soluzione numerica ottenuta al passo precedente deve poi essere valutata praticamente.
Questa fase di costruzione del modello non deve essere sottovalutata. I motivi di inattendibilità di un modello possono essere molti; in particolare la maggior difficoltà consiste nell'ottenere dati e/o informazioni validi. Spesso questo è legato al diverso linguaggio utilizzato dagli esperti del problema reale e dagli esperti di ottimizzazione. Informazioni essenziali sono spesso trascurate perché talmente scontate per l'esperto del problema da non dover essere raccontate. O viceversa modelli matematici troppo dettagliati possono produrre soluzioni incomprensibili.
La "validazione" del modello può avvenire attraverso una *verifica sperimentale* oppure con metodi di *simulazione*, allo scopo di ottenere, in questa fase di interazione con l'esperto, un modello matematico sempre più attendibile.

La definizione di un modello si configura quindi come un processo di raffinamento iterativo, che può essere schematizzato come rappresentato in Figura 1.1.

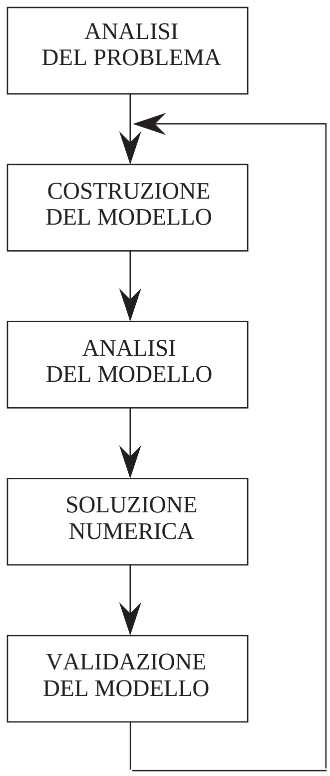

Figura 1.1: Fasi dell'approccio modellistico

Vantaggi dell'approccio modellistico
Esistono diverse ragioni per adottare l'approccio modellistico per la soluzione di problemi: si riassumono di seguito le principali.

– *Maggiore comprensione del problema.*

  Il modello è una rappresentazione semplificata del problema e spesso la sua costruzione consente di individuare proprietà strutturali del problema che altrimenti non sarebbero affatto evidenti.

– *Possibilità di risolvere matematicamente il problema.*

  Grazie al modello è possibile analizzare matematicamente il problema ed ottenere così una soluzione che, soprattutto in riferimento a scopi di pianificazione, permette di adottare strategie che da una sola analisi strutturale del problema non apparirebbero evidenti o che a volte potrebbero essere perfino controintuitive.

– *Deduzione analitica di importanti proprietà.*

  Nella fase di analisi del modello è possibile dedurre per via analitica alcune importanti proprietà del problema sulla base dei risultati disponibili per la classe di problemi a cui si fa riferimento.

– *Possibilità di simulazioni.*

  Con un modello è possibile effettuare esperimenti che spesso non è possibile effettuare direttamente nella realtà. La fase di simulazione è un passo fondamentale nella costruzione di un modello che può essere utilizzata per verificare l'effetto di una decisione, non necessariamente quella ottima, su un prototitpo del sistema e non sul sistema stesso; ad esempio, l'uso di un modello consente di studiare gli effetti dell'adozione di una particolare misura economica in un paese senza la necessità di sperimentarla direttamente. Il decision maker ha uno strumento che gli consente di valutare l'effetto di una sua decisione senza essere necessariamente in grado di capire l'aspetto modellistico del problema.

Critiche all'approccio modellistico
Le principali critiche all'approccio modellistico possono essere sintetizzate nei seguenti due punti:

– Impossibilità di quantificare soddisfacentemente con opportuni valori numerici alcuni dati richiesti dal modello; questo accade, ad esempio, nel tentativo di quantificare con un costo o con un profitto alcuni valori sociali soprattutto in relazione a scopi di pianificazione.

– La qualità delle risposte che un modello produce potrebbero dipendere profondamente dall'accuratezza dei dati introdotti.

La qualità delle risposte fornite dal modello dipende dall'accuratezza della sua definizione: la fase di validazione é cruciale per valutare la soluzione numerica ottenuta e completare il modello introducendo elementi trascurati in una prima fase.

1.2 Un primo esempio di costruzione di un modello matematico

Come primo esempio di costruzione di un modello matematico analizziamo un semplice problema di pianificazione degli investimenti.

Esempio 1.2.1 – CAPITAL BUDGETING. *Supponiamo di dover investire £1000 sul mercato finanziario. Supponiamo inoltre che il mercato offra tre tipi diversi di investimenti* $\mathbf{A}, \mathbf{B}, \mathbf{C}$ *ciascuno caratterizzato da un prezzo d'acquisto e da un rendimento netto, che sono riassunti nella seguente tabella:*

	A	**B**	**C**
costo	*750*	*200*	*800*
rendimento	*20*	*5*	*10*

Si vuole decidere quali degli investimenti effettuare per massimizzare il rendimento sapendo che gli investimenti **A**, **B**, **C** *non si possono effettuare in modo parziale cioà non sono frazionabili.*

Analisi del problema e costruzione del modello.

Si tratta di un problema di pianificazione degli investimenti. Si devono definire formalmente le variabili di decisione, l'insieme delle soluzioni ammissibili e la funzione obiettivo.
– *Variabili di decisione.* In questo caso il decisore vuole semplicemente sapere, per ogni investimento, se tale investimento deve essere effettuato oppure no. Una scelta naturale delle variabili di decisione è la seguente:

$$x_i = \begin{cases} 0 & \text{non si effetua l'investimento } i\text{−esimo} \\ 1 & \text{si effettua l'investimento } i\text{−esimo} \end{cases} \qquad i = \mathbf{A}, \mathbf{B}, \mathbf{C} \tag{1.1}$$

– *Insieme ammissibile.* In base alla definizione delle variabili, le possibili scelte compatibili con il nostro budget sono:

(0) non si effettuano investimenti $x_A = x_B = x_C = 0$

(1) si effettua l'investimento **A**; $x_A = 1, x_B = x_C = 0$

(2) si effettua l'investimento **B**; $x_A = 0, x_B = 1, x_C = 0$

(3) si effettua l'investimento **C**; $x_A = x_B = 0, x_C = 1$

(4) si effettuano gli investimenti **A** e **B**; $x_A = x_B = 1, x_C = 0$

(5) si effettuano gli investimenti **B** e **C**; $x_A = 0, x_B = x_C = 1.$

(6) si effettuano gli investimenti **A** e **C**; $x_A = 1, x_B = 0, x_C = 1.$

(7) si effettuano gli investimenti **A**, **B** e **C**; $x_A = x_B = x_C = 1.$

Tra queste solo alcune sono compatibili con il nostro budget. In particolare, notiamo che le possibilità **A**, **C** e **A**, **B**, **C** non sono ammissibili in quanto il costo supera la nostra disponibilità, come si evince dalla Tabella 1.2.

L'insieme ammissibile oè costituito dalle scelte $(0) - (5)$ ed è dato da:

$$S = \left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\}$$

Si tratta quindi di un sottoinsieme dei vettori di $\mathbb{R}^3$ a componenti $0 - 1$ ovvero

$$S \subseteq \{0, 1\}^3$$

– *Funzione obiettivo.* L'obiettivo che ci proponiamo è la massimizzazione del rendimento totale. Quindi dobbiamo esprimere la funzione obiettivo che corrisponde al rendimento netto relativo alla scelta di $x = (x_A, x_B, x_C)^T$ in S È possibile ottenere la soluzione ottima valutando esaustivamente la funzione obiettivo per ogni elemento di S, ottenendo in relazione alle possibili scelte i valori riportati nella Tabella 1.2.
La soluzione ottima è ovviamente quella corrispondente alla scelta (4), cioè all'effettuare gli investimenti **A** e **B**, con valore della funzione obiettivo pari a £25.

Questo rappresentazione del problema ha alcuni difetti, in particolare:

	Investimento							
	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
costi	0	750	200	800	950	1000	1550	1750
rendimento	0	20	5	10	25	15	30	35

Tabella 1.1: Tabella costi e rendimenti relativi a tutte le possibili combinazioni di investimento

1. *L'insieme ammissibile S è rappresentato in modo estensivo*, cioè elencando tutte le soluzioni ammissibili. In questo caso la cardinalità dell'insieme ammissibile è al più quella di $\{0,1\}^3$ cioè 2^3, ma in generale, se la dimensione del problema fosse più grande sarebbe impossibile valutare esaustivamente le soluzioni del problema. Se, ad esempio, il numero degli investimenti fosse stato 100 (che dal punto di vista delle applicazioni reali è del tutto verosimile) la cardinalità dell'insieme ammissibile sarebbe stata 2^{100} e per la valutazione di 2^{100} possibilità anche supponendo di utilizzare un calcolatore che effettui 10^{10} valutazioni al secondo (velocità superiore a quella raggiungibile dai calcolatori attuali) occorrerebbero 10^{20} secondi, cioè 3000 miliardi di anni !
2. *Il modello non è indipendente dai dati del problema*, cioè cambiando i dati del problema (prezzi e/o rendimenti) sarebbe necessario cambiare completamente il modello.

In genere si cerca di dare una *rappresentazione intensiva* dell'insieme ammissibile S, cioè individuare le proprietà $P(x)$ che consentono di distinguere le soluzioni ammissibili dagli elementi dell'insieme $\{0,1\}^3$ che non lo sono. Si vuole quindi scrivere l'insieme S in una forma del tipo:

$$S = \left\{x \in \{0,1\}^3 : \text{ vale la proprietà } \; P(x)\right\}.$$

Nell'esempio, la proprietà distintiva degli elementi di S è il costo complessivo che non deve essere superiore a £1000. Possiamo esprimere matematicamente questa relazione come:

$$P(x): \; 750x_A + 200x_B + 800x_C \leq 1000$$

e quindi l'insieme ammissibile si può scrivere

$$S = \left\{x = (x_A, x_B, x_C)^T \in \{0,1\}^3 : 750x_A + 200x_B + 800x_C \leq 1000\right\}.$$

Anche la funzione obiettivo può essere scritta in forma più sintetica come:

$$f(x) = 20x_A + 5x_B + 10x_C.$$

In conclusione, il problema di decisione può essere posto nella forma:

$$\begin{array}{ll} \max & (20x_A + 5x_B + 10x_C) \\ & 750x_A + 200x_B + 800x_C \leq 1000 \\ & x_i \in \{0,1\} \quad i = \mathbf{A}, \mathbf{B}, \mathbf{C}. \end{array}$$

□

Capitolo 2

Modelli di Ottimizzazione

2.1 Introduzione

In questo capitolo ci occuperemo più nel dettaglio di quei particolari modelli matematici noti come *Modelli di Ottimizzazione* che rivestono un ruolo centrale nella RO. In termini generali, data una funzione $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$, ed $S \subseteq \mathbb{R}^n$, un *problema di Ottimizzazione* può essere formulato nella forma:

$$\begin{array}{ll} \min & f(x) \\ & x \in S. \end{array} \qquad (PO)$$

Quindi un problema di Ottimizzazione consiste nel determinare, se esiste, un punto di minimo della funzione f tra i punti dell'insieme S. I problemi di ottimizzazione sono spesso denominati, con terminologia equivalente, problemi di Programmazione Matematica.
Osserviamo subito che un problema di massimo si può sempre ricondurre a un problema di minimo, cambiando di segno la funzione obiettivo. Infatti, i punti di massimo (ove esistano) del problema

$$\begin{array}{ll} max & f(x) \\ & x \in S \end{array}$$

coincidono con i punti di minimo del problema

$$\begin{array}{ll} min & -f(x) \\ & x \in S \end{array}$$

e risulta: $\max_{x \in S} f(x) = -\min_{x \in S}(-f(x))$.

In base a tale osservazione ci si può riferire esclusivamente, senza perdita di generalità, a problemi di minimizzazione.

La funzione f viene chiamata *funzione obiettivo* e l'insieme S *insieme ammissibile* cioè l'insieme delle possibili soluzioni del problema. Un punto $x \in S$ si chiama *soluzione ammissibile*.

L'insieme ammissibile S è un sottoinsieme di $\mathbb{R}^n$ e quindi $x = (x_1, x_2, \ldots, x_n)^T$ è una variabile vettoriale n-dimensionale e la funzione obiettivo f è una funzione di n variabili reali $f(x_1, x_2, \ldots, x_n)$.

2.2 Definizioni preliminari

Si riportano di seguito alcune definizioni fondamentali riguardanti i problemi di Ottimizzazione.

Definizione 2.2.1 (Problema inammissibile) *Il problema di ottimizzazione (PO) si dice* inammissibile *se $S = \emptyset$, cioè se non esistono soluzioni ammissibili.*

Definizione 2.2.2 (Problema illimitato) *Il problema di ottimizzazione (PO) si dice* illimitato (inferiormente) *se comunque scelto un valore $M > 0$ esiste un punto $x \in S$ tale che $f(x) < -M$.*

Un esempio di PO illimitato inferiormente è dato da $f(x) = x^3$ e $S = \{x : x \leq 2\}$. Infatti, al tendere di x a $-\infty$ la funzione obiettivo tende anch'essa a $-\infty$. Notiamo che se, con la stessa funzione obiettivo, si cambia l'insieme S, e si pone $S = \{x : x \geq 0\}$, il problema non è più illimitato inferiormente.

Definizione 2.2.3 (Punto di minimo globale) *Si dice che il problema di ottimizzazione (PO) am*mette soluzione ottima (finita) *se esiste un $x^\star \in S$ tale che risulti*

$$f(x^*) \leq f(x), \quad \text{per ogni} \quad x \in S.$$

Il punto $x^\star$ è detto soluzione ottima *o* minimo globale *e il corrispondente valore $f(x^\star)$ di dice* valore ottimo.

Per esempio, se si pone $f = x^2$ e $S = I\!R$, l'ottimo è l'origine, e il corrispondente valore ottimo è zero. Se si prende $S = \{x : x \geq 2\}$, l'ottimo è 2 e il valore ottimo 4.

In molti problemi di ottimo, in cui la ricerca di soluzioni globali può risultare difficile, può avere interesse anche la ricerca di soluzioni di tipo "locale". Indichiamo allora con $U(x)$ un *intorno di un punto* x e consideriamo la definizione seguente.

Definizione 2.2.4 (Punto di minimo locale) *Un punto $x^* \in S$ si dice* punto di minimo locale *di f su S se* esiste *un intorno $U(x^*)$ di x^* tale che:*

$$f(x^*) \leq f(x), \quad \text{per ogni} \quad x \in S \cap U(x^*),$$

e, in tal caso, si dice che $f(x^*)$ è un minimo locale *di f su S.*
Si dice che $x^* \in S$ è un punto di minimo locale stretto *di f su S se* esiste *un intorno $U(x^*)$ di x^* tale che:*

$$f(x^*) < f(x), \quad \text{per ogni} \quad x \in S \cap U(x^*), \; x \neq x^*.$$

Per esempio, se si pone $f = -x^2$ e $S = \{x : -2 \leq x \leq 1\}$, l'ottimo (minimo globale) è $x = -2$ di valore ottimo -4, ed il punto $x = 1$ è un minimo *locale* di valore pari a 1.

È immediato rendersi conto del fatto che

un punto di minimo globale è anche un punto di minimo locale, ma non è vero in generale il viceversa.

Naturalmente si può anche presentare il caso in cui la funzione obiettivo è limitata inferiormente su S ossia:

$$\inf_{x \in S} f(x) > -\infty,$$

ma tuttavia *non esistono punti di minimo globale* di f su S.
"Risolvere" un problema di ottimizzazione significa quindi, in pratica:

- stabilire se l'insieme ammissibile è non vuoto, oppure concludere che non esistono soluzioni ammissibili;
- stabilire se esistono soluzioni ottime, oppure dimostrare che il problema non ammette soluzioni ottime;
- determinare (eventualmente in modo approssimato) una soluzione ottima.

All'interno dei problemi di Ottimizzazione si possono distinguere le seguenti importanti classi di problemi:

- **Problemi di Ottimizzazione Continua.**
  Le variabili possono assumere tutti i valori reali ($x \in \mathbb{R}^n$); ed inoltre si parla di problemi di ottimizzazione continua
  - *vincolata* se $S \subset \mathbb{R}^n$
  - *non vincolata* se $S = \mathbb{R}^n$.
- **Problemi di Ottimizzazione Discreta.**
  Le variabili sono vincolate ad essere numeri interi ($x \in \mathbf{Z}^n$); si possono distinguere all'interno di questa classe di problemi altre due classi:
  - *programmazione a numeri interi* se $S \subseteq \mathbf{Z}^n$
  - *ottimizzazione booleana* se $S \subseteq \{0,1\}^n$.
- **Problemi misti.**
  Solo alcune delle variabili sono vincolate ad essere intere.

2.3 Problemi di Programmazione Matematica

Di solito l'insieme ammissibile S viene descritto da una numero finito di diseguaglianze del tipo $g(x) \leq 0$, dove g è una funzione definita su $\mathbb{R}^n$ a valori reali. Cioè, formalmente, date m funzioni $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, m$ si esprime S nella forma

$$S = \{x \in \mathbb{R}^n \mid g_1(x) \leq 0, \quad g_2(x) \leq 0, \quad \dots \quad , \quad g_m(x) \leq 0\}.$$

Ogni diseguaglianza $g_i(x) \leq 0$ prende nome di *vincolo* e l'insieme ammissibile è quindi formato da tutti quei punti $x \in \mathbb{R}^n$ che sono soluzione del sistema di diseguaglianze

$$
\begin{array}{rcl}
g_1(x) & \leq & 0 \\
g_2(x) & \leq & 0 \\
g_3(x) & \leq & 0 \\
 & \vdots & \\
g_m(x) & \leq & 0
\end{array}
$$

Osservazione 2.3.1 In questa formulazione dell'insieme S si sono utilizzati vincoli di diseguaglianza nella forma di minore o uguale, ma è chiaro che questa notazione include i casi in cui i vincoli sono espressi con vincoli di disuguaglianza nella forma di maggiore o uguale e vincoli di uguaglianza; infatti si può sempre trasformare un vincolo di maggiore o uguale del tipo $g(x) \geq 0$ in un vincolo di minore o uguale semplicemente riscrivendolo nella forma $-g(x) \leq 0$. Inoltre un vincolo di uguaglianza $g(x) = 0$ può essere riscritto nella forma equivalente delle due diseguaglianze $g(x) \leq 0$ e $-g(x) \leq 0$.

Quindi si può riscrivere il problema di ottimizzazione (PO) nella forma

$$
\begin{array}{ll}
\min & f(x) \\
 & g_i(x) \leq 0, \quad i = 1, \dots, m.
\end{array}
\tag{2.1}
$$

Un problema di questo tipo viene chiamato *problema di Programmazione Matematica.*

I problemi di Programmazione Matematica si possono classificare in base alle proprietà della funzione obiettivo e dei vincoli prendendo in considerazione, tra le più significative la *linearità*, la *convessità*.
Una prima distinzione è quella che fa riferimeno all'ipotesi di linearità. Da tale punto di vista, possiamo distinguere:

- problemi di Programmazione Lineare (PL), in cui l'obiettivo è una funzione lineare del tipo

  $$c_1x_1 + c_2x_2 + \cdots + c_nx_n,$$

  e i vincoli sono espressi da un sistema di equazioni e disequazioni lineari, cioè esprimibili nella forma

  $$a_{i1}x_1 + \quad \dots \quad + a_{in}x_n \geq (\leq / =) b_i$$

- problemi di Programmazione Non Lineare (PNL), in cui l'obiettivo oppure i vincoli non sono tutti lineari.

I problemi di PNL corrispondono alla situazione più generale. I problemi non vincolati rientrano nella classe di PNL, infatti, in tal caso f è necessariamente una funzione *non lineare* da R^n in R; è facile verificare che se f fosse lineare il problema non ammetterebbe soluzione.
Inoltre, in linea di principio, possono essere formulati come problemi di PNL anche i problemi combinatori.

La classificazione dei problemi di Programmazione Matematica in base alla proprietà di convessità sarà affrontata nel Capitolo 3.

Alcuni esempi di problemi di Programmazione Matematica sono i seguenti:

Esempio 2.3.2 *Si consideri una funzione obiettivo di due variabili* $f(x_1, x_2) = 2x_1 + x_2$ *che si vuole minimizzare, con i vincoli* $2x_1 + x_2 \leq 1, x_1 \geq 0, x_2 \geq 0$. *Si ottiene il problema*

$$
\begin{array}{ll}
\min & 2x_1 + x_2 \\
 & x_1 + x_2 \leq 1 \\
 & x_1 \geq 0 \\
 & x_2 \geq 0
\end{array}
$$

che è nella forma (2.1) dove $g_1(x_1, x_2) = x_1 + x_2 - 1$, $g_2(x_1, x_2) = -x_1$, $g_3(x_1, x_2) = -x_2$. *L'insieme ammissibile è descritto attraverso da tre vincoli. Poiché tutte le funzioni che compaiono sono lineari nella variabili* x_1 *e* x_2, *questo problema è un problema di Programmazione Lineare.*

Esempio 2.3.3 *Si consideri una funzione obiettivo* $f(x_1, x_2) = (x_1 - \frac{1}{2})^2 + (x_2 - \frac{1}{2})^2$ *che si vuole massimizzare, con i vincoli* $x_1 + x_2 \geq 1$, $x_1 \leq 1$, $x_2 \leq 1$. *Si ottiene il problema*

$$\begin{array}{ll} \min & -(x_1 - \frac{1}{2})^2 - (x_2 - \frac{1}{2})^2 \\ & x_1 + x_2 \geq 1 \\ & x_1 \leq 1 \\ & x_2 \leq 1 \end{array}$$

che è nella forma (2.1) dove $g_1(x_1, x_2) = 1 - x_1 - x_2$, $g_2(x_1, x_2) = x_1 - 1$, $g_3(x_1, x_2) = x_2 - 1$. *L'insieme ammissibile è un poliedro; la funzione obiettivo è quadratica. Si tratta quindi di un problema di Programmazione Non Lineare.*

Esempio 2.3.4 *Si consideri una funzione obiettivo* $f(x_1, x_2) = 3x_1^3 + 7x_1^2 + x_2$ *che si vuole minimizzare, con vincoli* $x_1 + x_2 \leq \frac{1}{2}$, $x_1 \geq 0$, $x_2 \geq 1$. *Si ottiene il problema*

$$\begin{array}{ll} \min & 3x_1^3 + 7x_1^2 + x_2 \\ & x_1 + x_2 \leq \frac{1}{2} \\ & x_1 \geq 0 \\ & x_2 \geq 1 \end{array}$$

che è un problema di Programmazione Non Lineare che può essere facilmente ricondotto nella forma (2.1) riscrivendo gli ultimi due vincoli nella forma $-x_1 \leq 0$ *e* $-x_2 \leq -1$.

Esempio 2.3.5 *Si consideri una funzione obiettivo* $f(x_1, x_2) = x_1 + x_2$ *che si vuole minimizzare sulla regione ammissibile descritta dal vincolo di uguaglianza* $4x_1 - x_2 = -2$. *Il problema risultante è:*

$$\begin{array}{ll} \min & x_1 + x_2 \\ & 4x_1 - x_2 = -2 \end{array}$$

che è un problema di Programmazione Lineare con un solo vincolo di uguaglianza.

2.4 Esempi di modelli di Programmazione Matematica

Come primi esempi di costruzione di modelli verranno ora analizzati un semplice problema di pianificazione della produzione, un problema di pianificazione degli investimenti e un problema di progettazione industriale.

Esempio 2.4.1 *Un'industria chimica fabbrica 4 tipi di fertilizzanti,* **Tipo 1**, **Tipo 2**, **Tipo 3**, **Tipo 4**, *la cui lavorazione è affidata a due reparti dell'industria: il reparto produzione e il reparto confezione. Per ottenere fertilizzante pronto per la vendita è necessaria naturalmente la lavorazione in entrambi i reparti. La tabella che segue riporta, per ciascun tipo di fertilizzante i tempi (in ore) necessari di lavorazione in ciascuno dei reparti per avere una tonnellata di fertilizzante pronto per la vendita.*

	Tipo 1	**Tipo 2**	**Tipo 3**	**Tipo 4**
Reparto produzione	2	1.5	0.5	2.5
Reparto confezionamento	0.5	0.25	0.25	1

Dopo aver dedotto il costo del materiale grezzo, ciascuna tonnellata di fertilizzante dà i seguenti profitti (prezzi espressi in Euro per tonnellata)

	Tipo 1	**Tipo 2**	**Tipo 3**	**Tipo 4**
profitti netti	250	230	110	350

Determinare le quantità che si devono produrre settimanalmente di ciascun tipo di fertilizzante in modo da massimizzare il profitto complessivo, sapendo che settimanalmente il reparto produzione può lavorare al più 100 ore mentre il reparto confezionamento può lavorare al più 50 ore settimanali.

Analisi del problema e costruzione del modello.

Si tratta di un problema di pianificazione della produzione industriale. Costruiamo un modello di Programmazione Matematica rappresentante il problema in analisi supponendo di voler pianificare la produzione settimanale.

– *Variabili di decisione.* La scelta delle variabili di decisione é molto delicata: infatti la qualitá dell'intero modello dipenderá da essa. In genere, per stabilire l'opportuno insieme di variabili di decisione, conviene porsi la seguente domanda: che cosa vuole sapere il decisore alla fine del processo di ottimizzazione? Ancora meglio, cosa gli é *sufficiente* sapere, per prendere le sue decisioni? In questo caso, ad esempio, tutto ció che il decisore deve conoscere sono le quantitá di fertilizzante da produrre per ciascun tipo. Dunque introduciamo le variabili reali x_1, x_2, x_3, x_4 rappresentanti rispettivamente le quantità di prodotto del **Tipo 1**, **Tipo 2**, **Tipo 3**, **Tipo 4** da fabbricare in una settimana.

– *Funzione Obiettivo.* Ciascuna tonnellata di fertilizzante contribuisce al profitto totale secondo la tabella data. Quindi il profitto totale sarà

$$250x_1 + 230x_2 + 110x_3 + 350x_4. \tag{2.2}$$

L'obiettivo dell'industria sarà quello di scegliere le variabili x_1, x_2, x_3, x_4 in modo che l'espressione (2.2) del profitto sia massimizzata. La (2.2) rappresenta la funzione obiettivo.

– *Vincoli.* Ovviamente la capacità produttiva della fabbrica limita i valori che possono assumere le variabili x_j, $j = 1, \dots, 4$; infatti si ha una capacità massima lavorativa in ore settimanali di ciascun reparto. In particolare per il reparto produzione si hanno a disposizione al più 100 ore settimanali e poiché ogni tonnellata di fertilizzante di **Tipo 1** utilizza il reparto produzione per 2 ore, ogni tonnellata di fertilizzante di **Tipo 2** utilizza il reparto produzione per 1.5 ore e così via per gli altri tipi di fertilizzanti si dovrà avere

$$2x_1 + 1.5x_2 + 0.5x_3 + 2.5x_4 \le 100. \tag{2.3}$$

Ragionando in modo analogo per il reparto confezionamento si ottiene

$$0.5x_1 + 0.25x_2 + 0.25x_3 + x_4 \le 50. \tag{2.4}$$

Le espressioni (2.3), (2.4) costituiscono i vincoli del modello. Si devono inoltre esplicitare vincoli dovuti al fatto che le variabili x_j, $j = 1, \dots 4$ rappresentando quantità di prodotto non possono essere negative e quindi vanno aggiunti i vincoli di non negatività

$$x_1 \ge 0, x_2 \ge 0, x_3 \ge 0, x_4 \ge 0.$$

La formulazione finale sarà quindi

$$\begin{array}{ll} \max & 25x_1 + 23x_2 + 11x_3 + 35x_4 \\ & 2x_1 + 1.5x_2 + 0.5x_3 + 2.5x_4 \le 100 \\ & 0.5x_1 + 0.25x_2 + 0.25x_3 + x_4 \le 50 \\ & x_1 \ge 0, x_2 \ge 0, x_3 \ge 0, x_4 \ge 0. \end{array}$$

Questa formulazione è un problema matematico ben definito e costituisce il modello di Programmazione Matematica rappresentante il problema di pianificazione della produzione industriale in analisi. Si tratta, in questo caso, di un problema di programmazione lineare. □

Esempio 2.4.2 *Un'industria deve costruire un silos di forma cilindrica. Tale silos deve essere posto in un magazzino appoggiato su una delle basi. Tale magazzino è a pianta rettangolare di dimensioni metri* 20×10 *ed ha un tetto spiovente lungo il lato di 10 metri, che ha altezza massima di metri* 5 *e altezza minima di metri* 3. *Per costruire questo silos deve essere usato del materiale plastico sottile flessibile che può essere tagliato, modellato e incollato saldamente. Sapendo che si dispone di non più di* $200\ m^2$ *di tale materiale plastico si costruisca un modello che permetta di determinare le dimensioni del silos (raggio di base ed altezza) in modo da massimizzare la quantità di liquido che può esservi contenuto.*

Analisi del problema e costruzione del modello.

Si tratta di determinare il dimensionamento ottimale di un contenitore cilindrico per uso industriale cercando di massimizzare il suo volume tenendo presente che deve essere contenuto in un magazzino di dimensioni fissate.

Si devono definire formalmente le variabili di decisione, l'insieme delle soluzioni ammissibili e la funzione obiettivo.

– *Variabili di decisione.* È immediato introdurre due variabili x e y che rappresentano rispettivamente la lunghezza (in metri) del raggio di base e dell'altezza del contenitore cilindrico.

– *Funzione obiettivo.* La funzione obiettivo è rappresentata dal volume del contenitore cilindrico ed è data da

$$\pi x^2 y.$$

– *Vincoli.* Il diametro della base non può superare le dimensioni del magazzino e quindi deve essere

$$2x \le 10.$$

La limitazione dell'altezza del contenitore varia al variare del diametro di base in quanto il tetto è spiovente. Dato che la pendenza del tetto è del 20%, dovrà risultare

$$y \le 5 - 0.2 \cdot 2x.$$

Inoltre disponendo solo di una quantità limitata di materiale plastico la superficie totale del contenitore cilindrico non può superare $200m^2$ e quindi deve risultare

$$2\pi x^2 + 2\pi xy \le 200.$$

Si devono infine esplicitare i vincoli di non negatività $x \ge 0$, $y \ge 0$.

La formulazione complessiva risulta quindi

$$\begin{array}{ll} \max & x^2 y \\ & x \le 5 \\ & y \le 5 - 0.2 \cdot 2x \\ & 2\pi x^2 + 2\pi xy \le 200 \\ & x \ge 0, \quad y \ge 0. \end{array}$$

Il modello è quindi un modello di programmazione non lineare. □

Capitolo 3

Problemi di ottimizzazione convessa e concava

Tra i problemi di Ottimizzazione sono di particolare interesse i cosiddetti problemi *convessi.* Per poter definire correttamente cosa si intende per problema convesso è necessario introdurre alcuni concetti preliminari.

3.1 Insiemi Convessi

Definizione 3.1.1 *Un vettore y in $\mathbb{R}^n$ si dice* combinazione convessa *di $p \geq 1$ vettori $x^1, \ldots, x^p$ se risulta*

$$y = \sum_{i=1}^{p} \lambda_i x^i \qquad \sum_{i=1}^{p} \lambda_i = 1 \qquad \lambda_i \geq 0 \quad i = 1, \ldots, p.$$

Possiamo allora introdurre la seguente definizione

Definizione 3.1.2 *Siano x e y due punti in $\mathbb{R}^n$. L'insieme dei punti di $\mathbb{R}^n$ ottenuti come*

$$z = (1 - \beta)x + \beta y,$$

al variare di β nell'intervallo $[0, 1]$ viene definito come segmento chiuso *di estremi x e y e viene sinteticamente indicato con $[x, y]$.*

Esempio 3.1.3 Nella figura 3.1 è rappresentato il segmento in $\mathbb{R}^2$ avente per estremi i punti $x = (1, 1)^T$ e $y = (8, 5)^T$. Per $\beta = 0$ ritroviamo il punto x, mentre per $\beta = 1$ ritroviamo il punto y; i punti segnati nella figura come x_a, x_b e x_c corrispondono rispettivamente a valori di β pari a 0.25, 0.5 e 0.75.

Dalla figura 3.1 risulta ovvio che il concetto di segmento è la generalizzazione, al caso di $\mathbb{R}^n$ dell' usuale concetto di segmento valido nel piano.

Notiamo anche come, nel caso in cui gli estremi appartengano ad $\mathbb{R}^1$, e sono quindi due numeri (scalari), diciamo a e b, il concetto di segmento di estremi a e b coincida con quello di intervallo $[a, b]$, fatto che giustifica la notazione $[x, y]$ impiegata per indicare il segmento.

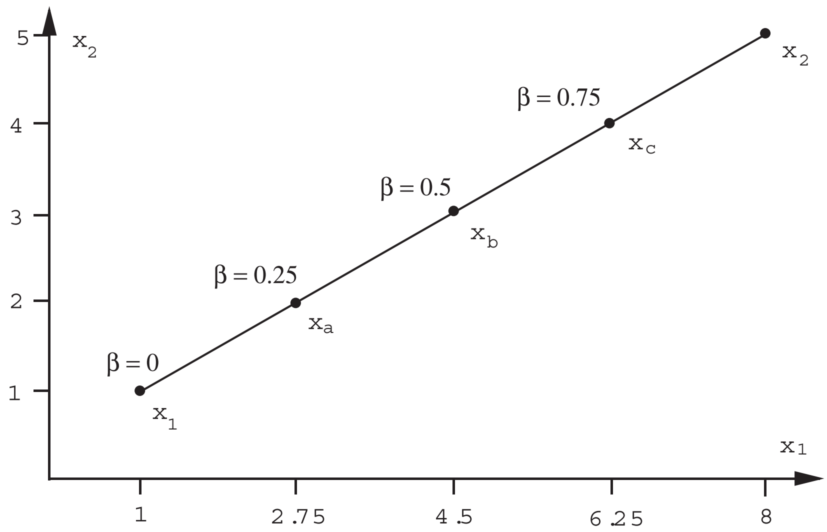

Figura 3.1: Esempio di segmento.

Definizione 3.1.4 *Un insieme* $X \subseteq I\!R^n$ *è* convesso *se per ogni coppia di punti appartenenti all'insieme, appartengono all'insieme anche tutti i vettori ottenibili come loro combinazione convessa.*

Utilizzando il concetto di segmento chiuso, la definizione di insieme convesso può essere riformulata nel modo seguente:

Un insieme X *è convesso se per ogni coppia di vettori* $x, y \in X$ *si ha* $[x, y] \subseteq X$.

Dalla definizione segue che l'insieme vuoto e l'insieme costituito da un solo vettore sono insiemi convessi (banali). Il più semplice insieme convesso non banale è il *segmento* di estremi $x, y \in I\!R^n$.

Esempio 3.1.5 In $I\!R^2$ gli insiemi (a) , (b) della figura 3.2 sono convessi, mentre gli insiemi (c), (d) della stessa figura non lo sono. Infatti agli insiemi (c),(d) appartengono coppie di punti, quali quelle segnate nella figura, tali che il segmento che li congiunge presenta dei punti non appartenenti all'insieme; ciò non avviene invece comunque si prendano coppie di punti negli insiemi (a) e (b).

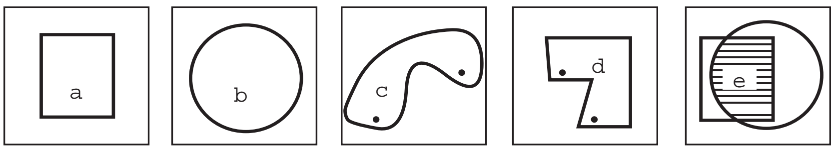

Figura 3.2: Insiemi convessi e non convessi.

Una importante proprietà degli insiemi convessi è espressa dal seguente teorema.

Teorema 3.1.6 *L'intersezione di due insiemi convessi è un insieme convesso.*

Dimostrazione: Siano $X_1, X_2 \subseteq I\!R^n$ due insiemi convessi e sia $X = X_1 \cap X_2$ la loro intersezione. Siano x ed y due vettori in X, allora $x, y \in X_1$ ed $x, y \in X_2$. Poiché X_1 ed X_2 sono insiemi convessi abbiamo che $[x, y] \subseteq X_1$ e che $[x, y] \subseteq X_2$. Ma allora $[x, y] \subseteq X$ e l'insieme X è convesso $\square$

Esempio 3.1.7 L'insieme (e) della figura 3.2 è dato dall'intersezione di due insiemi convessi ed è convesso

Dal Teorema (3.1.6) si può derivare, con un semplice ragionamento induttivo, il seguente corollario.

Corollario 3.1.8 *L'intersezione di un numero finito di insiemi convessi è un insieme convesso.*

Passiamo ora a considerare dei particolari insiemi convessi che rivestono un ruolo importante nella teoria della programmazione lineare.
Consideriamo l'espressione

$$a_1x_1 + \quad \dots \quad + a_nx_n = b$$

che, con notazione vettoriale possiamo scrivere $a^Tx = b$, con $a \in I\!R^n$ e $b \in I\!R$. Possiamo allora dare le seguenti definizioni.

Definizione 3.1.9 *Sia a un vettore di $I\!R^n$ e b un numero reale. L'insieme*

$$H = \{x \in I\!R^n : a^Tx = b\}$$

è detto iperpiano definito dall'equazione $a^Tx = b$. Gli insiemi

$$\begin{aligned} S^{\leq} &= \{x \in I\!R^n : a^Tx \leq b\} \\ S^{\geq} &= \{x \in I\!R^n : a^Tx \geq b\} \end{aligned}$$

sono detti semispazi chiusi definiti dalle disequazioni $a^Tx \leq b$ e $a^Tx \geq b$.

Nel caso dello spazio $I\!R^2$ il concetto di iperpiano coincide con quello di retta, mentre nel caso dello spazio $I\!R^3$ il concetto di iperpiano coincide con quello di piano. In maniera intuitiva, i semispazi possono essere pensati come l' insieme dei punti che "giacciono" da una stessa parte rispetto all'iperpiano.

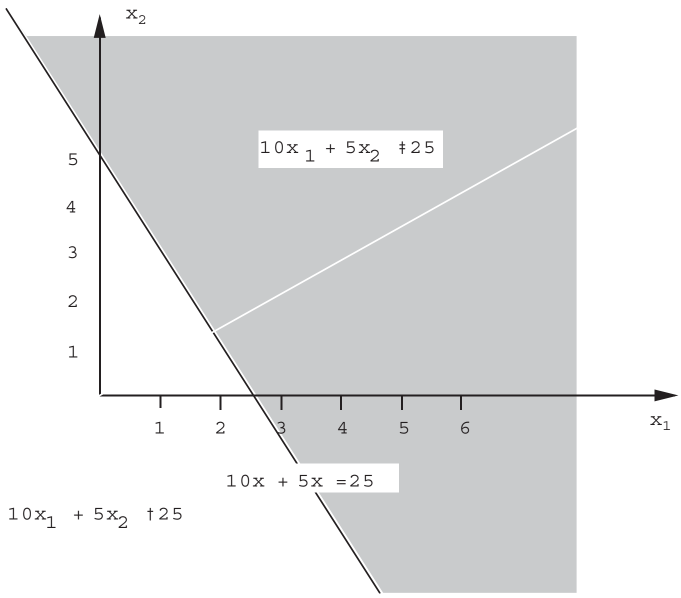

Figura 3.3: Retta e semipiani individuati da un'equazione lineare.

Esempio 3.1.10 Con riferimento alla figura 3.3, l'iperpiano (= retta) $10x_1 + 5x_2 = 25$ divide lo spazio (= piano) in due semispazi: $S^{\geq} = \{x \in I\!R^2 : 10x_1 + 5x_2 \geq 25\}$, indicato in grigio nella figura, e $S^{\leq} = \{x \in I\!R^2 : 10x_1 + 5x_2 \leq 25\}$, indicato in bianco nella figura.

Notiamo che l'iperpiano H fa parte di tutti e due i semispazi e che l'intersezione dei due semispazi coincide con l'iperpiano. In termini insiemistici abbiamo che

$$H \subset S^{\geq}, \qquad H \subset S^{\leq}, \qquad S^{\geq} \cap S^{\leq} = H.$$

I semispazi e gli iperpiani sono insiemi convessi.

Teorema 3.1.11 *Un semispazio chiuso è un insieme convesso.*

Dimostrazione: Dimostreremo il teorema per un semispazio $S^{\leq} = \{x \in I\!R^n : a^T x \leq b\}$, la dimostrazione per il semispazio $S^{\geq}$ ottenuto invertendo il verso della disequazione è analoga. Consideriamo due generici vettori x ed y appartenenti ad $S^{\leq}$, vogliamo dimostrare che ogni vettore $z \in [x, y]$ appartiene ad $S^{\leq}$, ovvero soddisfa la relazione $a^T z \leq b$.

Sia $z = \beta x + (1 - \beta) y$ con $0 \leq \beta \leq 1$. Poiché x ed y appartengono ad $S^{\leq}$ abbiamo che $a^T x \leq b$ e $a^T y \leq b$. Inoltre, poiché β ed $1 - \beta$ sono reali non negativi abbiamo che

$$a^T(\beta x + (1-\beta)y) = \beta a^T x + (1-\beta) a^T y \leq \beta b + (1-\beta) b = b$$

e quindi che $a^T z \leq b$ $\square$

Utilizzando il Teorema (3.1.11) e il Teorema (3.1.6) e ora facile dimostrare che anche un iperpiano è un insieme convesso.

Corollario 3.1.12 *Un iperpiano è un insieme convesso.*

Dimostrazione: Un iperpiano è l'intersezione di due semispazi chiusi ($S^{\leq}$ e $S^{\geq}$). Per il Teorema (3.1.11) un semispazio chiuso è un insieme convesso mentre, per il Teorema (3.1.6), l'intersezione di due insiemi convessi è un insieme convesso. $\square$

In particolare è usuale introdurre la seguente definizione:

Definizione 3.1.13 *Un insieme $P \subseteq I\!R^n$ è un* poliedro *se è l'intersezione di un numero finito di semispazi chiusi e iperpiani.*

In $I\!R^3$ possibili poliedri sono i cubi, i tetraedi ecc. In R^2 sono poliedri, ad esempio, i segmenti, le rette, poligoni piani, ecc.

Naturalmente, risulta:

Un poliedro è un insieme convesso.

Introduciamo ora il concetto di *punto estremo*, che ha un ruolo molto importante nello studio dei problemi di Programmazione lineare.

Definizione 3.1.14 (Punto estremo) *Sia $C \subseteq R^n$ un insieme convesso. Un punto $x \in C$ si dice* punto estremo *di C se* non *può essere espresso come combinazione convessa di due punti di C distinti da x, o, equivalentemente, se non esistono $y, z \in C$ con $z \neq y$ tali che*

$$x = (1-\lambda)y + \lambda z, \quad \text{con} \quad 0 < \lambda < 1.$$

Sono esempi di punti estremi: per un cerchio, i punti della circonferenza che lo delimita; per un segmento $[x_1, x_2]$, gli estremi x_1, x_2; per un triangolo, i suoi vertici (non sono punti estremi i punti dei lati che non siano vertici). L'insieme dei punti estremi di un insieme convesso C verrà indicato con il simbolo $\text{Ext}(C)$, ossia:

$$\text{Ext}(C) = \{x \in C : x \ \text{ è punto estremo di } \ C\}.$$

Un insieme convesso può non ammettere punti estremi. Ad esempio, un iperpiano, un semispazio, una sfera aperta non hanno punti estremi.

Osserviamo inoltre che $\text{Ext}(C)$ è contenuto nella frontiera di C e quindi nessun insieme convesso aperto può ammettere punti estremi.

Esempio 3.1.15 Nell'insieme di figura 3.4 il punto A non è un punto estremo, in quanto è interno al segmento che congiunge i punti B e C, anch'essi appartenenti all'insieme; lo stesso vale per il punto D, interno al segmento [E,F]. Sono invece punti estremi dell'insieme i punti E, F, G, H.

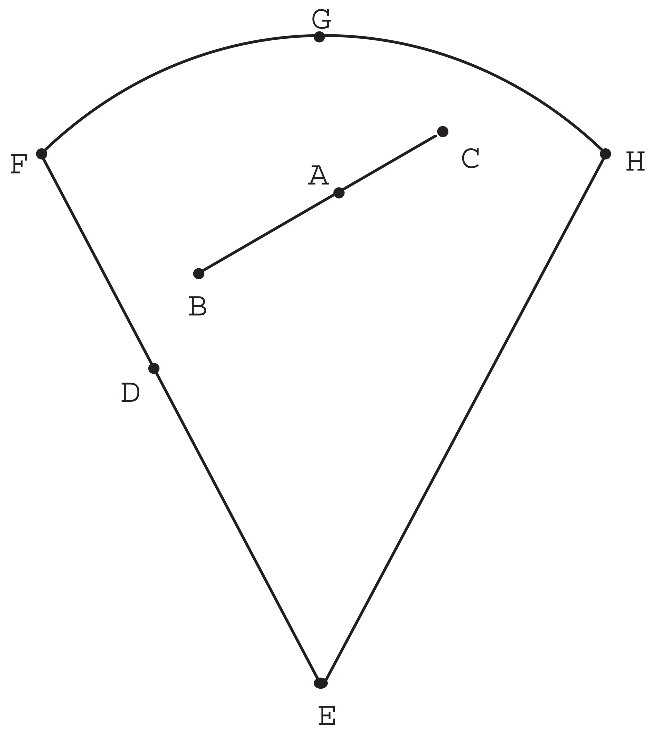

Figura 3.4: Punti estremi di un insieme.

Nel caso in cui l'insieme convesso sia un poliedro, i punti estremi sono detti anche *vertici*.[1]

[1] In realtà la definizione di vertice è concettualmente diversa, ma si dimostra che un punto di un poliedro è punto estremo se e solo se è un vertice.

3.2 Funzioni convesse e concave

Definizione 3.2.1 *Una funzione $f(x)$ si dice* convessa *su un insieme convesso $\mathcal{C}$ se, presi comunque due punti $y, z \in \mathcal{C}$ risulta che:*

$$f((1-\beta)y + \beta z) \leq (1-\beta)f(y) + \beta f(z), \quad \beta \in [0,1]. \tag{3.1}$$

La funzione $f(x)$ si dice poi strettamente convessa *se, per $y, z \in \mathcal{C}$, $y \neq z$,risulta*

$$f((1-\beta)y + \beta z) < (1-\beta)f(y) + \beta f(z), \quad \beta \in (0,1).$$

Una funzione $f(x)$ si dice *(strettamente) concava* su un insieme convesso $\mathcal{C}$ se la funzione $-f(x)$ è (strettamente) convessa su $\mathcal{C}$.

Definizione 3.2.2 *Una funzione $f(x)$ si dice* concava *su un insieme convesso $\mathcal{C}$ se, presi comunque due punti $y, z \in \mathcal{C}$ risulta che:*

$$f((1-\beta)y + \beta z) \geq (1-\beta)f(y) + \beta f(z), \quad \beta \in [0,1]. \tag{3.2}$$

La funzione $f(x)$ si dice poi strettamente concava *se, per $y, z \in \mathcal{C}$, $y \neq z$,risulta*

$$f((1-\beta)y + \beta z) > (1-\beta)f(y) + \beta f(z), \quad \beta \in (0,1).$$

Nella (3.2) $y, z, f(y), f(z)$ sono dati, e β varia tra 0 e 1. Se mettiamo in esplicita evidenza la dipendenza da β, introducendo la funzione

$$\phi(\beta) = f((1-\beta)y + \beta z)$$

otteniamo per una funzione strettamente convessa che:

$$\phi(\beta) < (1-\beta)\phi(0) + \beta\phi(1) \quad \beta \in (0,1)$$

Quest'ultima relazione mette in evidenza che, se si rappresenta graficamente nel piano (β, ϕ) la funzione $\phi(\beta)$, il grafico della funzione nell'intervallo $(0,1)$ si trova al di sotto del segmento, detto *secante*, che congiunge i punti $(0, \phi(0))$ e $(1, \phi(1))$ e coincide solo negli estremi del segmento. Si può concludere che una funzione strettamente convessa è caratterizzata dalla proprietà di avere il grafico sempre al di sotto di ogni sua secante.

Una funzione lineare del tipo $c^T x + b$ è sia convessa che concava (ma NON è strettamente convessa o strettamente concava).

Se f è una funzione convessa differenziabile possiamo dare condizioni necessarie e sufficienti di convessità espresse per mezzo delle derivate prime o seconde dalla funzione. Ci limitiamo a riportare i risultati seguenti.

Teorema 3.2.3 (Condizioni necessarie e sufficienti di convessità) *Sia C un insieme convesso aperto, sia $f : C \to R$ e supponiamo che ∇f sia continuo su C. Allora f è convessa su C se e solo se, per tutte le coppie di punti $x, y \in C$ si ha:*

$$f(y) \geq f(x) + \nabla f(x)^T (y - x). \tag{3.3}$$

Inoltre f è strettamente convessa su C se e solo se, per tutte le coppie di punti $x, y \in C$ con $y \neq x$, si ha:

$$f(y) > f(x) + \nabla f(x)^T (y - x). \tag{3.4}$$

Dal punto di vista geometrico, la condizione del teorema precedente esprime il fatto che una funzione è convessa su C se e solo se in un qualsiasi punto $y \in C$ l'ordinata $f(y)$ della funzione non è inferiore alle ordinate dei punti del piano tangente al grafo della funzione in un qualsiasi altro punto x di C.

Nel teorema successivo riportiamo una condizione necessaria e sufficiente di convessità espressa per mezzo delle derivate seconde.

Teorema 3.2.4 (Condizioni necessarie e sufficienti di convessità) *Sia C un insieme convesso aperto, sia $f : C \to R$ e supponiamo che la matrice Hessiana $\nabla^2 f$ sia continua su C. Allora f è convessa su C se e solo se, per ogni $x \in C$, la matrice $\nabla^2 f(x)$ è semidefinita positiva.*

Se si prendono in considerazione le derivate seconde, non è vero, in generale, che una condizione necessaria di convessità stretta è la definita positività della matrice Hessiana (basti pensare alla funzione $y = x^4$ in $x = 0$). Si può stabilire tuttavia che se la matrice Hessiana è definita positiva allora f è strettamente convessa.

Teorema 3.2.5 (Condizione sufficiente di convessità stretta) *Sia C un insieme convesso aperto, sia $f : C \to R$ e supponiamo che la matrice Hessiana $\nabla^2 f$ sia continua e definita positiva su C. Allora f è strettamente convessa su C.*

3.2.1 Funzioni quadratiche

Tra le funzioni di particolare interesse in problemi di Programmazione Matematica ci sono le funzioni quadratiche.

Le *funzione quadratica* è una funzione del tipo

$$q(x) = \frac{1}{2}x^T Ax + c^T x$$

Data una matrice A quadrata e simmetrica di dimensione $(n \times n)$, con elementi $a_{ij} = a_{ji}$, $i, j = 1, \dots, n$, si definisce *forma quadratica* associata alla matrice A la funzione

$$x^T Ax = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j. \tag{3.5}$$

Una forma quadratica è quindi una particolare funzione quadratica in cui il termine lineare è identicamente nullo ($c = 0$).

Si verifica facilmente che il gradiente e l'Hessiano della forma quadratica sono dati rispettivamente da Ax e A.

La forma quadratica $x^T Ax$ si dice:

- *definita positiva*, se risulta $x^T Ax > 0 \ \forall x \in I\!R^n, x \neq 0$;
- *semidefinita positiva* se risulta $x^T Ax \geq 0 \ \forall x \in I\!R^n$;
- *indefinita* se per qualche x risulta $x^T Ax > 0$, e per altri x risulta $x^T Ax < 0$.

Corrispondentemente, si dice che la matrice A associata alla forma quadratica è rispettivamente definita positiva, semidefinita positiva, indefinita.

La forma quadratica $x^T Ax$ si dice (semi)definita negativa se $-x^T Ax$ è (semi)definita positiva.

Nel caso di funzioni quadratiche, la matrice hessiana A è costante e la condizione di convessità dei Teoremi 3.2.4 e diventa:

Teorema 3.2.6 *Una funzione quadratica $q(x)$ è convessa su $\mathbb{R}^n$ se, e solo se, risulta:*

$$\frac{1}{2}y^T Ay \geq 0, \quad \text{per ogni } y \in \mathbb{R}^n;$$

inoltre, la funzione $q(x)$ è strettamente convessa su $\mathbb{R}^n$ se e solo se risulta

$$\frac{1}{2}y^T Ay > 0, \quad \text{per ogni } y \in \mathbb{R}^n.$$

Notare che nel caso di funzione quadratica la condizione di stretta convessità è condizione necessaria e sufficiente.

Per verificare se una matrice A è definita positiva, si può utilizzare un semplice test.

Criterio 1 *Siano $A_k, k = 1\ldots, n$ gli n minori principali della matrice A, detti* sottomatrici principali di nord-ovest*, cioè le n sottomatrici con elementi $a_{ij}, i, j = 1, \ldots, k$,ottenute da A eliminando le ultime $n-k$ righe e colonne. Denotato con $\det A_k$ il determinante di A_k, risulta che:*

- *A è definita positiva se, e solo se, $\det A_k > 0$, per $k = 1, \ldots, n$.*

Si osserva che se A è semidefinita positiva allora risulta $\det A_k \geq 0$, ma non è vero in generale il viceversa. Basta prendere come esempio la matrice

$$A = \begin{pmatrix} 0 & 0 \\ 0 & 2a_{22} \end{pmatrix}$$

con $a_{22} < 0$ e la forma quadratica associata

$$a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2$$

in cui $a_{11} = a_{12} = 0$. I minori principali A_1 e A_2 hanno determinante nullo, e quindi soddisfano il criterio $\det A_k \geq 0$, ma $q(x)$ è semidefinita negativa !!!

Per verificare se una matrice A è semidefinita positiva si deve applicare un criterio molto più oneroso riportato di seguito.

Criterio 2 *Siano D_{i_k,j_k} le sottomatrici di A con elementi a_{ij},ottenute da A eliminando $n-k$ righe e colonne in tutti i possibili modi, dette* minori principali *di A, ovvero le sottomatrici con elementi $a_{ij}, i = i_1, \ldots, i_k,\ j = j_1, \ldots, j_k$, per $1 \leq k \leq n$. Denotato con $\det D_{i_k,j_k}$ il determinante di D_{i_k,j_k}, si ha:*

- *A è semidefinita positiva se, e solo se, $D_{i_k,j_k} \geq 0$ per ogni $1 \leq k \leq n$.*

Per analizzare se una matrice A è definita/semidefinita negativa si possono applicare i criteri precedenti alla matrice $-A$ associata alla forma quadratica.

Un altro test per verificare il segno di una forma quadratica consiste nel determinare gli *autovalori* della matrice A, cioe' i valori $\alpha_i,\ i = 1, \ldots, n$ che risolvono l'equazione di grado n:

$$\det(A - \alpha I) = 0,$$

ove I è la matrice identica di ordine n. Se la matrice A è simmetrica, si ha che gli autovalori sono tutti reali. Risulta che:

- A è definita positiva se, e solo se, $\alpha_i > 0$, per $i = 1, \ldots, n$;
- A è semidefinita positiva se, e solo se, $\alpha_i \geq 0$ per $i = 1, \ldots, n$;
- A è indefinita, altrimenti.

È evidente che il test basato sui minori è di più semplice impiego di quello basato sugli autovalori: infatti calcolare determinanti, fino all'ordine n, è piu' semplice che risolvere un'equazione di grado n.

Nel seguito indicheremo $\alpha_{\min}(A)$ e $\alpha_{\max}(A)$ rispettivamente il più piccolo e il più grande autovalore di una matrice A.

Esempio 3.2.7 Un esempio di funzione quadratica strettamente convessa è la funzione

$$4x_1^2 + 8x_2^2 - x_1$$

rappresentata in figura 3.5.

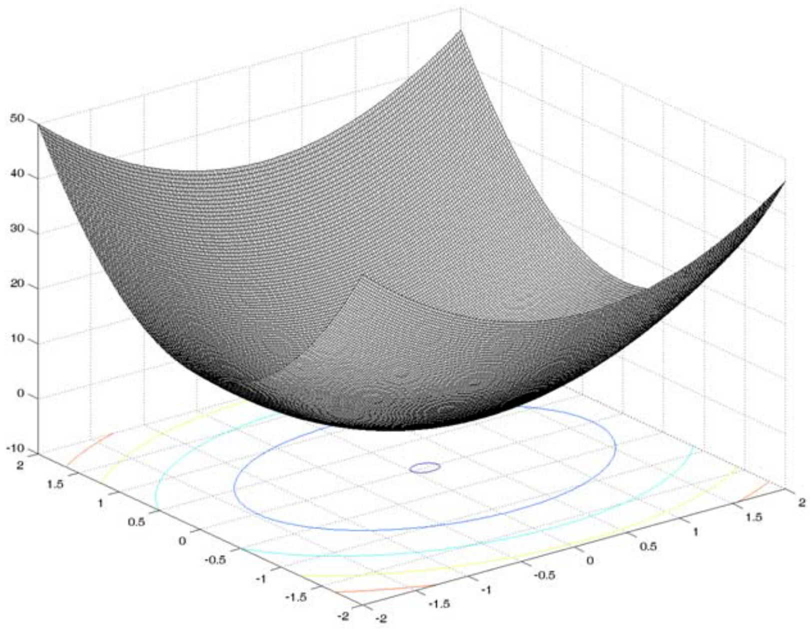

Figura 3.5: Grafico della funzione quadratica convessa Dell'Esempio 3.2.7.

3.3 Problemi di ottimizzazione

Dal punto di vista delle proprietà di convessità possiamo distinguere:

- problemi di programmazione convessa: sono i problemi di minimo in cui la funzione obiettivo è convessa e l'insieme ammissibile è un insieme convesso (o anche i problemi di massimo in cui la funzione obiettivo è concava e l'insieme ammissibile è convesso)
- problemi di programmazione concava: sono i problemi di minimo in cui la funzione obiettivo è concava e l'insieme ammissibile è un insieme convesso (o anche i problemi di massimo in cui la funzione obiettivo è convessa e l'insieme ammissibile è convesso)
- problemi generali, in cui non sono soddisfatte tali condizioni.

Riconoscere che un problema di ottimizzazione è convesso o concavo fornisce importanti informazioni qualitative sulle sue soluzioni. Osserviamo che

Una problema di Programmazione Lineare è sia convesso che concavo.

Esempio 3.3.1 Sia dato il problema

$$
\begin{array}{ll}
\max & x_2 \\
 & x_2 - x_1^3 \geq 0 \\
 & x_1 + x_2 \leq 1 \\
 & x_1, x_2 \geq 0,
\end{array}
$$

dire se è convesso.
Si tratta di un problema di massimo con funzione obiettivo lineare e quindi concava. L'insieme ammissibile è rappresentato in figura 3.6. Si tratta di un insieme convesso. Quindi il problema dato è convesso.

Notiamo che, poiché la funzione obiettivo è lineare, si tratta ache di un problema di massimizzazione di una funzione convessa su insieme convesso e cioè di un problema concavo !

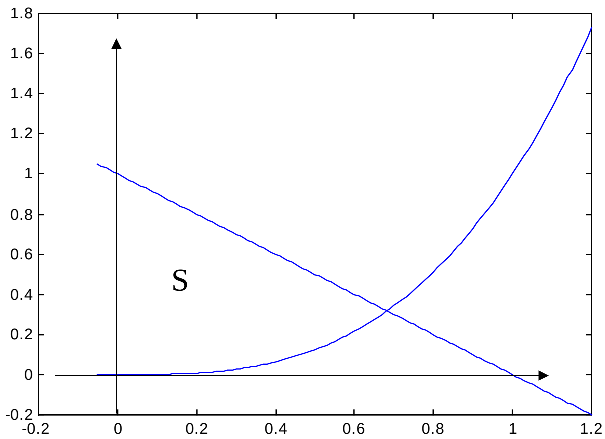

Figura 3.6: Insieme ammissibile S dell'Esempio 3.3.1

Per riconoscere se un problema generale è convesso o concavo dobbiamo verificare che S sia un insieme convesso, e che $f(x)$ sia convessa/concava su S, il che non è sempre facile. Consideriamo un insieme S definito da vincoli di uguaglianza e disuguaglianza:

$$
\begin{array}{ll}
g_i(x) \leq 0, & i = 1, \dots, m \\
h_j(x) = 0, & j = 1, \dots, p.
\end{array}
\tag{3.6}
$$

Nel caso in cui g_i per $i = 1, \dots, m$ e h_j per $j = 1, \dots, p$ siano funzioni lineari, S è un poliedro e quindi un insieme convesso. Se invece qualche g_i o h_j è non lineare possiamo utilizzare la seguente proposizione fornisce una condizione solo sufficiente per la convessità di S.

Teorema 3.3.2 *Sia*

$$S = \{x \in R^n : g_i(x) \leq 0, \quad i = 1, \dots, m \quad h_j(x) = 0, \quad j = 1, \dots, p\}.$$

Se per ogni i le funzioni $g_i(x)$ sono convesse in $I\!R^n$, e per ogni j le funzioni h_j sono funzioni del tipo $a_j^T x - b_j$, allora l'insieme S è convesso.

Dimostrazione. L'insieme ammissibile S è ottenuto come intersezione degli insiemi $S^i = \{x \in R^n : \ g_i(x) \leq 0\}$ e $H^j = \{x \in R^n : \ a_j^T x - b_j = 0\}$. Dimostriamo quindi che S^i e H^j sono insiemi convessi. Il risultato segue poi dal Corollario 3.1.8.
Consideriamo l'insieme $S^i = \{x \in R^n : \ g_i(x) \leq 0\}$ e siano y e z due punti appartenenti a S^i. Per la convessità di g_i si ha

$$g_i((1-\beta)y + \beta z) \leq (1-\beta)g_i(y) + \beta g_i(z) \qquad \text{con } \beta \in [0,1],$$

Quindi poiché $g_i(y) \leq 0$ e $g_i(z) \leq 0$, si ottiene

$$g_i((1-\beta)y + \beta z) \leq 0 \quad \text{con } \beta \in [0,1],$$

Ovvero tutti i punti del segmento $[y, z]$ sono in S^i che dimostra che l'insieme S^i è convesso.
Per quanto riguarda gli insiemi H^j, essi sono convessi per il Corollario 3.1.12.
Poichè l'intersezione di due insiemi convessi è un insieme convesso, possiamo concludere che, nelle ipotesi poste, l'insieme S risulta convesso. □

Otteniamo quindi la seguente condizione sufficiente per riconoscere se un problema è convesso o concavo.

Teorema 3.3.3 *Sia dato il Problema*

$$\begin{array}{ll} \min & f(x) \\ & g_i(x) \leq 0, \quad i = 1, \dots, m \\ & h_j(x) = 0, \quad j = 1, \dots, p. \end{array} \tag{3.7}$$

Si assuma che i vincoli di disuguaglianza siano dati da funzioni $g_i(x)$ convesse in $I\!R^n$, e che i vincoli di uguaglianza siani dati da funzioni del tipo $a_j^T x - b_j$.
Se la funzione obiettivo $f(x)$ è una funzione convessa in $I\!R^n$, il Problema (3.7) è convesso.
Se la funzione obiettivo $f(x)$ è una funzione concava in $I\!R^n$, il Problema (3.7) è concavo.

Una problema con insieme ammissibile convesso S e funzione obiettivo lineare è sia convesso che concavo.

Notiamo che, la condizione sufficiente del Teorema 3.3.2 non è soddisfatta nell'Esempio 3.3.1 a causa della non convessità del vincolo $x_1^3 - x_2 \leq 0$. □

I problemi di ottimizzazione convessi sono di particolare importanza per due motivi. Il primo è che la grande maggioranza dei problemi di ottimizzazione che si incontrano nella pratica sono convessi. Il secondo è che la convessità induce alcune proprietà che semplificano l'analisi e la soluzione di un problema convesso.

Una delle proprietà più significative è la seguente:

Teorema 3.3.4 **[Assenza di ottimi locali]** *Sia $S \subseteq R^n$ un insieme convesso e f una funzione convessa su S. Allora, il problema*

$$\begin{array}{ll} \min & f(x) \\ & x \in S \end{array}$$

o non ha soluzione, o ha solo soluzioni globali; non può avere soluzioni esclusivamente locali.

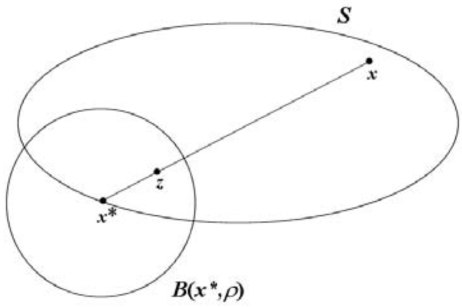

Figura 3.7: Costruzione geometrica utilizzata nella dimostrazione del Teorema 3.3.4

Dimostrazione. La dimostrazione è per assurdo. Ammettiamo che esista un minimo globale e che x^* sia una soluzione locale, ma non globale, di $min_{x \in S} f(x)$. Allora esisterà un altro punto $x \in S$ tale che $f(x) < f(x^*)$.

La costruzione utilizzata nella dimostrazione è illustrata schematicamente nella figura 3.7.

Consideriamo il segmento $[x^*, x]$: per la convessità di f, si ha:

$$f((1-\beta)x^* + \beta x) \leq (1-\beta)f(x^*) + \beta f(x) = f(x^*) + \beta(f(x) - f(x^*)), \text{per ogni } \beta \in [0,1].$$

Il termine $\beta(f(x) - f(x^*))$ risulta < 0, per ogni $\beta \in (0,1]$, e si annulla solo per $\beta = 0$. Quindi, poichè in ogni punto $x \in (x^*, x]$ risulta $f(z) < f(x^*)$, non esiste nessun intorno di raggio $\rho > 0$ in cui x^* può soddisfare la definizione di minimo locale. $\square$

I problemi di programmazione concava costituiscono una classe molto ampia di problemi non convessi, che include numerose classi di problemi inerentemente "difficili". La difficoltà principale che si manifesta nella soluzione di problemi di tipo concavo risiede nel fatto che possono esssere presenti molti punti di minimo locale che non sono punti di minimo globale. In casi del genere la ricerca delle soluzioni globali può divenire un problema di natura combinatoria, che può richiedere, nel caso peggiore, tempi di calcolo esponenzialmente crescenti con le dimensioni del problema (È possibile dimostrare, in particolare, che, sotto opportune ipotesi, molti problemi di ottimizzazione combinatoria possono essere riformulati come problemi (continui) di programmazione concava.).

Per i problemi di programmazione concava è possibile dimostrare che le soluzioni ottime, ove esistano, appartengono alla frontiera dell'insieme ammissibile. Più precisamente vale il risultato seguente.

Teorema 3.3.5 [Assenza di soluzioni ottime interne] *Sia $S \subseteq R^n$ un insieme convesso e f una funzione concava su S. Allora, se esiste un punto di minimo globale del problema*

$$\begin{array}{ll} \min & f(x) \\ & x \in S \end{array}$$

e se la funzione obiettivo non è costante su S, ogni punto di minimo globale appartiene alla frontiera di S.

Dimostrazione. Supponiamo che il problema ammetta soluzione e che x^* sia una soluzione ottima. Poichè, per ipotesi, f non è costante su S deve esistere un punto $\hat{x} \in S$ tale che

$$f(\hat{x}) > f(x^*) = \min_{x \in S} f(x).$$

Supponiamo ora che $z \in S$ sia un punto interno all'insieme ammissibile. Deve allora esistere una sfera aperta $B(z; \rho)$ con centro in z e raggio $\rho > 0$ tutta contenuta in S. Sulla retta congiungente $\hat{x}$ con z possiamo allora determinare un $y \in B(z; \rho) \subseteq S$ tale che z appartenga al segmento $[\hat{x}, y]$ e risulti $y \neq z$, ossia possiamo trovare un λ con $0 \leq \lambda < 1$ tale che

$$z = (1 - \lambda)\hat{x} + \lambda y.$$

Per la concavità di f e l'ipotesi che sia $f(\hat{x}) > f(x^*)$, tenendo conto del fatto che $f(y) \geq f(x^*)$ e che $1 - \lambda > 0$ (perchè $y \neq z$), si ottiene:

$$f(z) \geq (1 - \lambda)f(\hat{x}) + \lambda f(y) > (1 - \lambda)f(x^*) + \lambda f(x^*) = f(x^*).$$

Ciò dimostra che $f(z) > f(x^*)$ e quindi che non può esistere una soluzione ottima in un punto interno. $\square$

Per quanto riguarda la Programmazione Lineare, poiché si tratta di un problema sia convesso che concavo, abbiamo, come immediata conseguenza, il seguente corollario

Corollario 3.3.6 *Se un problema di Programmazione Lineare ammette soluzione, allora il minimo globale si trova sulla frontiera del poliedro ammissibile.*

Nel seguito vedremo che nel caso in cui l'insieme ammissibile sia un poliedro, vale una proprietà più forte. In particolare, nel caso di problemi di Programmazione Lineare (Capitolo 5), almeno un punto di minimo globale deve trovarsi su un vertice del poliedro ammissibile.

Capitolo 4

Modelli di Programmazione Lineare e soluzione grafica

In questo capitolo esaminiamo in modo più dettagliato la Programmazione Lineare. In particolare saranno presentati alcuni modelli di PL più o meno classici, sarà illustrata una tecnica risolutiva per il caso di due sole variabili che aiuta a comprendere alcune delle caratteristiche più importanti dei problemi di Programmazione Lineare.

4.1 Struttura di un problema di Programmazione Lineare

Come abbiamo già visto nel Capitolo 2 un problema di Programmazione Lineare è caratterizzato da una funzione obiettivo lineare (da minimizzare o massimizzare) della forma

$$f(x) = c_1x_1 + \ldots + c_nx_n = \sum_{j=1}^{n} c_jx_j$$

e da un numero finito m di vincoli lineari della forma

$$\begin{array}{cccc} a_{11}x_1+ & \ldots & +a_{1n}x_n & \simeq b_1 \\ a_{21}x_1+ & \ldots & +a_{2n}x_n & \simeq b_2 \\ \vdots & \ldots & \vdots & \vdots \\ a_{m1}x_1+ & \ldots & +a_{mn}x_n & \simeq b_m. \end{array} \tag{4.1}$$

dove con $\simeq$ intendiamo $\leq, \geq$ oppure $=$.
Introducendo il vettore $c \in \mathbb{R}^n$, definito $c = (c_1, \ldots, c_n)^T$, $x \in \mathbb{R}^n$ definito $x = (x_1, \ldots, x_n)^T$, il vettore $b = (b_1, \ldots, b_m)^T$ e la matrice $(m \times n)$

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

un generico problema di Programmazione Lineare può essere scritto nella forma

$$\begin{array}{ll} \min & c^Tx \\ & Ax \simeq b. \end{array}$$

Si osservi che, indicando con a_i^T, $i = 1, \ldots, m$, le righe della matrice A, ciascun vincolo del problema, ovvero ciascuna disuguaglianza della (4.1) può essere scritto nella forma $a_i^Tx \simeq b_i$, $i = 1, \ldots, m$ con $\simeq$ pari a $\leq, \geq$ oppure $=$.

Un problema di PL consiste nel minimizzare (massimizzare) una funzione obiettivo lineare su un poliedro.

Tra le forme di poliedri più usate nella definizione di problemi di PL abbiamo le seguenti:

(a) $S = \{x \in R^n : \; Ax = b\}$

(b) $S = \{x \in R^n : \; Ax \geq b\}$

(c) $S = \{x \in R^n : \; Ax = b, \; x \geq 0\}$ (forma standard)

In R^2 sono poliedri, ad esempio, gli insiemi definiti come segue.

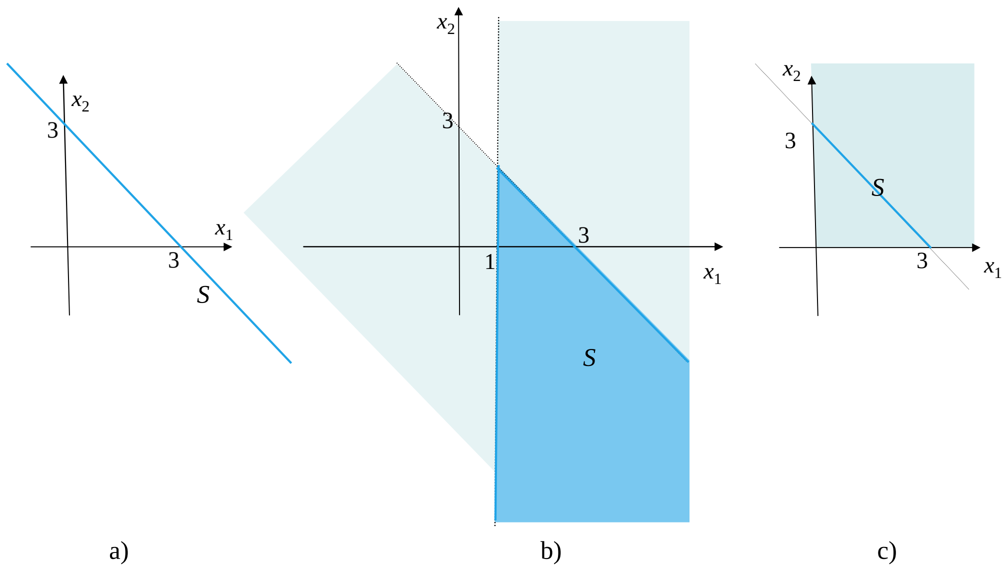

Figura 4.1: Esempi di poliedri in $I\!R^2$.

(a) $S = \{x \in R^2 : \; x_1 + x_2 = 3\}$ rappresentato in a) di Figura 4.1. La matrice $A = (1 \; 1)$ e $b = 3$.

(b) $S = \{x \in R^2 : \; x_1 + x_2 \leq 3, \; x_1 \geq 1\}$ rappresentato in b) di Figura 4.1.

La matrice $A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ e $b = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$

(c) (forma standard) $S = \{x \in R^2 : \; x_1 + x_2 = 3, \; x_1 \geq 0, \; x_2 \geq 0\}$ rappresentato in c) di Figura 4.1. La matrice $A = (1 \; 1)$ e $b = 3$.

In generale però è sempre possibile sempre riportarsi ad una delle forme (b) o (c) con semplici trasformazioni dei vincoli o delle variabili. È da notare che non è possibile ricondursi nel caso generale a problemi con soli vincoli di uguaglianza del tipo

$$\begin{array}{ll} \min & c^T x \\ & Ax = b. \end{array}$$

Tali problemi sono in effetti di scarsa rilevanza pratica in quanto è possibile dimostrare che se esiste una soluzione ammissibile e il problema non è illimitato allora *tutte* le soluzioni ammissibili sono ottime. In particolare vale il teorema 5.4.1.

Trasformazioni dei vincoli Osserviamo che si può sempre supporre che il secondo membro di un vincolo di uguaglianza o disuguaglianza sia non negativo. Infatti, se così non fosse, basta moltiplicare per (-1) (ovvero cambiare di segno) ad entrambi i membri del vincolo e, eventualmente cambiare il verso della disuguaglianza. Supponiamo quindi nel seguito, senza perdere di generalità che $b_j \geq 0$.

Da disuguaglianza a uguaglianza Con l'introduzione di opportune variabili aggiuntive non negative, ogni vincolo di disuguaglianza può sempre essere posto nella forma di vincolo di uguaglianza. Infatti un vincolo del tipo

$$a_j^T x \leq b_j, \tag{4.2}$$

può essere riscritto nella forma

$$a_j^T x + x_{n+1} = b_j, \qquad \text{con} \quad x_{n+1} \geq 0$$

ove la variabile x_{n+1} rappresenta la differenza, non negativa, tra il secondo e il primo membro della disuguaglianza (4.2), e viene detta *variabile di slack.* D'altra parte una disuguaglianza del tipo

$$a_j^T x \geq b_j, \tag{4.3}$$

può essere posta nella forma

$$a_j^T x - x_{n+1} = b_j, \qquad \text{con} \quad x_{n+1} \geq 0$$

ove la variabile x_{n+1} rappresenta la differenza, non negativa, tra il primo e il secondo membro della disuguaglianza (4.3), e viene detta *variabile di surplus.* Quando la distinzione non è necessaria, le variabili slack e di surplus vengono chiamate variabili ausiliarie.

Esempio 4.1.1 Il sistema di vincoli

$$\begin{array}{rcrcrcr} 3x_1 & & & + & 4x_3 & \leq & 5 \\ -5x_1 & + & x_2 & - & 6x_3 & = & -7 \\ 8x_1 & & & + & 9x_3 & \geq & 2 \end{array}$$

può essere riscritto nella forma

$$\begin{array}{rcrcrcrcrcr} 3x_1 & & & + & 4x_3 & + & x_4 & & & = & 5 \\ 5x_1 & - & x_2 & + & 6x_3 & & & & & = & 7 \\ 8x_1 & & & + & 9x_3 & & & - & x_5 & = & 2 \\ x_4 & \geq & 0 & & & & & & & & \\ x_5 & \geq & 0 & & & & & & & & \end{array}$$

avendo introdotto la variabile slack x_4 e la variabile di surplus x_5, entrambe non negative. □

Da uguaglianza a disuguaglianza È anche possibile trasformare vincoli di uguaglianza in vincoli di disuguaglianza. Considerato il vincolo $a_j^T x = b_j$ si pu]'o trasformare in due vincoli di disuguaglianza

$$\begin{array}{l} a_j^T x \leq b_j \\ -a_j^T x \leq -b_j \end{array}$$

Trasformazioni delle variabili Nei problemi di programmazione lineare si può sempre assumere che le variabili di decisione siano non negative, e cioè che per ogni i risulti $x_i \geq 0$. Infatti se per qualche i deve risultare $x_i \leq 0$, basta effettuare la sostituzione $x_i = -x_i'$, con $x_i' \geq 0$; se per qualche i la variabile x_i non è vincolata in segno, basta sostituire ad essa la differenza tra due variabili aggiuntive entrambe vincolate in segno, cioè basta porre $x_i = x_i^+ - x_i^-$, con $x_i^+ \geq 0$, e $x_i^- \geq 0$.

Esempio 4.1.2 Sia dato il vincolo

$$x_1 + 0.5x_2 \leq 3$$

con $x_2 \leq 0$ e con x_1 non vincolata.
Ponendo $x_1 = x_1 - x_3$ e $x_2 = -x_2$, equivale al vincolo

$$x_1 - 0.5x_2 - x_3 \leq 3$$

con $x_i \geq 0$, $i = 1, 2, 3$. □

4.2 Semplici esempi di problemi di programmazione lineare

Chiariremo meglio i concetti finora esposti con alcuni esempi. I problemi descritti hanno una funzione essenzialmente esemplificativa. In casi concreti, un problema di programmazione lineare può avere un numero di variabili di decisione e un numero di vincoli dell'ordine delle decine e centinaia di migliaia.

4.2.1 Un problema di produzione (allocazione di risorse)

Un colorificio produce due tipi di coloranti **C1** e **C2** utilizzando 3 preparati base in polvere **P1**, **P2**, **P3** che vengono sciolti in acqua. La differente concentrazione dei preparati base dà origine ai due diversi tipi di coloranti. Le quantità (in ettogrammi) di preparati base necessarie per produrre un litro di colorante di ciascuno dei due tipi è riportato nella seguente tabella

	C1	**C2**
P1	1	1
P2	1	2
P3	-	1

Ogni giorno la quantità di ciascuno dei preparati base (in ettogrammi) della quale il colorificio può disporre è la seguente

P1	**P2**	**P3**
750	1000	400

Il prezzo di vendita del colorante **C1** è di 7 Euro al litro, mentre il colorante **C2** viene venduto a 10 Euro al litro. Determinare la strategia ottimale di produzione giornaliera in modo da massimizzare i ricavi ottenuti dalla vendita dei due coloranti. **Formulazione.**
Si vuole costruire il modello di Programmazione Lineare che rappresenti il problema in analisi considerando le limitazioni date dalle produzioni effettivamente realizzabili.

È immediato associare le variabili di decisione ai quantitativi di coloranti prodotti. Siano, quindi, rispettivamente x_1 e x_2 i quantitativi (in litri) da produrre giornalmente dei due coloranti.

Nel formulare il modello di Programmazione Lineare si deve verificare che siano soddisfatte le ipotesi fondamentali:

- *Proporzionalità.*

  I consumi dei preparati base e i ricavi ottenibili sono proporzionali ai quantitativi di coloranti prodotti. Ad esempio, per produrre una quantità x_2 di colorante **C2** si consumano $2x_2$ ettogrammi di **P2** e dalla vendita di x_2 litri di **C2** si ricavano $10x_2$ migliaia di lire indipendentemente dalla quantità prodotta e venduta dell'altro tipo di colorante.

- *Additività.*

  I consumi dei preparati base e i ricavi rispettivamente associati alla produzione dei due coloranti sono additivi, nel senso che per produrre x_1 litri di colorante **C1** e x_2 di **C2** si consumano $x_1 + 2x_2$ ettogrammi di preparato di base **P2** e si ricavano $7x_1 + 10x_2$ migliaia di lire.

- *Continuità.*

  Ogni variabile introdotta nel modello può assumere tutti i valori reali nell'intervallo di ammissibilità.

– *Variabili.* Come già detto, prendiamo come variabili di decisione x_1 e x_2, rispettivamente i quantitativi (in litri) di colorante **C1** e **C2** da produrre giornalmente.
– *Funzione obiettivo.* È rappresentata dal profitto totale che per le ipotesi fatte è dato (in Euro) da $7x_1 + 10x_2$.
– *Vincoli.* Poiché il consumo di preparati base non può essere superiore alla disponibilità si deve avere

$$\begin{aligned} x_1 + x_2 &\leq 750 \\ x_1 + 2x_2 &\leq 1000 \\ x_2 &\leq 400. \end{aligned}$$

Inoltre si deve esplicitare il vincolo di non negatività sulle variabili

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

Quindi la formulazione finale è

$$\begin{array}{ll} \max & 7x_1 + 10x_2 \\ & x_1 + x_2 \leq 750 \\ & x_1 + 2x_2 \leq 1000 \\ & x_2 \leq 400 \\ & x_1 \geq 0, x_2 \geq 0. \end{array}$$

4.2.2 Un problema di miscelazione

Un'industria conserviera deve produrre succhi di frutta mescolando polpa di frutta e dolcificante ottenendo un prodotto finale che deve soddisfare alcuni requisiti riguardanti il contenuto di vitamina C, di sali minerali e di zucchero. La polpa di frutta e il dolcificante vengono acquistati al costo rispettivamente di Lire 400 e Lire 600 ogni ettogrammo. Inoltre dalle etichette si ricava che 100 grammi di polpa di frutta contengono 140 mg di vitamina C, 20 mg di sali minerali e 25 grammi di zucchero, mentre 100 grammi di dolcificante contengono 10 mg di sali minerali, 50 grammi di zucchero e non contengono vitamina C. I requisiti che il prodotto finale (cioè il succo di frutta pronto per la vendita) deve avere sono i seguenti: il succo di frutta deve contenere almeno 70 mg di vitamina C, almeno 30 mg di sali minerali e almeno 75 grammi di zucchero. Si devono determinare le quantità di polpa di frutta e di dolcificante da utilizzare nella produzione del succo di frutta in modo da minimizzare il costo complessivo dell'acquisto dei due componenti base.

Formulazione.
Si vuole costruire un modello di Programmazione Lineare che rappresenti il problema in analisi tenendo presente i requisiti di qualità richiesti. Si verifica facilmente che le ipotesi fondamentali di un modello di Programmazione Lineare sono soddisfatte.

– *Variabili.* È naturale associare la variabili di decisione alle quantità di polpa di frutta e di dolcificante da utilizzare per la produzione del succo di frutta. Quindi siano x_1 e x_2 rispettivamente le quantità espresse in ettogrammi di polpa di frutta e di dolcificante che devono essere utilizzate.

– *Funzione obiettivo.* È rappresentata dal costo complessivo dell'acquisto dei due componenti base e quindi è data da $400x_1 + 600x_2$. Questa espressione naturalmente deve essere minimizzata.

– *Vincoli.* Poiché un ettogrammo di polpa contiene 140 mg di vitamina C e il dolcificante non contiene vitamina C, il primo vincolo da considerare riguardante il contenuto di vitamina C del succo di frutta si può scrivere nella forma

$$140x_1 \geq 70.$$

Analogamente per rispettare il requisito sul contenuto di sali minerali del succo di frutta si dovrà imporre il vincolo

$$20x_1 + 10x_2 \geq 30.$$

Infine il vincolo sul contenuto di zucchero del succo di frutta si può esprimere nella forma

$$25x_1 + 50x_2 \geq 75.$$

Infine si deve esplicitare il vincolo di non negatività sulle variabili cioè

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

Quindi la formulazione finale è

$$\begin{array}{ll} \min & 400x_1 + 600x_2 \\ & 140x_1 \geq 70 \\ & 20x_1 + 10x_2 \geq 30 \\ & 25x_1 + 50x_2 \geq 75 \\ & x_1 \geq 0, x_2 \geq 0 \end{array}$$

4.2.3 Un problema di trasporto

Consideriamo un'industria che produce un bene di consumo in due stabilimenti di produzione, situati rispettivamente a Pomezia e a Caserta. La produzione viene prima immagazzinata in due depositi, situati uno a Roma e l'altro a Napoli. Quindi i prodotti vengono distribuiti alla rete di vendita al dettaglio. Per ogni unità di prodotto, il costo del trasporto dallo stabilimento al deposito è dato dalla Tabella 1:

Trasporto	**lire**
Pomezia-Roma	1000
Pomezia-Napoli	3000
Caserta-Napoli	500
Caserta-Roma	3500

Tab. 1
Costi di trasporto

La capacità produttiva dei due stabilimenti è limitata, per cui ogni settimana il bene in questione non può essere prodotto in più di 10000 unità nello stabilimento di Pomezia e in più di 8000 unità nello stabilimento di Caserta. Inoltre le statistiche di vendita informano che ogni settimana vengono vendute mediamente 11000 unità tramite il deposito di Roma e 4600 unità tramite il deposito di Napoli. L'industria vuole minimizzare il costo del trasporto della merce dagli stabilimenti ai depositi, assicurando che i depositi ricevano settimanalmente le quantità medie prima indicate.

Le variabili di decisione sono le quantità del bene di consumo trasportate settimanalmente, che possiamo associare alle variabili x_1, x_2, x_3, x_4 nel seguente modo:

quantità trasportata da Pomezia a Roma	x_1
quantità trasportata da Pomezia a Napoli	x_2
quantità trasportata da Caserta a Napoli	x_3
quantità trasportata da Caserta a Roma	x_4

La funzione obiettivo è il costo sostenuto settimanalmente per il trasporto:

$$z(x_1, x_2, x_3, x_4) = 1000x_1 + 3000x_2 + 500x_3 + 3500x_4.$$

Poichè i due stabilimenti hanno capacità produttiva limitata deve essere

$$\begin{array}{l} x_1 + x_2 \leq 10000 \\ x_3 + x_4 \leq 8000. \end{array}$$

Poichè si vuole garantire il rifornimento medio settimanale, deve essere

$$\begin{array}{l} x_1 + x_4 = 11000 \\ x_2 + x_3 = 4600. \end{array}$$

Infine evidentemente deve risultare $x_i \geq 0$, $i = 1, 2, 3, 4$, e quindi il problema di programmazione lineare per l'industria dell'esempio è il seguente:

$$\begin{array}{rcl} \min z & = & 1000x_1 + 3000x_2 + 500x_3 + 3500x_4 \\ & & x_1 + x_2 \leq 10000 \\ & & x_3 + x_4 \leq 8000 \\ & & x_1 + x_4 = 11000 \\ & & x_2 + x_3 = 4600 \\ & & x_i \geq 0, \quad i = 1, 2, 3, 4. \end{array}$$

4.2.4 Un problema di Yield Management ferroviario

Una compagnia ferroviaria vende i biglietti per il treno che effettua il percorso dalla città A (Roma) alla B (Bologna) effettuando una fermata intermedia F_1 (Firenze).

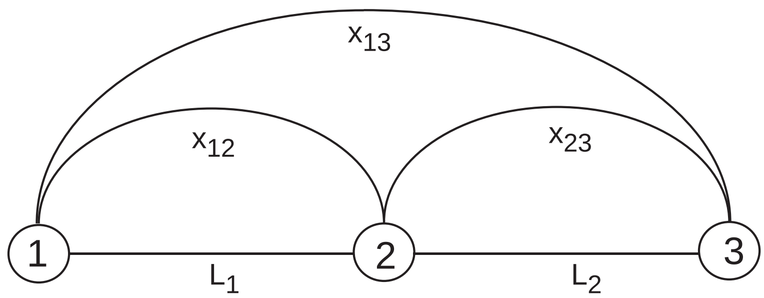

Figura 4.2: Un treno con percorso composto da due tratte L_1, L_2 e tre Origini-Destinazione (1,2), (2,3), (1,3)

Le tariffe sono fissate a priori e dipendono solo dalla distanza tra le origini-destinazioni. In particolare i prezzi in Euro (1^a classe) per per ciascuna possibile Origine-Destinazione (O-D) coperta dal treno sono riportati nella Tabella 4.2.4.

O-D	A-F_1	A-B	F_1-B
Tariffa	42,35	53,20	18,59

Tabella 4.1: Tariffe per le tre possibili O-D coperte dal treno

O-D	A-F_1	A-B	F_1-B
μ	420	355	335

Tabella 4.2: Domanda prevista per le tre possibili O-D coperte dal treno

Il numero di posti disponibili sul treno è pari a 700.

All'inizio del periodo di prenotazione la domanda per ciascuna origine-destinazione (O-D) è incerta, ma è nota una previsione sulla domanda il cui valor (medio) μ è riportato in Tabella 4.2.4.

Il problema consiste nel determinare quanti posti vendere su ciascuna Origine -Destinazione (O-D) coperta dal treno in modo da massimizzare il profitto.

Formulazione.
Si vuole costruire un modello di Programmazione Lineare che rappresenti il problema in analisi tenendo presente i requisiti richiesti. Si assume che la domanda effettiva per ciascuna origine-Destinazione sarà almeno pari al valor medio μ.

– *Variabili.* È naturale associare la variabili di decisione alle quantità di posti prenotabili su ciascuna Origine-Destinazione possibile. In questo caso, sono le tre variabili x_{12}, x_{13} e x_{23} che rappresentano rispettivamente i posti prenotabili sulla tratta A-F_1, A-B, F_1-B.

– *Funzione obiettivo.* È rappresentata dal profitto (atteso) relativo alla vendita e quindi è data da $42,35x_{12} + 53,20x_{13} + 18,59x_{23}$. Questa espressione naturalmente deve essere massimizzata.

– *Vincoli.* Osserviamo innanzitutto che, poiché per ipotesi si avranno almeno μ domande per ciascuna Origine-Destinazione, è necessario imporre i vincoli

$$\begin{aligned} x_{12} &\leq 420 \\ x_{13} &\leq 355 \\ x_{23} &\leq 335 \end{aligned}$$

in quanto non è sensato prenotare più posti della domanda effettiva.

Inoltre devono essere imposti i vincoli legato alla capacità di posti del treno. In particolare, nel tratto da A ad F_1 sono presenti sia i viaggiatori che si recano da A ad F_1 che quelli che vanno da A ad B. cosí nel tratto dal F_1 a B sono presenti sia i viaggiatori che vanno da A ad B che quelli che vanno da F_1 a B. Avremo quindi 2 vincoli

$$\begin{aligned} x_{12} + x_{13} &\leq 700 \\ x_{23} + x_{13} &\leq 700 \end{aligned}$$

Infine si deve esplicitare il vincolo di non negatività sulle variabili cioè $x_{12}, x_{23}, x_{13} \geq 0$.
In principio le variabili rappresentano dei posti, e quindi si dovrebbero esplicitare anche i vincoli di interezza. Questi vincoli possono essere omessi come sarà chiarito in un prossimo capitolo.

Quindi la formulazione finale è

$$\begin{array}{ll} \max & 42,35x_{12} + 53,20x_{13} + 18,59x_{23} \\ & x_{12} + x_{13} \leq 700 \\ & x_{23} + x_{13} \leq 700 \\ & 0 \leq x_{12} \leq 420 \\ & 0 \leq x_{13} \leq 355 \\ & 0 \leq x_{23} \leq 335 \\ & (x_{12}, x_{23}, x_{13} \text{ intere}) \end{array}$$

4.3 Interpretazione geometrica di un Problema di Programmazione Lineare

In questo paragrafo si vuole fornire una interpretazione geometrica di un problema di Programmazione Lineare. In particolare, quando un problema di Programmazione Lineare contiene solamente due variabili, si può rappresentare efficacemente il problema sul piano cartesiano e si può determinare una sua soluzione in maniera elementare con semplici deduzioni geometriche. Le situazioni che verranno presentate nel seguito vogliono rappresentare un punto di partenza intuitivo per la trattazione di problemi di Programmazione Lineare in n variabili; i risultati che verranno dedotti per via elementare nel caso bidimensionale trovano, infatti, una generalizzazione consistente nel caso di un generico problema di Programmazione Lineare.

A questo scopo verranno considerati due esempi di problemi di Programmazione Lineare già ottenuti come formulazione di un semplice problema di allocazione ottima di risorse (Esempio 4.2.1) e di un semplice problema di miscelazione (Esempio 4.2.2).

4.3.1 Rappresentazione di vincoli lineari

Preliminarmente, si richiama il fatto che sul piano cartesiano Ox_1x_2 l'equazione

$$a_1x_1 + a_2x_2 = c \tag{4.4}$$

rappresenta una retta che partiziona il piano in due semipiani. Ciascun semipiano è caratterizzato da punti $P(x_1, x_2)$ che soddisfano la disequazione $ax_1 + bx_2 \geq c$ oppure la disequazione $ax_1 + bx_2 \leq c$. Quindi ogni disequazione del tipo

$$ax_1 + bx_2 \geq c \qquad \text{oppure} \qquad ax_1 + bx_2 \leq c$$

individua univocamente un semipiano. Si riporta, ora, un semplice risultato geometrico che verrà utilizzato nel seguito.

Lemma 4.3.1 *Si considera una famiglia di rette parallele*

$$a_1x_1 + a_2x_2 = c \tag{4.5}$$

con $a_1, a_2 \in \mathbb{R}$ fissati e con $c \in \mathbb{R}$. Il vettore $a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ individua una direzione ortogonale alle rette della famiglia (4.5) ed è orientato dalla parte in cui sono le rette della famiglia ottenute per valori crescenti della c, cioè verso il semipiano in cui risulta $a_1x_1 + a_2x_2 \geq c$.

Dimostrazione. Non in programma Sia x un vettore di componenti x_1 e x_2. La famiglia di rette (4.5) può essere scritta in forma vettoriale

$$a^T x = c.$$

Siano ora $\bar{x}$ e $\bar{z}$ due punti appartenenti alla retta $a^T x = c$. Risulta quindi

$$a^T \bar{z} = c \qquad \text{e} \qquad a^T \bar{x} = c$$

e sottraendo membro a membro queste due uguaglianze si ottiene

$$a^T (\bar{z} - \bar{x}) = 0.$$

Quindi il vettore a è ortogonale al vettore $\bar{z} - \bar{x}$ che individua la direzione della famiglia di rette cioè il vettore a rappresenta una direzione ortogonale alla famiglia di rette (4.5).

Si consideri ora un punto $\bar{x} = (x_1, x_2)^T$ appartenente alla retta $a^T x = c$ e un punto $\bar{y} = (y_1, y_2)^T$ tale che $a^T y \geq c$. Si vuole dimostrare che il punto $\bar{y}$ appartiene al semipiano individuato dalla retta $a^T x = c$ verso il quale è orientato il vettore a. Infatti, per le ipotesi fatte, si ha

$$a^T \bar{y} \geq c \qquad \text{e} \qquad a^T \bar{x} = c$$

e sottraendo membro a membro queste due relazioni si ottiene

$$a^T \left(\bar{y} - \bar{x}\right) \geq 0$$

e questo significa che l'angolo θ che il vettore $\bar{y} - \bar{x}$ forma con il vettore a deve essere acuto e quindi il vettore a deve essere orientato verso il semipiano ove si trova il punto $\bar{y}$, cioè il semipiano individuato dalla diseguaglianza $a^T x \geq c$. (Figura 4.3) $\square$

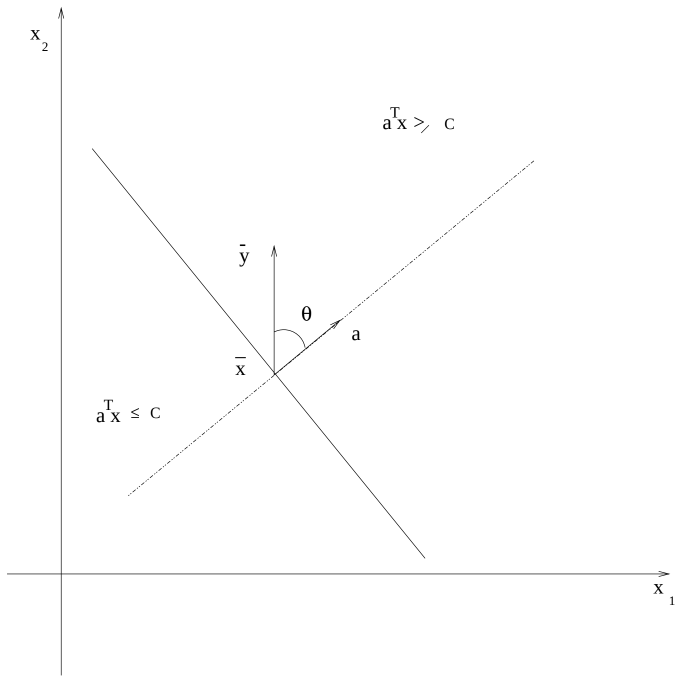

Figura 4.3: Interpretazione geometrica del Lemma 4.3.1

Come esempio del risultato riportato dal lemma appena dimostrato, si consideri la disuguaglianza lineare $3x_1 + x_2 \geq 6$. Il vettore $a = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$ individua una direzione ortogonale alla retta $3x_1 + x_2 = 6$ ed è orientato verso il semipiano individuato dalla disuguaglianza $3x_1 + x_2 \geq 6$ (Figura 4.4).

Nella pratica, per determinare quale dei due semipiani è individuato dalla disuguaglianza lineare $a_1x_1 + a_2x_2 \geq c$ si può procedere semplicemente in questo modo: dopo aver rappresentato la retta $a_1x_1 + a_2x_2 = c$ per individuare qual è il semipiano di interesse, si può scegliere un punto P del piano (l'origine degli assi è il piú semplice) e valutare l'espressione $a_1x_1 + a_2x_2$ in questo punto; se il valore così ottenuto è maggiore o uguale di c allora il semipiano individuato dalla disuguaglianza lineare $a_1x_1 + a_2x_2 \geq c$ è quello contenente il punto P; in caso contrario è quello opposto.

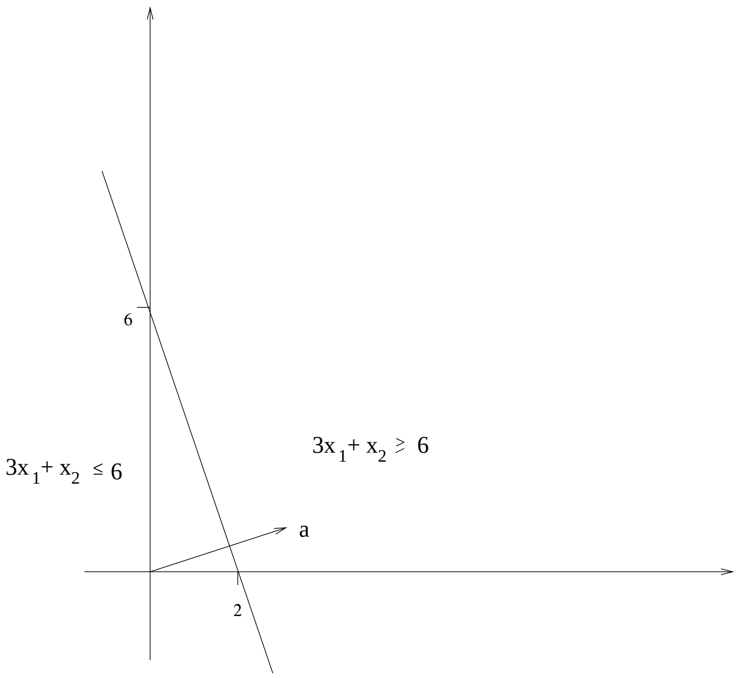

Figura 4.4: Rappresentazione del vincolo lineare $3x_1 + x_2 \geq 6$

4.3.2 Rappresentazione di funzioni obiettivo lineari

Quanto esposto nel paragrafo precedente è utile anche per esaminare la variazione di una funzione lineare che rappresenta la funzione obiettivo di un problema di Programmazione Lineare. In due variabili, la funzione obiettivo di un problema di Programazione Lineare è un'espressione del tipo $c_1x_1 + c_2x_2$ da massimizzare o da minimizzare. Per rappresentare questa funzione obiettivo su un piano cartesiano Ox_1x_2 si considera la famiglia di rette parallele

$$c_1x_1 + c_2x_2 = C \tag{4.6}$$

ottenuta al variare di C, che rappresentano le *curve di livello* della funzione $f(x_1, x_2) = c_1x_1 + c_2x_2$ che ovviamente in questo caso sono rette. Se il problema è di minimizzazione, si cercherà di ottenere un valore piú basso possibile per la C in corrispondenza di valori ammissibili per x_1 e x_2; viceversa, se il problema è di massimizzazione, si cercherà ottenere un valore piú alto possibile per la C. Sulla base di quanto esposto nel paragrafo precedente, valori superiori della C si determinano traslando le rette nel verso individuato dal vettore $\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$ che rappresenta, quindi, una *direzione di crescita* per la funzione $c_1x_1 + c_2x_2$. Ovviamente, la direzione opposta sarà una *direzione di decrescita.*

Quindi, geometricamente, un problema di massimizzazione consisterà nel considerare la traslazione nel verso della direzione di crescita della funzione obiettivo, mentre in un problema di minimizzazione si considera la traslazione nel verso opposto (Figura 4.5)

4.3.3 Esempi di risoluzione grafica

Esempio 4.3.2
Si consideri ora il problema di allocazione ottima di risorse del Paragrafo 4.2.1 che è rappresentato dal seguente problema di Programmazione Lineare:

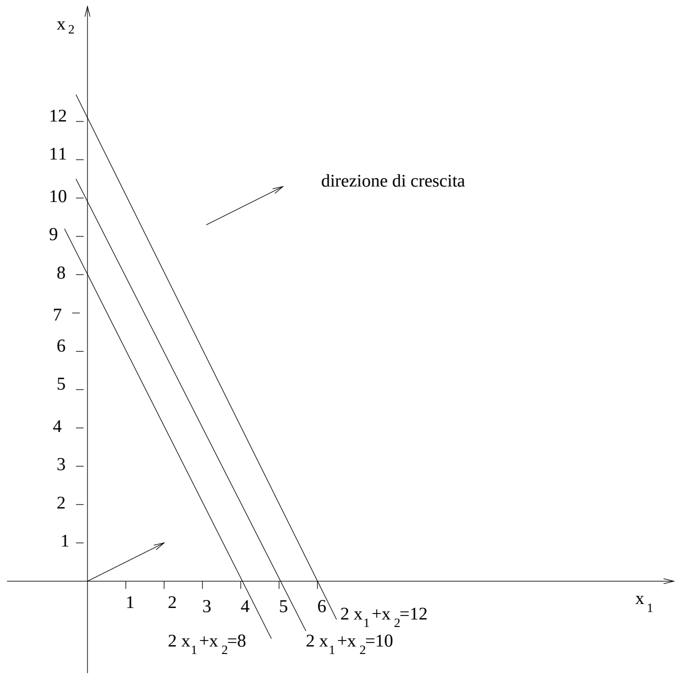

Figura 4.5: Rette di livello della funzione $2x_1 + x_2$

$$
\begin{array}{ll}
\max & 7x_1 + 10x_2 \\
 & x_1 + x_2 \leq 750 \\
 & x_1 + 2x_2 \leq 1000 \\
 & x_2 \leq 400 \\
 & x_1 \geq 0, x_2 \geq 0.
\end{array}
$$

Sul piano cartesiano Ox_1x_2 ciascun vincolo individua un semipiano. In particolare, in Figura 4.6 è evidenziato il semipiano individuato dal primo vincolo $x_1 + x_2 \leq 750$.

In Figura 4.7 è evidenziato il semipiano individuato dal secondo vincolo $x_1 + 2x_2 \leq 1000$.

Infine in Figura 4.8 è evidenziato il semipiano individuato dal terzo vincolo $x_2 \leq 400$.

Ovviamente i vincoli di non negatività delle variabili $x_1 \geq 0$ e $x_2 \geq 0$ rappresentano rispettivamente il semipiano delle ascisse non negative e il semipiano delle ordinate non negative.

L'insieme ammissibile del problema di Programmazione Lineare che stiamo esaminando è dato quindi dall'intersezione di tali semipiani e si può indicare con

$$S = \Big\{(x_1, x_2) \in \mathbb{R}^2 \;\;|\;\; x_1 + x_2 \leq 750, x_1 + 2x_2 \leq 1000, x_2 \leq 400, x_1 \geq 0, x_2 \geq 0\Big\}.$$

Tale regione di piano prende nome di *regione ammissibile*, è un insieme convesso ed è rappresentato in Figura 4.9.

Tutti i punti $P(x_1, x_2)$ appartenenti a questa regione sono punti dell'insieme ammissibile del problema e quindi tutti i punti di questa regione costituiscono soluzioni ammissibili del problema.

Si consideri ora la funzione obiettivo $7x_1 + 10x_2$ e si consideri la famiglia di rette

$$7x_1 + 10x_2 = C$$

ottenute al variare del parametro C. Esse costituiscono le curve di livello della funzione in due variabili $f(x_1, x_2) = 7x_1 + 10x_2$ che sono ovviamente delle rette come rappresentato in Figura 4.10.

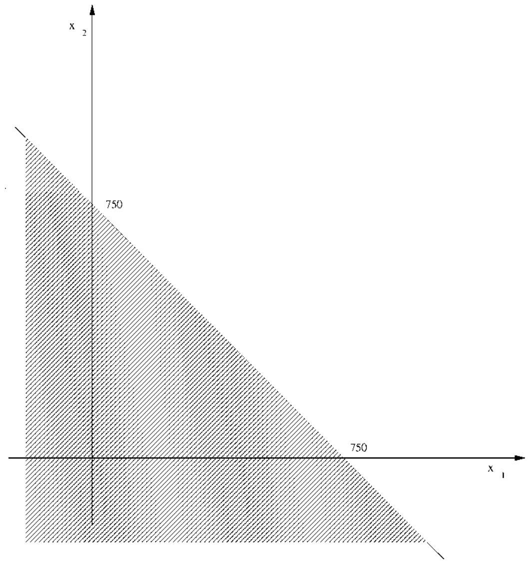

Figura 4.6: Semipiano individuato dal vincolo $x_1 + x_2 \leq 750$

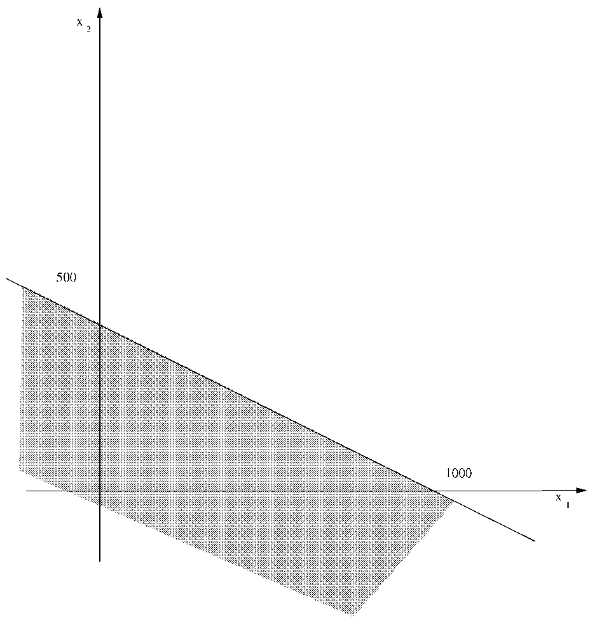

Figura 4.7: Semipiano individuato dal vincolo $x_1 + 2x_2 \leq 1000$

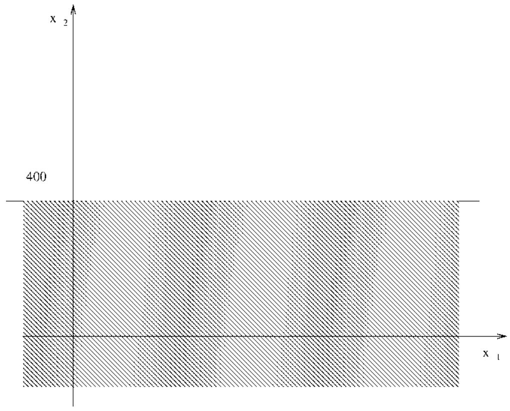

Figura 4.8: Semipiano individuato dal vincolo $x_2 \leq 400$

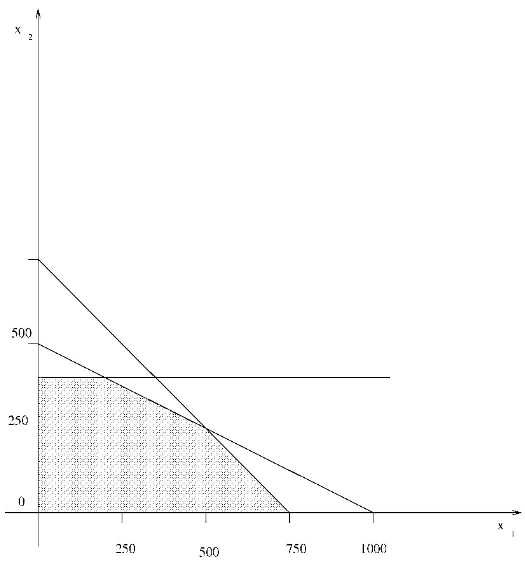

Figura 4.9: La regione ammissibile S

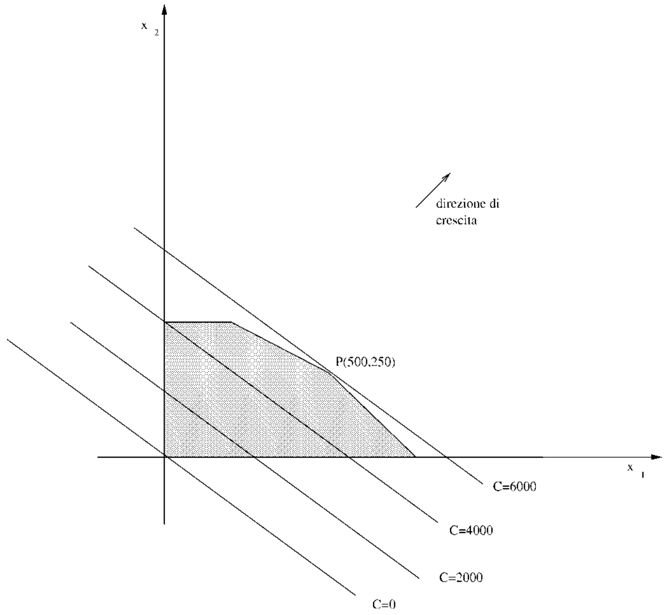

Figura 4.10: Curve di livello della funzione $f(x_1, x_2) = 7x_1 + 10x_2$ e punto di ottimo

In riferimento al problema di allocazione ottima rappresentato dal problema di Programmazione Lineare che stiamo esaminando, il parametro C rappresenta il profitto totale che deve essere massimizzato. Per $C = 0$ si ottiene la linea di livello passante per l'origine del piano Ox_1x_2. Ovviamente, scegliendo $x_1 = 0$ e $x_2 = 0$ (che è un punto ammissibile in quanto $(0, 0) \in S$) si ottiene il profitto totale nullo. All'aumentare del valore di tale profitto, cioè all'aumentare del valore della costante C, graficamente si ottengono rette parallele alla retta di livello passante per l'origine traslate nella direzione di crescita della funzione $7x_1 + 10x_2$ data dal vettore $\begin{pmatrix} 7 \\ 10 \end{pmatrix}$ (Figura 4.10). Poiché si desidera massimizzare la funzione obiettivo, si cercherà di raggiungere il valore piú alto per la C ottenuto scegliendo per x_1 e x_2 valori ammissibili cioè tali che $(x_1, x_2) \in S$. Osservando la rappresentazione grafica della regione ammissibile S si deduce che all'aumentare di C, le rette di livello della funzione obiettivo intersecano la regione ammissibile finché $C \leq 6000$. Tale valore è ottenuto per $x_1 = 500$ e $x_2 = 250$ e non esistono altri punti della regione ammissibile in cui la funzione obiettivo assume valori maggiori. Il valore 6000 è, quindi, il massimo valore che la funzione obiettivo può raggiungere soddisfacendo i vincoli. Tale soluzione ottima è raggiunta in corrispondenza del punto P di coordinate $(x_1, x_2) = (500, 250)$; tale punto non è un punto qualsiasi, ma costituisce quello che nella geometria piana viene detto vertice del poligono convesso che delimita la regione ammissibile. Il fatto che l'ottimo del problema è raggiunto in corrispondenza di un vertice della regione ammissibile non è casuale, ma come si vedrà in seguito, è una caratteristica fondamentale di un generico problema di Programmazione Lineare. Si osservi fin d'ora che la frontiera della regione ammissibile è definita dalle rette

$$x_1 + x_2 = 750, \quad x_1 + 2x_2 = 1000, \quad x_2 = 400, \quad x_1 = 0, \quad x_2 = 0$$

e che ogni intersezione di due di queste rette è un vertice della regione ammissibile; il numero di queste possibili intersezioni (non tutte necessariamente appartenenti alla regione ammissibile) è ovviamente pari al piú a 10. Si osservi, infine, che nel punto di ottimo sono attivi i vincoli $x_1 + x_2 \leq 750$ e $x_1 + 2x_2 \leq 1000$ mentre non è attivo il vincolo $x_2 \leq 400$.

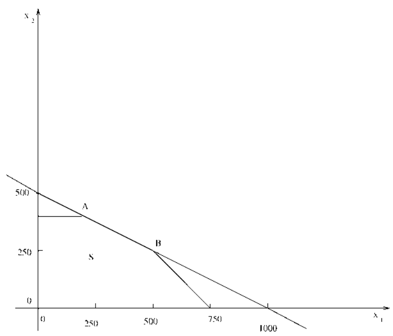

Figura 4.11: Esempio di soluzione non unica

Nel caso particolare che abbiamo esaminando, la soluzione ottima determinata è unica, ma in generale può accadere che le rette di livello della funzione obiettivo siano parallele ad un segmento del perimetro del poligono che delimita la regione ammissibile; in questo caso potrebbe accadere che esistano piú punti ammissibili in cui la funzione assume lo stesso valore ottimo e quindi la soluzione non sarebbe piú unica; nel problema in esame, ciò accadrebbe, ad esempio, se la funzione obiettivo fosse $cx_1 + 2cx_2$ con c costante reale positiva (Figura 4.11); infatti, tutti i punti del segmento $\overline{AB}$ rappresentano soluzioni ottime.

Tuttavia, anche in questo caso si può sempre trovare un vertice che costituisce una soluzione ottima.

Esempio 4.3.3

Consideriamo ora il problema di miscelazione del paragrafo 4.2.2 che è rappresentato dal seguente problema di Programmazione Lineare:

$$\begin{cases} \min(400x_1 + 600x_2) \\ 140x_1 \geq 70 \\ 20x_1 + 10x_2 \geq 30 \\ 25x_1 + 50x_2 \geq 75 \\ x_1 \geq 0, x_2 \geq 0 \end{cases}$$

Nelle figure che seguono rappresentati i vincoli lineari di questo problema; In particolare nella Figura 4.12 è evidenziato il semipiano individuato dal vincolo $140x_1 \geq 70$. Nella Figura 4.13 e nella Figura 4.14 sono evidenziati rispettivamente i semipiani individuati dai vincoli $20x_1 + 10x_2 \geq 30$ e $25x_1 + 50x_2 \geq 75$. Ovviamente i vincoli di non negatività delle variabili $x_1 \geq 0$ e $x_2 \geq 0$ rappresentano rispettivamente il semipiano delle ascisse non negative e il semipiano delle ordinate non negative.

L'insieme ammissibile del problema di Programmazione Lineare che stiamo esaminando è dato quindi dall'intersezione di tali semipiani e si può indicare con

$$\widetilde{S} = \Big\{(x_1, x_2) \in \mathbb{R}^2 \;\;|\;\; 140x_1 \geq 70, 20x_1 + 10x_2 \geq 30, 25x_1 + 50x_2 \geq 75, x_1 \geq 0, x_2 \geq 0\Big\}.$$

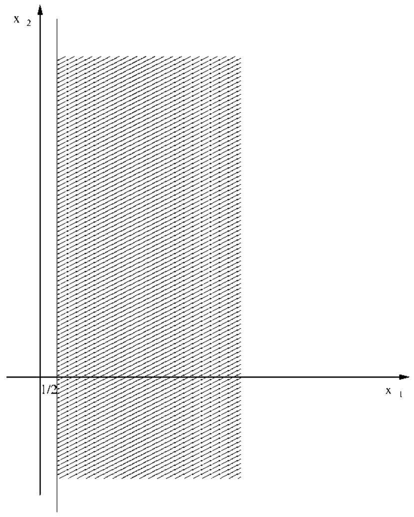

Figura 4.12: Semipiano individuato dal vincolo $140x_1 \geq 70$

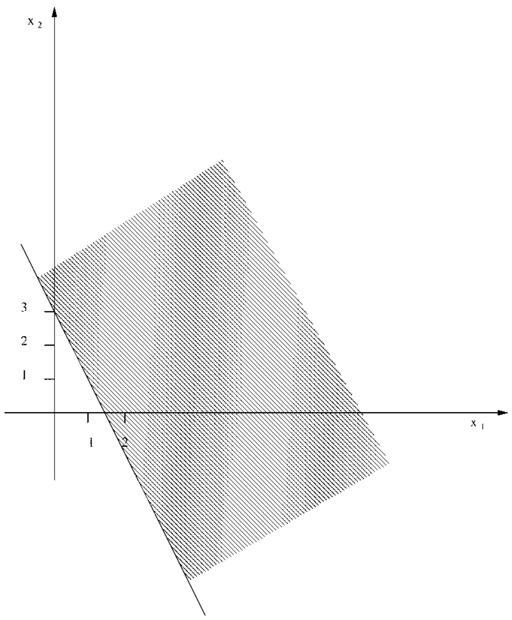

Figura 4.13: Semipiano individuato dal vincolo $20x_1 + 10x_2 \geq 30$

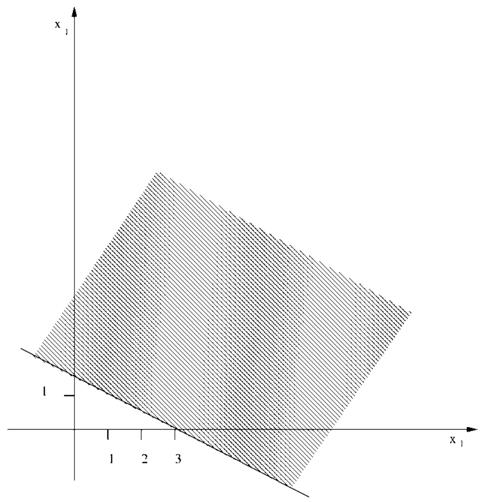

Figura 4.14: Semipiano individuato dal vincolo $25x_1 + 50x_2 \geq 75$

L'insieme $\widetilde{S}$ rappresenta la regione ammissibile del problema di Programmazione Lineare in esame ed è rappresentata in Figura 4.15.

Tutti i punti $P(x_1, x_2)$ appartenenti a questa regione sono punti dell'insieme ammissibile del problema e quindi tutti i punti di questa regione costituiscono soluzioni ammissibili del problema. Si osservi che, a differenza della regione ammissibile del problema considerato nell'esempio precedente, la regione ammissibile $\widetilde{S}$ è illimitata.

Ora, tracciando le curve di livello della funzione obiettivo $400x_1 + 600x_2$ si ottiene la famiglia di rette

$$400x_1 + 600x_2 = C.$$

Trattandosi di un problema di minimizzazione si vuole determinare il valore piú basso di C ottenuto scegliendo per x_1 e x_2 valori ammissibili cioè tali che $(x_1, x_2) \in \widetilde{S}$. Osservando la rappresentazione grafica della regione ammissibile $\widetilde{S}$ e osservando che la direzione di decrescita è quella opposta al vettore $\begin{pmatrix} 400 \\ 600 \end{pmatrix}$, si deduce che al diminuire di C, le rette di livello della funzione obiettivo intersecano la regione ammissibile finché $C \geq 1000$ (Figura 4.16)

Tale valore è ottenuto per $x_1 = 1$ e $x_2 = 1$ e non esistono altri punti della regione ammissibile in cui la funzione obiettivo assume valori minori. Il valore 1000 è, quindi, il minimo valore che la funzione obiettivo può raggiungere soddisfacendo i vincoli. Tale soluzione ottima è raggiunta in corrispondenza del punto P di coordinate $(x_1, x_2) = (1, 1)$; si osservi che anche in questo caso tale punto è un punto particolare della regione ammissibile. Si osservi, infine che in questo punto sono attivi i vincoli $20x_1 + 10x_2 \geq 30$ e $25x_1 + 50x_2 \geq 75$ mentre risulta non attivo il vincolo $140x_1 \geq 70$.

Abbiamo esaminato due esempi di interpretazione geometrica e soluzione grafica di problemi di Programmazione Lineare in due variabili. In entrambe i problemi è stato possibile determinare una soluzione ottima. Tuttavia è facile dedurre, sempre per via geometrica, che un problema di Programmazione Lineare

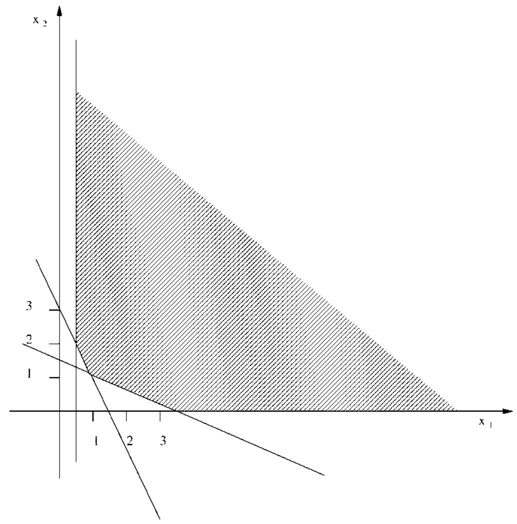

Figura 4.15: La regione ammissibile $\widetilde{S}$

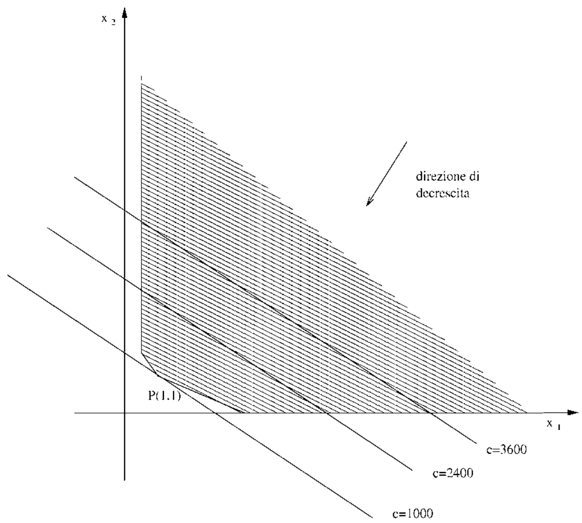

Figura 4.16: Curve di livello della funzione $400x_1 + 600x_2$ e punto di ottimo

può non ammettere soluzione ottima. Ad esempio, se nell'Esempio 4.3.2 sostituiamo il vincolo $x_2 \leq 400$ con il vincolo $x_2 \geq 1000$, la regione ammissibile sarebbe vuota nel senso che non esisterebbe nessun punto del piano che soddisfa tutti i vincoli. In questo caso il problema risulterebbe inammissibile e questo indipendentemente dalla funzione obiettivo e dal fatto che il problema è in forma di minimizzazione o massimizzazione.

Un altro esempio di problema di Programmazione Lineare che non ammette soluzione ottima si può ottenere considerando il problema dell'Esempio 4.3.3 e supponendo che la funzione obiettivo debba essere massimizzata anziché minimizzata. In questo caso nella regione ammissibile (che è illimitata) la funzione obiettivo può assumere valori arbitrariamente grandi cioè tali che comunque scelto un valore $M > 0$ esiste un punto in cui la funzione obiettivo assume valore maggiore di M; questo significa che il problema è illimitato superiormente e quindi non può esistere una soluzione ottima.

Sulla base di queste considerazioni sulla geometria di un problema di Programmazione Lineare in due variabili si può intuire che le situazioni che si possono verificare sono le seguenti:

- *il problema ammette soluzione ottima* (che può essere o non essere unica) in un vertice del poligono convesso che delimita la regione ammissibile;
- *il problema non ammette soluzione ottima* perché
  - la regione ammissibile è vuota
  - la regione ammissibile è illimitata e la funzione obiettivo è illimitata superiormente (se il problema è di massimizzazione) o illimitata inferiormente (se il problema è di minimizzazione).

Quindi se si suppone che esiste un punto ammissibile, cioè che la regione ammissibile sia non vuota, allora sembrerebbe di poter dedurre che o il problema di Programmazione Lineare ammette soluzione ottima in un vertice del poligono convesso che delimita la regione ammissibile oppure è illimitato.

Questi asserti, ora semplicemente dedotti intuitivamente per via geometrica, hanno in effetti una validità generale e verranno enunciati e dimostrati in maniera rigorosa nel capitolo 5. Come ultima considerazione intuitiva si vuole citare la possibilità che la regione ammissibile sia costituita da una striscia di piano, cioè dalla porzione di piano compresa tra due rette parallele (Figura 4.17). In questo caso non esistono vertici per la regione ammissibile e il problema risulta illimitato ad eccezione del caso particolare in cui le rette di livello della funzione obiettivo sono parallele alle rette che delimitano la striscia di piano; in questo caso si hanno infinite soluzioni.

La non esistenza di vertici in questo caso si intuisce essere legata al fatto che la regione ammissibile costituita da una striscia di piano *contiene rette*. Infatti nei casi delle regioni ammissibili S e $\widetilde{S}$ dei problemi di Programmazione Lineare dell'Esempio 4.3.2 e dell'Esempio 4.3.3 non esistono rette contenute in S o in $\widetilde{S}$. Anche la regione illimitata $\widetilde{S}$ può contenere semirette, ma non rette.

Ad eccezione del caso particolare della regione ammissibile rappresentata da una striscia di piano, il caso in cui è possibile garantire l'esistenza di almeno una soluzione ottima è quello della regione ammissibile che non contiene nemmeno semirette (che è il caso della regione S dell'Esempio 4.3.2). Si tratta di applicare il teorema di Weierstrass sull'esistenza di un minimo (e un massimo) assoluto di una funzione continua su un insieme chiuso e limitato.

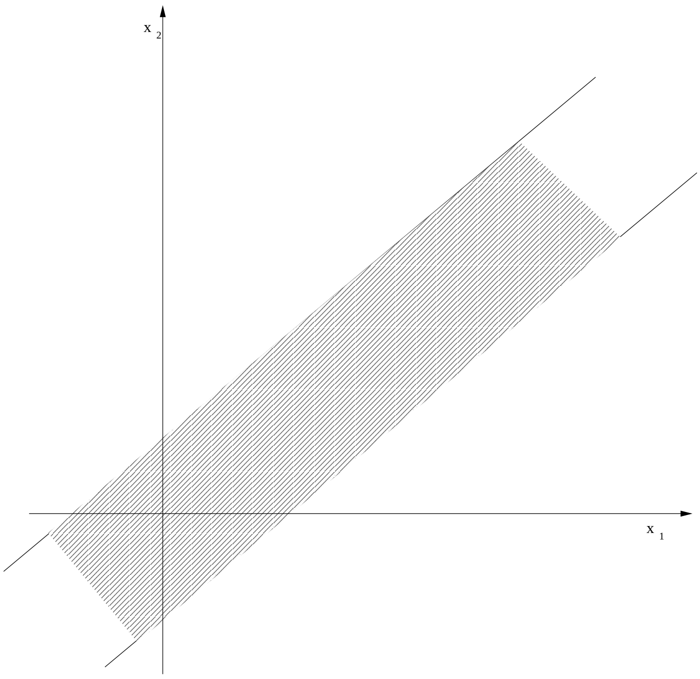

Figura 4.17: Regione ammissibile costituita da una striscia di piano

Capitolo 5

Teoria della Programmazione Lineare

In questo capitolo saranno approfonditi, formalizzati ed generalizzati i concetti espressi a titolo esemplificativo nel capitolo precedente.
Abbiamo già visto che l'insieme ammissibile di un problema di programmazione lineare è un poliedro (convesso). Ci proponiamo ora di studiare alcune proprietà dei poliedri che consentano di caratterizzare le soluzioni ottime di un problema di programmazione Lineare.

5.1 Direzione ammissibile

Introduciamo il concetto di direzione ammissibile. Intuitivamente una direzione è "ammissibile" in un punto $\bar{x} \in S$ se è possibile muoversi lungo tale direzione mantenendosi in S. Più formalmente possiamo dare la seguente definizione riferita ad un generico insieme $S \subseteq \mathbb{R}^n$ (non necessariamente un poliedro).

Definizione 5.1.1 (Direzione ammissibile) *Sia S un sottoinsieme di R^n e $\bar{x} \in S$. Si dice che un vettore $d \in R^n$, $d \neq 0$ è una direzione ammissibile per S in $\bar{x}$ se esiste $t_{\max} > 0$ tale che*

$$\bar{x} + td \in S, \quad \textit{per ogni} \quad t \in [0, t_{\max}\,].$$

Esempio 5.1.2 Sia dato il poliedro

$$S = \{x \in R^2 : \; x_1 + x_2 \leq 3, \; x_1 \geq 1, \; x_2 \geq 0\}$$

rappresentato in Figura 5.1 e consideriamo il punto ammissibile $\bar{x} = (1, \; 1)$ (indicato con un puntino rosso in figura).

È facile convincersi che la direzione $d = (1, \; 0)$ è ammissibile in $\bar{x}$. Risulta infatti

$$\bar{x} + td = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + t \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1+t \\ 1 \end{pmatrix}$$

Sostituendo nei vincoli si ha:

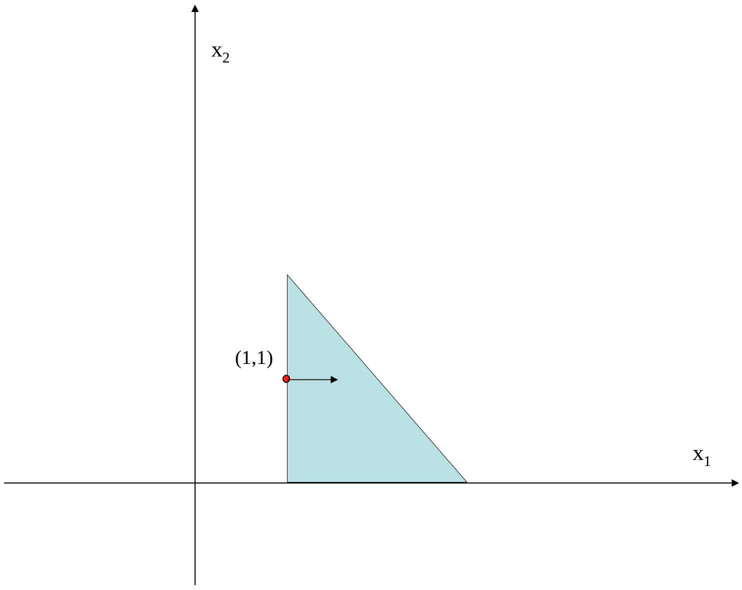

Figura 5.1: Poliedro Esempio 5.1.2.

1. $x_1 + x_2 = (1 + t) + 1 = 2 + t$ che risulta ≤ 3 per valori di $t \leq 1$.
2. $x_1 = 1 + t$ che risulta ≥ 1 per ogni $t \geq 0$;
3. $x_2 = 1 > 0$.

Quindi per ogni $t \in [0, 1]$ ($t_{\max} = 1$) risulta $\bar{x} + td \in S$.

È altrettanto facile convincersi che la direzione opposta $-d = (-1,\ 0)$ non è ammissibile in $\bar{x}$. Risulta infatti

$$y = \bar{x} + t(-d) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 - t \\ 1 \end{pmatrix}$$

Sostituendo nei vincoli si ha:

1. $x_1 + x_2 = (1 - t) + 1 = 2 - t$ che risulta ≤ 3 per valori di $t \geq -1$.
2. $x_1 = 1 - t$ che risulta ≥ 1 solo per $t \leq 0$;
3. $x_2 = 1 > 0$ per ogni t.

Quindi NON esiste un $t_{\max} > 0$ per cui il punto y appartiene ad S. La direzione non è dunque ammissibile. $\square$

Il concetto di direzione ammissibile è riferito ad un generico insieme S non necessariamente poliedrale né con particolari proprietà. Nel Capitolo 6 vedremo l'utilizzo di questo concetto per generici insiemi. In questo capitolo ci occupiamo di caratterizzare le direzione ammissibili nel caso in cui l'insieme S sia un poliedro.

5.2 Direzioni ammissibili di un poliedro

Consideriamo ora il caso di poliedri.

In questo caso, per caratterizzare le direzioni ammissibili è necessario introdurre ulteriori definizione. A tal scopo analizziamo nuovamente l'esempio 5.1.2.

Esempio 5.2.1 Consideriamo nuovamente il poliedro

$$S = \{x \in R^2 : \ x_1 + x_2 \le 3, \ x_1 \ge 1, \ x - 2 \ge 0\}$$

in Figura 5.1 e il punto ammissibile $\bar{x} = (1, \ 1)$.

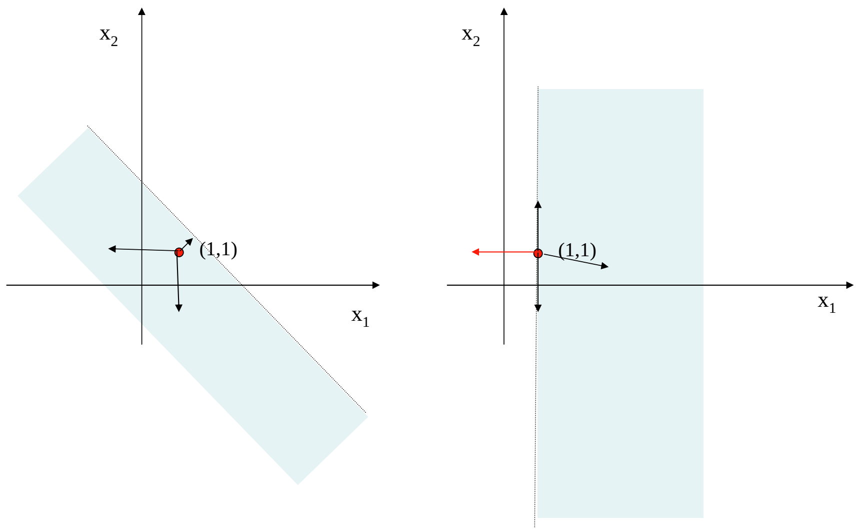

Figura 5.2: Semispazi individuati rispettivamente dai vincoli $x_1 + x_2 \le 3$ e $x_1 \ge 1$ dell'Esempio 5.2.1.

In Figura 5.2 sono rappresentati i due semispazi definiti singolarmente dai due vincoli $x_1 + x_2 \le 3$ e $x_1 \ge 1$ che definiscono il poliedro S. Se valutiamo i vincoli in $\bar{x} = (1, \ 1)^T$ (in rosso nella figura 5.2) abbiamo:

$$\begin{aligned} &\bar{x}_1 + \bar{x}_2 < 3, \\ &\bar{x}_1 = 1 \\ &\bar{x}_2 > 0 \end{aligned}$$

ovvero solo il vincolo $x_1 \le 1$ è soddisfatto all'uguaglianza.

Dalla figura 5.2 è evidente che nel punto $\bar{x}$ ogni direzione $d \in I\!R^n$ è ammissibile rispetto al vincolo $x_1 + x_2 \le 3$ a patto di scegliere adeguatamente il valore dello spostamento t. Analogamente rispetto al vincolo $x_2 \ge 0$ (non rappresentato in figura 5.2) che individua il primo e secondo quadrante, nel punto $\bar{x}$ ogni direzione è ammissibile. Se invece consideriamo il semispazio individuato dal vincolo $x_1 \ge 1$ è evidente che non tutte le direzioni sono ammissibili. Ad esempio la direzione $d = (-1 \ 0)^T$ (considerata nell'esempio 5.1.2 e disegnata in rosso in figura 5.2) non consente di mantenere l'ammissibilità del vincolo per nessun valore di $t > 0$.

Da queste semplici osservazioni appare evidente che in un punto ammissibile, le direzioni ammissibili sono "determinate" da quei vincoli che sono soddisfatti all'uguaglianza. Mentre il valore di $t_{\max}$ dipende dal valore degli altri vincoli.
Introduciamo allora la seguente definizione:

Definizione 5.2.2 (Vincoli attivi) *Sia $S = \{x \in R^n : \ Ax \geq b\}$ e sia $\bar{x} \in S$. Se $\bar{x}$ soddisfa il $a_i^T \bar{x} = b_i$ si dice che il vincolo i-esimo è* attivo *in $\bar{x}$.*
Dato $\bar{x} \in S$ si indica con $I(\bar{x})$ l'insieme degli indici di tutti i vincoli attivi n $\bar{x}$, ovvero:

$$I(\bar{x}) = \{i \in \{1, \ldots, m\} : \ a_i^T \bar{x} = b_i\}$$

Nell'esempio 5.1.2, è attivo il secondo vincolo e cioè $I(\bar{x}) = \{2\}$.

La definizione 5.2.2 di vincolo attivo non è legata alla particolare forma del poliedro, ma può essere estesa al caso generale. In particolare, notiamo che i vincoli di uguaglianza sono, per definizione, sempre attivi in un punto ammissibile. Possiamo quindi affermare

Dato un poliedro definito come $S = \{x \in I\!R^n : \ Ax = b\}$, risulta $I(\bar{x}) = \{1, \ldots, m\}$.

Dato un poliedro in forma standard $S = \{x \in I\!R^n : \ Ax = b, \ x \geq 0\}$, i vincoli attivi sono quelli di uguaglianza più gli eventuali componenti di $\bar{x}_j = 0$, ovvero risulta $I(\bar{x}) = \{1, \ldots, m\} \cup \{j \in \{1, \ldots, n\} : \ \bar{x}_j = 0\}$.

Possiamo ora caratterizzare le direzioni ammissibili di un poliedro.

Teorema 5.2.3 *Dato il poliedro $S = \{x \in R^n : \ Ax \geq b\}$. Sia $\bar{x}$ un punto ammissibile e sia $I(\bar{x}) = \{i \in \{1, \ldots, m\} : \ a_i^T \bar{x} = b_i\}$ l'insieme degli indici attivi in $\bar{x}$. Una direzione d è ammissibile in $\bar{x}$ se e solo se risulta*

$$a_i^T d \geq 0, \qquad \textit{per ogni} \ \ i \in I(\bar{x}). \tag{5.1}$$

Inoltre per ogni direzione ammissibile d, il punto $\bar{x} + td$ è ammissibile per ogni valore di t che soddisfa la condizione

$$0 < t \leq t_{\max} = \min_{j \notin I(\bar{x}): \ a_j^T d < 0} \frac{a_j^T \bar{x} - b_j}{|a_j^T d|} \tag{5.2}$$

dove si intende che, se l'insieme $\{j \notin I(\bar{x}) : \ a_j^T d < 0\}$ su cui si effettua il minimo è vuoto, il valore $t_{\max} = \infty$.

Dimostrazione. Sia $\bar{x}$ tale che $a_i^T \bar{x} \geq b_i$ per ogni $i = 1, \ldots, m$. Definiamo un punto $x = \bar{x} + td$ con $t > 0$.
Verifichiamo che le direzione ammissibili sono tutte e sole quelle per cui vale la (5.1). Si tratta quindi di verificare le condizioni su d per cui risulta

$$a_i^T (\bar{x} + td) = a_i^T \bar{x} + t a_i^T d \geq b_i \qquad \text{per ogni } t > 0.$$

Per ogni $i \in I(\bar{x})$, vale $a_i^T \bar{x} = b_i$ da cui $a_i^T \bar{x} + t a_i^T d = b_i + t a_i^T d$. Quindi si ottiene per $i \in I(\bar{x})$:

$$a_i^T \bar{x} + t a_i^T d \geq b_i \ \ \text{per ogni } t > 0 \qquad \text{se e solo se} \quad a_i^T d \geq 0.$$

Per ogni $j \notin I(\bar{x})$, vale $a_j^T \bar{x} > b_i$ da cui $a_j^T \bar{x} + t a_j^T d > b_j + t a_j^T d$. Abbiamo due possibili casi: 1.) $a_j^T d \geq 0$ oppure 2.) $a_j^T d < 0$.
Se $a_j^T d \geq 0$, allora per ogni $t > 0$ risulta $b_j + t a_j^T d \geq b_j$ e il punto è sempre ammissibile rispetto al vincolo $j-$esimo. Se invece $a_j^T d < 0$, è necessario scegliere t abbastanza piccolo in modo che $a_j^T \bar{x} - t|a_j^T d|$ continui ad essere $\geq b_j$ per $t > 0$. In ogni caso (1. o 2.) la direzione risulta ammissibile.

Verifichiamo che il valore di $t_{\max}$ deve soddisfare la (5.2).
Sia quindi data d ammissibile. Dalle considerazioni precedenti si ha che per ogni $i \in I(\bar{x})$, vale $a_i^T(\bar{x}+td) \geq b_i$ per ogni $t > 0$.
Consideriamo allora gli indici $j \notin I(\bar{x})$. Abbiamo già visto che nel caso $a_j^T d \geq 0$, possiamo scrivere per ogni $t > 0$ $a_j^T(\bar{x} + td) = a_j^T \bar{x} + t a_j^T d \geq a_j^T \bar{x} > b_j$ per ogni $t > 0$.
Consideriamo allora il caso $a_j^T d < 0$ in cui sappiamo che t non può assumere tutti i valori positivi. Dobbiamo trovare il valore di t che soddisfa la condizione di ammissibilità espressa da $a_j^T(\bar{x} + td) = a_j^T \bar{x} + t a_j^T d = a_j^T \bar{x} - t|a_j^T d| \geq b_j$. Risolvendo la condizione $a_i^T \bar{x} - t|a_j^T d| \geq b_j$ si ottiene che

$$t \leq \frac{a_j^T \bar{x} - b_j}{|a_i^T d|}.$$

Poiché questo deve valere per ogni $j \notin I(\bar{x})$ e tale che $a_j^T d < 0$, si ottiene la limitazione data dalla (5.2).
□

Illustriamo con un esempio.

Esempio 5.2.4 Sia dato il poliedro

$$\begin{aligned} &3x_1 - 2x_2 \geq -30 \\ &2x_1 - x_2 \geq -12 \\ &x_1 \geq 0 \; x_2 \geq 0 \end{aligned}$$

ed il punto ammissibile $\bar{x} = (0, \ 12)^T$. Risulta

$$\begin{aligned} &3\bar{x}_1 - 2\bar{x}_2 = -12 > -30 \\ &2\bar{x}_1 - \bar{x}_2 = -12 \\ &\bar{x}_1 = 0 \; \bar{x}_2 = 12 > 0 \end{aligned}$$

Quindi $I(\bar{x}) = \{2, 3\}$ e le direzioni ammissibili $d = (d_1, \ d_2)$ devono soddisfare il sistema

$$\begin{aligned} &2d_1 - d_2 \geq 0 \\ &d_1 \geq 0 \end{aligned}$$

Il valore di $t_{\max}$ si ottiene considerando i vincoli non attivi e dipende ovviamente dalla particolare direzione ammissibile considerata. Sia, ad esempio $\bar{d} = (1, \ 2)^T$. Risulta $a_1^T \bar{d} = 3 - 4 < 0$ e $a_4^T \bar{d} = 1 > 0$. Quindi il valore di $t_{\max}$ è individuato dal primo vincolo e si ha in particolare

$$t_{\max} = \frac{a_1^T \bar{x} - b_1}{|a_1^T d|} = \frac{3\bar{x}_1 - 2\bar{x}_2 + 30}{3\bar{d}_1 - 2\bar{d}_2} = \frac{6}{1} = 6.$$

In figura 5.3 è rappresentato il poliedro, il punto $\bar{x}$, l'insieme delle direzioni ammissibili (insieme compreso tra le due frecce blu tratteggiate), i punto ottenibili spostandosi da $\bar{x}$ lungo le direzioni ammissibili (regione compresa tra le due frecce blu a tratto intero) e il valore del passo $t_{\max}$ relativo alla direzione ammissibile $\bar{d} = (1, \ 2)^T$. Il punto $\bar{x} + t_{\max}\bar{d} = (6, \ 24)^T$ è indicato in rosso in figura 5.3. □

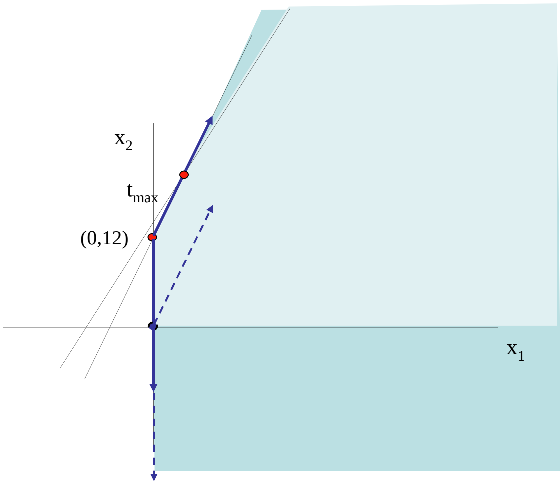

Figura 5.3: Poliedro Esempio 5.2.4.

Corollario 5.2.5 *Dato il poliedro $S = \{x \in I\!R^n : \ Ax \geq b\}$. Sia $\bar{x}$ un punto ammissibile e sia $I(\bar{x}) = \{i \in \{1, \ldots, m\} : \ a_i^T \bar{x} = b_i\}$ l'insieme degli indici attivi in $\bar{x}$ e supponiamo che esista un vettore d non nullo tale che*

$$a_i^T d = 0, \qquad \textit{per ogni} \ \ i \in I(\bar{x}).$$

Allora le direzioni d e $-d$ sono ammissibili in $\bar{x}$ ed esistono valori $t^+_{\max} > 0$ e $t^-_{\max} > 0$ tali che

$$\begin{array}{l} \bar{x} + td \in S \ \textit{per ogni} \ t \in (0, t^+_{\max}) \\ \bar{x} - td \in S \ \textit{per ogni} \ t \in (0, t^-_{\max}) \end{array}$$

con

$$t^+_{\max} = \min_{j \notin I(\bar{x}):\ a_j^T d < 0} \frac{a_j^T \bar{x} - b_j}{|a_j^T d|} \qquad t^-_{\max} = \min_{j \notin I(\bar{x}):\ a_j^T d > 0} \frac{a_j^T \bar{x} - b_j}{a_j^T d} \tag{5.3}$$

Si intende che se l'insieme degli indici su cui viene calcolato il minimo è vuoto, il corrispondente valore $t^+_{\max}$ e/o $t^-_{\max}$ vale ∞.

Dimostrazione. La dimostrazione segue facilmente dal Teorema 5.2.3, applicandolo separatamente alla due direzioni d e $-d$. □

Naturalmente è possibile caratterizzare anche le direzioni ammissibili nel caso di poliedri in forma standard $S = \{x \in I\!R^n : \ Ax = b, \ x \geq 0\}$.

Preventivamente consideriamo il caso di soli vincoli di uguaglianza $Ax = b$. Sia $\bar{x}$ un punto ammissibile, ovvero tale che $A\bar{x} = b$. La definizione di direzione ammissibile d richiede che $A(\bar{x} + td) = b$ per t sufficientemente piccolo. Sviluppando si ottiene

$$A\bar{x} + tAd = b + tAd = b \quad \Longleftrightarrow \quad Ad = 0 \quad \text{ per ogni } t.$$

Si osservi che in questo caso, t può assumere qualunque valore non nullo.

Dato un poliedro definito da un sistema di equazioni lineari del tipo $Ax = b$, e un punto ammissibile $\bar{x}$. Una direzione d è ammissibile se e solo se risulta

$$Ad = 0.$$

Inoltre ogni punto del tipo $\bar{x} + td$ è ammissibile per ogni valore di $t \neq 0$.

Riportiamo senza dimostrazione il seguente risultato relativo a poliedri in forma standard.

Teorema 5.2.6 (Direzioni ammissibili di un poliedro in forma standard) *Sia dato un poliedro del tipo $\{x \in \mathbb{R}^n : \ Ax = b, \ x \geq 0\}$ e sia $\bar{x}$ un punto ammissibile e sia $J(\bar{x}) = \{j \in \{1, \dots, n\} : \ \bar{x}_j = 0\}$ l'insieme degli indici attivi in $\bar{x}$. Una direzione d è ammissibile in $\bar{x}$ se e solo se risulta*

$$\begin{aligned} &Ad = 0, \\ &d_j \geq 0 \quad \text{per ogni} \ \ j \in J(\bar{x}). \end{aligned} \tag{5.4}$$

Inoltre per ogni direzione ammissibile d, un punto $\bar{x} + td$ è ammissibile per ogni valore di t che soddisfa la condizione

$$0 < t \leq t_{\max} = \min_{j: \ \bar{x}_j \neq 0: \ d_j < 0} \frac{\bar{x}_j}{|d_j|}$$

dove si intende che se l'insieme su cui si effettua il minimo $\{j : \ \bar{x}_j \neq 0 : \ d_j < 0\}$ è vuoto, il valore $t_{\max} = \infty$.

5.3 Caratterizzazione dei vertici di un poliedro

Passiamo ora a caratterizzare in modo algebrico i vertici di un poliedro in forma $Ax \geq b$.

Teorema 5.3.1 (Vertici di un poliedro) *Sia dato un poliedro $S = \{x \in R^n : \ Ax \geq b\}$. Un punto $\bar{x} \in S$ è un vertice se e solo se esistono n righe a_i^T della matrice A corrispondenti ai vincoli attivi in $\bar{x}$ linearmente indipendenti, cioè se e solo se risulta*

$$\textit{rango}\{a_i \ i \in I(\bar{x})\} = n.$$

Dimostrazione. Dimostriamo la parte necessaria, ovvero che se $\bar{x}$ è un vertice, allora rango$(\{a_i \ i \in I(\bar{x})\}) = n$. Per assurdo supponiamo che il rango sia $p < n$. Allora il sistema omogeneo

$$a_i^T d = 0 \quad i \in I(\bar{x})$$

in $I(\bar{x})$ equazioni e n incognite, ammette una soluzione non nulla. Dal teorema 5.2.3, sappiamo che d è una particolare direzione ammissibile; notiamo inoltre che, poiché anche $-d$ è soluzione del sistema omogeneo, anche $-d$ è una direzione ammissibile. Allora possiamo considerare i due punti

$$\begin{aligned} y &= \bar{x} + td \\ z &= \bar{x} + t(-d) \end{aligned}$$

e sappiamo che per valori di t sufficientemente piccoli sono entrambi ammissibili[1]. Osserviamo però che possiamo scrivere

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

ovvero $\bar{x}$ può essere ottenuto come combinazione convessa con coefficiente $\beta = \frac{1}{2}$ di due punti ammissibili e distinti. Ma questo contraddice che $\bar{x}$ sia un vertice.

[1] per trovare il valore di t basta applicare la formula (5.2) alle due direzioni d e $-d$.

Dimostriamo la parte sufficiente, ovvero che se rango$\{a_i \; i \in I(\bar{x})\} = n$ allora $\bar{x}$ è un vertice. Osserviamo preliminarmente che la condizione rango$(\{a_i \; i \in I(\bar{x})\}) = n$ implica che $\bar{x}$ sia l'unica soluzione del sistema

$$a_i^T x = b_i \quad \text{per ogni} \quad i \in I(\bar{x}).$$

Procediamo ora per assurdo e supponiamo che il punto $\bar{x}$ non sia un vertice. Allora non può essere l'unico punto ammissibile e in particolare esisteranno due punti ammissibili v e w distinti da $\bar{x}$ e tali che $\bar{x}$ possa essere espresso come combinazione convessa di v e w ovvero

$$\bar{x} = (1-\beta)v + \beta w \qquad \text{con } \beta \in (0,1).$$

Per ogni $i \in I(\bar{x})$ possiamo scrivere

$$b_i = a_i^T \bar{x} = (1-\beta)a_i^T v + \beta a_i^T w$$

Osserviamo che deve essere necessariamente $a_i^T v = b_i$ e $a_i^T w = b_i$ per ogni $i \in I(\bar{x})$. Se così non fosse, e $a_i^T v > b_i$ e/o $a_i^T w > b_i$ si avrebbe l'assurdo

$$b_i = a_i^T \bar{x} = (1-\beta)a_i^T v + \beta a_i^T w > (1-\beta)b_i + \beta b_i = b_i.$$

Ma allora otteniamo che sia v che w sono soluzioni del sistema

$$a_i^T x = b_i \quad \text{per ogni} \quad i \in I(\bar{x}).$$

Ma questo contraddice che $\bar{x}$ sia l'unica soluzione. □

Possiamo enunciare alcuni corollari che discendono direttamente dal teorema appena dimostrato.

Corollario 5.3.2 *Sia dato un poliedro $S = \{x \in I\!R^n : \; Ax \geq b\}$. Se la matrice A ha un numero di righe linearmente indipendenti minore di n, allora S non ha vertici.*
In particolare se $m < n$ allora S non ha vertici.

Corollario 5.3.3 *Un poliedro $S = \{x \in I\!R^n : \; Ax \geq b\}$ ha un numero finito di vertici, pari al massimo a*

$$\begin{pmatrix} m \\ n \end{pmatrix} = \frac{m!}{n!(m-n)!}$$

Esempio 5.3.4 Sia dato il poliedro dell'Esempio 5.2.4

$$\begin{array}{l} 3x_1 - 2x_2 \geq -30 \\ 2x_1 - x_2 \geq -12 \\ x_1 \geq 0 \; x_2 \geq 0 \end{array}$$

che possiamo scrivere in forma matriciale

$$\begin{pmatrix} -3 & -2 \\ 2 & -1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \geq \begin{pmatrix} -30 \\ -12 \\ 0 \\ 0 \end{pmatrix}.$$

Il poliedro è rappresentato in figura 5.4 e i vertici sono indicati con un puntino rosso. Verifichiamo che la condizione espressa dal Teorema 5.3.1 è verificata. I tre vertici sono i punti

$$v_1 = \begin{pmatrix} 0 \\ 12 \end{pmatrix} \quad v_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad v_3 = \begin{pmatrix} 6 \\ 24 \end{pmatrix}.$$

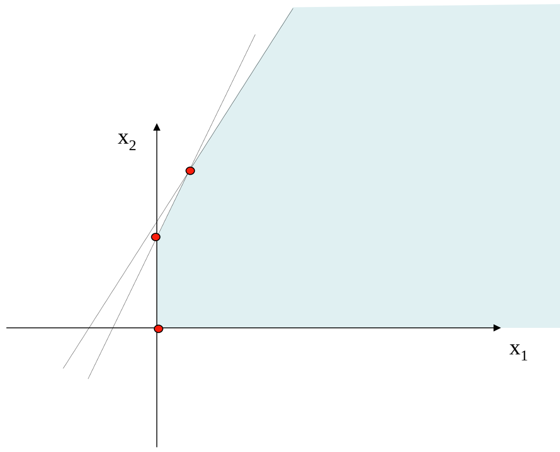

Figura 5.4: Poliedro Esempio 5.3.4.

In v_1 sono attivi i vincoli $2x_1 - x_2 \geq -12,\ x_1 \geq 0$, ovvero $I(v_1) = \{2, 3\}$. Consideriamo le righe della matrice A corrispondenti

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

che si verifica sono linearmente indipendenti.
In v_2 sono attivi i vincoli $3x_1 - 2x_2 \geq -30,\ x_1 \geq 0$, ovvero $I(v_1) = \{1, 3\}$. Consideriamo le righe della matrice A corrispondenti

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

che si verifica sono linearmente indipendenti.
In v_3 sono attivi i vincoli $2x_1 - x_2 \geq -12,\ 3x_1 - 2x_2 \geq -30$, ovvero $I(v_1) = \{1, 2\}$. Consideriamo le righe della matrice A corrispondenti

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

che si verifica sono linearmente indipendenti. □

Si noti che il teorema 5.3.1 non esclude che in un vertice siano attivi più di n vincoli.

Esempio 5.3.5 Consideriamo il poliedro dell'Esempio 5.3.4 con l'aggiunta di un vincolo $x_2 \leq 24$, ovvero:

$$\begin{array}{l} 3x_1 - 2x_2 \geq -30 \\ 2x_1 - x_2 \geq -12 \\ -x_2 \geq -24 \\ x_1 \geq 0, \\ x_2 \geq 0 \end{array}$$

Si verifica graficamente che il punto $v_3 = \begin{pmatrix} 6 \\ 24 \end{pmatrix}$ è ancora un vertice. Osserviamo che in v_3 sono ora attivi tre vincoli $I(v_3) = \{1, 2, 3\}$ e le righe della matrice A corrispondenti sono:

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

Si verifica facilmente che il rango di questa matrice è 2. □

Esempio 5.3.6 Sia dato il poliedro

$$\begin{aligned} x_1 + 2x_2 + x_3 &\leq 3 \\ 3x_1 - x_2 + x_3 &\leq 2 \\ 2x_1 + x_2 + x_3 &\leq 3 \\ 4x_1 + x_2 + 2x_3 &\leq 4 \end{aligned}$$

elenchiamo i vertici. In forma matriciale possiamo scrivere

$$\begin{pmatrix} -1 & -2 & -3 \\ -3 & 1 & -1 \\ -2 & -1 & -1 \\ -4 & -1 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \geq \begin{pmatrix} -3 \\ -2 \\ -3 \\ -4 \end{pmatrix}.$$

Si ha $n = 3$ e $m = 4$. Il numero massimo di vertici è quindi $\frac{m!}{n!(m-n)!} = \frac{4!}{3!} = 4$ e si ottengono considerando tutte le possibili combinazioni di tre righe della matrice A. consideriamo quindi i casi possibili:

1. $I = \{1, 2, 3\}$; il sistema $a_i^T x = b_i$ con $i \in I$ ha rango pari a 3 e l'unica soluzione è il punto $(1,\ 1,\ 0)^T$ che però non risulta ammissibile, perché risulta $(-4\ \ -1\ \ -2)(1,\ 1,\ 0)^T < -4$;
2. $I = \{1, 2, 4\}$ il rango della matrice è $2 < n$, quindi non può essere un vertice;
3. $I = \{2, 3, 4\}$ il sistema ammette l'unica soluzione $(3, 2\ , -5)$ che è ammissibile e quindi è un vertice;
4. $I = \{1, 3, 4\}$ il sistema ammette l'unica soluzione $(2, 2, -3)$ che è ammissibile e quindi è un vertice.

□

Possiamo anche dare una caratterizzare i vertici di un poliedro in forma standard. Vale il seguente teorema che si riporta senza dimostrazione.

Teorema 5.3.7 (Vertici di un poliedro in forma standard) *Sia dato un poliedro $S = \{x \in I\!R^n :\ Ax = b,\ x \geq 0\}$. Un punto $\bar{x} \in S$ è un vertice se e solo se le colonne della matrice A corrispondenti a componenti positive di $\bar{x}$, sono linearmente indipendenti.*

Tuttavia bisogna porre attenzione al fatto che esistono poliedri *che non contengono vertici*. Un esempio e' dato nella figura 5.5 in cui il poliedro è la parte di piano contenuta tra due rette parallele r_1 e r_2.

Vedremo nel prossimo paragrafo che il caso in cui il poliedro non ha vertici è l'unico caso in cui il problema di PL corrispondente può avere soluzione ottima (ovviamente sulla frontiera) senza che nessuna soluzione coincida con un vertice.

Risulta quindi interessante capire quando un poliedro può non ammettere vertice. A tal scopo introduciamo la seguente definizione

Definizione 5.3.8 (Retta) *Sia S un poliedro. Il poliedro contiene una retta se esiste un punto $\bar{x} \in S$ e una direzione $d \in I\!R^n$ tale che*

$$\bar{x} + td \in S \quad \text{per ogni} \quad t \in I\!R.$$

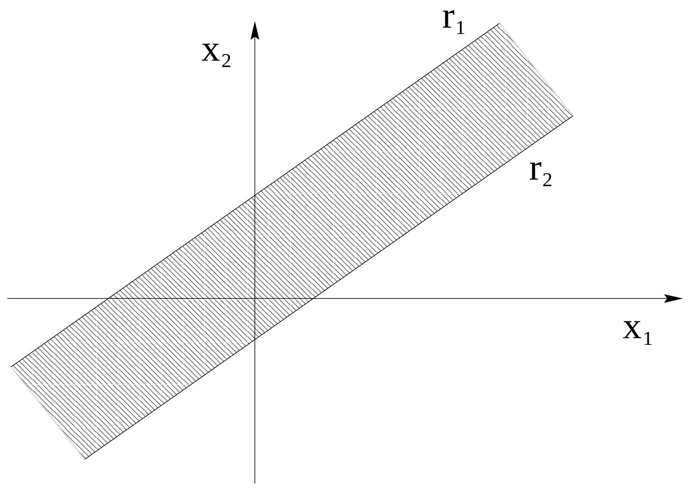

Figura 5.5: Poliedro senza vertici.

La caratterizzazione dei casi in cui un poliedro non ammette vertici è molto semplice ed è riportata nel seguente risultato, di cui omettiamo la prova.

Teorema 5.3.9 *Un poliedro P non vuoto non ha vertici se e solo se contiene una retta.*

È evidente che il poliedro nella Figura 5.5 contiene rette (in particolare contiene, per esempio, r_1, r_2) e quindi, non contiene vertici.

Nel caso le variabili del problema di PL siano vincolate ad essere tutte non negative ovvero tra i vincoli compaiono $x \geq 0$, questo implica che il poliedro ammissibile è interamente contenuta nel primo ortante e quindi *non può contenere rette.* Quindi, in base al teorema precedente, tutti i poliedri contenuti nel primo ortante o sono vuoti o hanno dei vertici. Notiamo che questa è sicuramente la classe di poliedri che più frequentemente si incontra nelle applicazioni.
In particolare quindi si ha il seguente risultato per problemi in forma standard.

Corollario 5.3.10 *Sia dato un poliedro del tipo $S = \{x \in I\!R^n : \; Ax = b, \; x \geq 0\}$. Se il poliedro non è vuoto, ammette sempre un vertice.*

5.4 Il teorema fondamentale della PL

Consideriamo ora il problema di Programmazione Lineare

$$\min_{x \in S} \; c^T x$$

con S poliedro. È da notare che non è possibile ricondursi nel caso generale a problemi con soli vincoli di uguaglianza del tipo

$$\begin{array}{ll} \min & c^T x \\ & Ax = b \end{array}$$

con A matrice $m \times n$. Tali problemi sono infatti di scarsa rilevanza pratica in quanto è possibile dimostrare che se esiste una soluzione ammissibile e il problema non è illimitato allora *tutte* le soluzioni ammissibili sono ottime. In particolare vale il seguente risultato.

Teorema 5.4.1 *Sia dato il problema di PL*

$$\begin{array}{ll} \min & c^T x \\ & Ax = b \end{array}$$

con A matrice $m \times n$. Se esiste un punto $\bar{x}$ tale che $A\bar{x} = b$ allora

1. *il problema è illimitato inferiormente (in questo caso esiste un direzione d ammissibile tale che $c^T d < 0$)*

   oppure

2. *tutte le soluzioni ammissibili sono ottime (in questo caso per ogni direzione d ammissibile risulta $c^T d = 0$)*

Dimostrazione. (non in programma) Se $\bar{x}$ è soluzione unica del sistema $Ax = b$, allora banalmente è anche ottima (e non esiste alcuna direzione $d \neq$ ammissibile perché il sistema $Ad = 0$ non ammette soluzione non nulla). Supponiamo quindi che $\bar{x}$ non sia l'unico punto ammissibile, allora per ogni d tale che $Ad = 0$ risulta $A(\bar{x}+td) = b$ per ogni $t > 0$. Se per ogni d ammissibile risulta $c^T d = 0$ allora possiamo scrivere

$$c^T(\bar{x} + td) = c^T \bar{x},$$

cioè tutte le soluzioni ammissibili hanno lo stesso valore della funzione obiettivo.
Se invece esiste una d ammissibile per cui $c^T d \neq 0$, possiamo senza perdere di generalità supporre che sia $c^T d < 0$. Notiamo infatti che, se d è ammissibile anche $-d$ lo è (se $Ad = 0$, risulta anche $A(-d) = 0$) e quindi, se fosse $c^T d > 0$, sarebbe sufficiente considerare la direzione $-d$. Possiamo allora scrivere

$$c^T(\bar{x} + td) = c^T \bar{x} + tc^T d.$$

Al tendere di t ad ∞ di ha $c^T(\bar{x} + td) \to -\infty$. □

Quindi il problema con soli vincoli di uguaglianza, si riduce essenzialmente allo studio di sistemi di equazioni lineari.
Nel seguito quindi faremo riferimento solo a problemi di Programmazione Lineare di tipo

$$\begin{array}{ll} \min & c^T x \\ & Ax \geq b \end{array} \tag{5.5}$$

$$\begin{array}{ll} \min & c^T x \\ & Ax = b \\ & x \geq 0 \end{array} \tag{5.6}$$

Si tratta quindi di stabilire se un problema di PL ammette soluzione e come caratterizzare la soluzione ottima.
Abbiamo già dimostrato con i Teoremi 3.3.4 e 3.3.5 che *se esiste* una soluzione ottima, allora si trova sulla frontiera del poliedro. I prossimi teoremi caratterizzano in modo più completo i problemi di Programmazione Lineare.
Faremo riferimento inizialmente a problemi di PL in forma (5.5) ed useremo l'ipotesi semplificativa che il poliedro non contenga rette[2], che, in base al Teorema 5.3.9, ci assicura l'esistenza di almeno un vertice del poliedro.

---

[2]Per completezza riportiamo anche il "vero" enunciato del Teorema Fondamentale della Programmazione Lineare

Teorema 5.4.2 *Sia dato un problema di PL. Allora una e una sola delle seguenti affermazioni è vera:*

1. *La regione ammissibile è vuota;*
2. *Il problema è illimitato;*

Teorema **5.4.3 (Teorema fondamentale della PL)** *Sia dato il problema di PL in forma (5.5). Supponiamo che il poliedro $S = \{x \in R^n : \ Ax \geq b\}$ non contenga rette. Allora è vera una e una sola delle delle seguenti tre affermazioni.*

(a) *Il problema non ammette soluzioni ammissibili (la regione ammissibile è vuota).*

(b) *Il problema è illimitato inferiormente sulla regione ammissibile.*

(c) *Esiste almeno una soluzione ottima e almeno una di esse è un vertice del poliedro.*

Dimostrazione. Ovviamente le tre affermazioni dell'enunciato sono incompatibili, nel senso che se è vera una, non possono essere vere le altre due. Quindi, per dimostrare il teorema, basterà mostrare che non può succedere che non si verifichi nessuna delle tre. Mostreremo questo facendo vedere che se non sono vere né (a) né (b), allora deve essere vera la (c).

Supponiamo quindi che la regione ammissibile sia non vuota e che il problema non sia illimitato inferiormente. La dimostrazione è divisa in due parti. Nella prima parte dimostriamo la seguente affermazione

Per ogni punto $x \in S$, esiste un vertice v^k tale che $c^T v^k \leq c^T x$.

La seconda parte utilizza invece risultati già noti. In particolare, dal Teorema 5.3.3, sappiamo che i vertici del poliedro sono in numero finito; li indichiamo con $v^1, \dots v^p$. Quindi tra tutti i vertici v^i possiamo scegliere quello per cui il valore della funzione obiettivo è minore, che indichiamo con v^*. Risulta quindi $c^T v^* \leq c^T v^i$ per ogni vertice v^i e quindi in particolare possiamo scrivere anche per il vertice v^k determinato nel prima parte

$$c^T v^* \leq c^T v^k$$

Mettendo insieme le due affermazioni in rosso, possiamo finalmente scrivere

$$c^T v^* \leq c^T v^k \leq c^T x, \quad \text{per ogni } x \in S$$

il che prova che l'affermazione (c) è vera.
Si tratta quindi di dimostrare che esiste un vertice v^* tale che $c^T v^* \leq c^T x$ per ogni x ammissibile.
Se il poliedro contiene solo un punto v^* allora v^* è ovviamente la soluzione ottima ed è anche un vertice.
Supponiamo allora che il poliedro contenga più di un punto (e quindi infiniti). Sia $\tilde{x}$ una soluzione ammissibile che non sia un vertice, dimostriamo che è possibile trovare un vertice v tale che $c^T v \leq c^T \tilde{x}$.
La dimostrazione è costruttiva. Poiché $\tilde{x}$ non è un vertice, per il Teorema 5.3.1, il sistema

$$a_i^T d = 0, \qquad i \in I(\tilde{x})$$

non ha rango massimo e quindi ammette una soluzione $\bar{d}$ non nulla. Abbiamo due casi possibili

(i) $c^T \bar{d} = 0$;

(ii) $c^T \bar{d} \neq 0$.

Se $c^T \bar{d} = 0$ ovviamente anche $c^T(-\bar{d}) = 0$. Inoltre per il Corollario 5.2.5 i punti $\tilde{x} + t\bar{d}$ con $t \in [0, t^+_{\max}]$ e $\tilde{x} - t\bar{d}$ con $t \in [0, t^-_{\max}]$ sono ammissibili. Poiché il poliedro non contiene rette deve risultare che almeno uno tra $t^+_{\max}$ e $t^-_{\max}$ è $< \infty$. Senza perdita di generalità possiamo assumere che sia $t^+_{\max} < \infty$.
Se invece $c^T \bar{d} \neq 0$, possiamo assumere, senza perdere di generalità, che $c^T \bar{d} < 0$ (altrimenti potremmo scegliere la direzione $-\bar{d}$ che, per il Corollario 5.2.5, è ammissibile e tale che $c^T(-\bar{d}) \leq 0$). In questo caso abbiamo che

$$c^T(\tilde{x} + td) = c^T\tilde{x} + tc^T d = c^T\tilde{x} - t|c^T d| < c^T\tilde{x} \qquad \text{per ogni } t > 0.$$

---

3. Il problema ammette soluzioni ottime.

Se il problema ammette soluzioni ottime e il poliedro che definisce la regione ammissibile ha dei vertici, allora almeno una soluzione ottima cade su un vertice.

Se $t \to \infty$, si avrebbe $c^T(\tilde{x} + td) \to -\infty$ che contraddice l'ipotesi che il problema non sia illimitato. Quindi necessariamente $t \leq t^+_{\max} < \infty$.
In entrambi i casi otteniamo che $\tilde{x} + t\bar{d}$ è ammissibile per $0 \leq t \leq t^+_{\max} < \infty$.
Definiamo il punto

$$y = \tilde{x} + t^+_{\max}\bar{d};$$

ricordando che $c^T\bar{d} \leq 0$ risulta

$$c^T y = c^T\tilde{x} - t^+_{\max}|c^T d| \leq c^T\tilde{x}.$$

Verifichiamo quali vincoli sono attivi in y. Ricordando che per $i \in I(\tilde{x})$ risulta $a_i^T\bar{d} = 0$, otteniamo

$$a_i^T y = a_i^T(\tilde{x} + t^+_{\max}\bar{d}) = a_i^T\tilde{x} = b_i \qquad i \in I(\tilde{x})$$

Quindi $I(\tilde{x}) \subseteq I(y)$. Facciamo ora vedere che $I(\tilde{x}) \subset I(y)$, ovvero che in y è attivo almeno un vincolo in più rispetto ad $\tilde{x}$. Consideriamo quindi $j \notin I(\tilde{x})$ e indico con $j_{\max}$ un indice per cui è raggiunto il minimo nella formula (5.3), cioè:

$$t^+_{\max} = \frac{a^T_{j_{\max}}\tilde{x} - b_{j_{\max}}}{|a^T_{j_{\max}}\bar{d}|}$$

Per definizione risulta che l'indice $j_{\max} \notin I(\tilde{x})$ e $a^T_{j_{\max}}\bar{d} < 0$. Possiamo allora scrivere:

$$a^T_{j_{\max}}x = a^T_{j_{\max}}\tilde{x} + t^+_{\max}a^T_{j_{\max}}\bar{d} = a^T_{j_{\max}}\tilde{x} - t^+_{\max}|a^T_{j_{\max}}d| = a^T_{j_{\max}}\tilde{x} - \frac{a^T_{j_{\max}}\tilde{x} - b_{j_{\max}}}{|a^T_{j_{\max}}\bar{d}|}|a^T_{j_{\max}}\bar{d}| = b_{j_{\max}}$$

Quindi abbiamo che $I(y) \supseteq I(\tilde{x}) \cup \{j_{\max}\}$, cioè in y è attivo almeno un vincolo che non era attivo in $\tilde{x}$. Osserviamo che potrebbe esistere più di un indice per cui è raggiunto il minimo nella formula (5.3). In questo caso avrei attivi in y tanti vincoli in più rispetto a $\tilde{x}$ quanti sono tali indici. Abbiamo quindi dimostrato che, a partire da un qualunque punto ammissibile $\tilde{x}$ che non è un vertice, possiamo determinare un nuovo punto y con valore della funzione obiettivo non superiore e con un numero di vincoli attivi linearmente indipendenti maggiore rispetto a $\tilde{x}$. Se y non è un vertice, possiamo ripetere lo stesso procedimento fino a quando non troviamo un punto in cui sono attivi n vincoli linearmente indipendenti, cioè un vertice. Quindi abbiamo dimostrato l'affermazione che ci serviva:

Per ogni punto $x \in S$, esiste un vertice v tale che $c^T v \leq c^T x$.

La dimostrazione è conclusa. □

Illustriamo con un esempio la tecnica costruttiva utilizzata nella dimostrazione del teorema precedente.

Esempio 5.4.4 Consideriamo il problema di PL

$$\begin{array}{ll} \min & 4x_1 - x_2 \\ & 3x_1 - 2x_2 \geq -30 \\ & 2x_1 - x_2 \geq -12 \\ & x_1 \geq 0, \\ & x_2 \geq 0 \end{array}$$

il cui poliedro è riportato in Figura 5.6.

Il poliedro ha tre vertici che abbiamo già calcolato nell'Esempio 5.3.4, che indichiamo con v^1, v^2, v^3 (puntini rossi in figura). Sia $x = (3,\ 0)$ un punto ammissibile (puntino blu in figura). Si verifica facilmente che $\tilde{x}$ non è un vertice. Infatti $I(\tilde{x}) = \{4\}$ (è attivo solo il vincolo $x_2 \geq 0$) e risulta ovviamente rango$\{a_i\ i \in I(\bar{x})\} = 1 < n = 2$. Consideriamo allora il sistema omogeneo

$$a_4^T d = 0 \qquad \text{ovvero} \quad d_2 = 0.$$

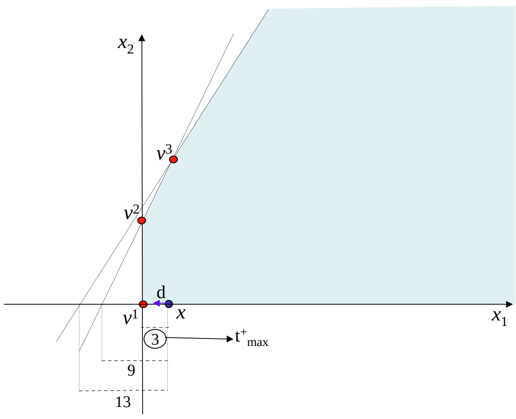

Figura 5.6: Figura relativa all'Esempio 5.4.4.

Quindi una qualunque direzione del tipo $(d_1,\ 0)^T$ con $d_1 \neq 0$ è ammissibile in $\tilde{x}$. Sia $\bar{d} = (-1,\ 0)$ una possibile soluzione (indicata in azzurro in Figura 5.6). Risulta $c^T\bar{d} = 4d_1 = -4 < 0$ e si consideri il punto

$$\tilde{x} + t\bar{d} = (3,\ 0)^T + t(-1,\ 0)^T = (3 - t,\ 0)^T.$$

Calcoliamo il valore di

$$t^+_{\max} = \min\left\{\frac{9+30}{3}, \frac{6+12}{2}, \frac{3}{1}\right\} = \min\{13, 9, 3\} = 3.$$

Si osservi in Figura 5.6 che i valori 13,9,3 che compaiono dentro il min per il calcolo di $t^+_{\max}$ corrispondono rispettivamente al valore del passo t per cui il punto $\tilde{x} + t\bar{d}$ "sfonda" rispettivamente il primo, il secondo e il terzo vincolo. Il valore di $t^+_{\max} = 3$ corrisponde al massimo valore del passo t per cui il punto rimane ammissibile. Sia allora

$$y = \tilde{x} + t^+_{\max}\bar{d} = (0,\ 0)^T$$

Risulta $I(y) = \{3, 4\} = I(\tilde{x}) \cup \{3\}$. Inoltre la riga a_3 è linearmente indipendente da a_4. Si tratta del vertice v^1. □

Nel caso di problemi di PL in forma standard (5.6), ricordando che in questo caso il poliedro ammissibile non contiene rette, possiamo enunciare il teorema fondamentale come segue:

Teorema 5.4.5 (Teorema fondamentale della PL) *Sia dato il problema di PL in forma (5.6). Allora è vera una e una sola delle delle seguenti tre affermazioni.*

(a) *Il problema non ammette soluzioni ammissibili (la regione ammissibile è vuota).*

(b) *Il problema è illimitato inferiormente sulla regione ammissibile.*

(c) *Esiste almeno una soluzione ottima e almeno una di esse è un vertice del poliedro.*

Capitolo 6

Le condizioni di ottimalità

6.1 Introduzione

Con riferimento al problema di ottimizzazione $\min_{x\in S} f(x)$, illustriamo nel seguito alcune condizioni di ottimalità in relazione alle classi di problemi di nostro interesse. In particolare ci riferiremo a problemi non vincolati, a problemi convessi con S generico, a problemi di programmazione matematica con particolare attenzione al caso di problemi con insieme ammissibile poliedrale.
In termini molto generali, una condizione di ottimalità è una condizione (necessaria, sufficiente, necessaria e sufficiente) perché un punto x^* risulti una soluzione ottima (locale o globale) del problema. Ovviamente, una condizione di ottimalità sarà significativa se la verifica della condizione risulta più "semplice" o più "vantaggiosa" (da qualche punto di vista) rispetto all'applicazione diretta della definizione. Le condizioni di ottimalità si esprimono infatti, tipicamente, attraverso sistemi di equazioni, sistemi di disequazioni, condizioni sugli autovalori di matrici opportune.

Lo studio delle condizioni di ottimalità ha sia motivazioni di natura teorica, sia motivazioni di natura algoritmica. Dal punto di vista teorico, una condizione di ottimalità può servire a caratterizzare analiticamente le soluzioni di un problema di ottimo e quindi consentire di svolgere analisi *qualitative*, anche in assenza di soluzioni numeriche esplicite; un esempio è l'analisi della sensibilità delle soluzioni di un problema di ottimo rispetto a variazioni parametriche. Dal punto di vista algoritmico, una condizione *necessaria* può servire a restringere l'insieme in cui ricercare le soluzioni del problema originario e a costruire algoritmi finalizzati al soddisfacimento di tale condizione; una condizione *sufficiente* può servire a dimostrare che un punto ottenuto per via numerica sia una soluzione ottima del problema e quindi a definire criteri di arresto del procedimento risolutivo.

6.2 Direzioni di discesa

Allo scopo di definire condizioni di ottimalità è necessario introdurre la seguente definizione.

Definizione 6.2.1 (Direzione di discesa) *Sia $f : R^n \to R$ e $x \in R^n$. Si dice che un vettore $d \in R^n$, $d \neq 0$ è una direzione di discesa per f in x se esiste $\tilde{t} > 0$ tale che*

$$f(x + td) < f(x), \quad \textit{per ogni} \quad t \in (0, \tilde{t}\,].$$

Il concetto di direzione di discesa insieme al concetto di direzione ammissibile introdotta in Def. 5.1.1, ci consentirà di caratterizzare i minimi locali di una funzione.

Nel caso di funzione lineare $f(x) = c^T x$, si ottiene facilmente dalla definizione ($c^T(x+td) = c^T x + tc^T d < c^T x$) che una direzione è di discesa se e solo risulta $c^T d < 0$.

> Se f è una funzione lineare, ossia se $f(x) \equiv c^T x$ allora d è una direzione di discesa (in un qualsiasi punto $x \in I\!R^n$) se e solo se $c^T d < 0$.

Osserviamo che abbiamo già implicitamente utilizzato il concetto di direzione di discesa nel capitolo 5 nella dimostrazione del Teorema fondamentale della Programmazione Lineare.
In generale, per funzioni continuamente differenziabili, è possibile caratterizzare la direzione di discesa utilizzando informazioni sulle derivate prime della funzione.
Vale in particolare la seguente condizioni sufficiente:

Teorema 6.2.2 (Condizione di discesa) *Supponiamo che f sia continuamente differenziabile nell'intorno di un punto $x \in R^n$ e sia $d \in R^n$ un vettore non nullo. Se risulta*

$$\nabla f(x)^T d < 0,$$

allora la direzione d è una direzione di discesa per f in x.

Dimostrazione. È noto (vedi Appendice A) che per una qualunque funzione continuamente differenziabile possiamo scrivere

$$f(x + td) = f(x) + t\nabla f(x)^T d + \alpha(x, t)$$

dove $\alpha(x, t)$ soddisfa $\lim_{t \to 0} \frac{\alpha(x, t)}{t} = 0$. Per valori sufficientemente piccoli di t possiamo dunque scrivere

$$f(x + td) - f(x) \simeq t\nabla f(x)^T d;$$

se $\nabla f(x)^T d < 0$ risulta allora $f(x + td) - f(x) < 0$ che quindi dimostra che d è una direzione di discesa.$\square$

Ricordando che $\nabla f(x)^T d = \|\nabla f(x)\| \|d\| \cos\theta$ dove θ è l'angolo compreso tra $\nabla f(x)$ e d, dal punto di vista geometrico la condizione $\nabla f(x)^T d < 0$ esprime il fatto che se una direzione d forma un angolo ottuso con il gradiente di f in x, allora d è una direzione di discesa per f in x.

> Sia f continuamente differenziabile e sia $d \in R^n$ un vettore non nullo. Se l'angolo θ tra $\nabla f(x)$ e d soddisfa
>
> $$\theta > 90^o$$
>
> allora d è di discesa per f in x.

Tra le direzioni di discesa, un ruolo particolarmente importante è svolto dalla direzione dell'*antigradiente* $d = -\nabla f(x)$. Se $\nabla f(x) \neq 0$, la direzione dell'antigradiente $d = -\nabla f(x)$ è sempre una direzione di discesa, infatti risulta

$$\nabla f(x)^T d = -\nabla f(x)^T \nabla f(x) = -\|\nabla f(x)\|^2 < 0.$$

La definizione è illustrata nella figura successiva, con riferimento agli insiemi di livello di una funzione in R^2.

Naturalmente è possibile dare una condizione sufficiente analoga al teorema 6.2.2 affinché una direzione sia di salita.

> Sia f sia continuamente differenziabile nell'intorno di un punto $x \in R^n$ e sia $d \in R^n$ un vettore non nullo. Se risulta
>
> $$\nabla f(x)^T d > 0,$$
>
> allora la direzione d è una direzione di salita per f in x.

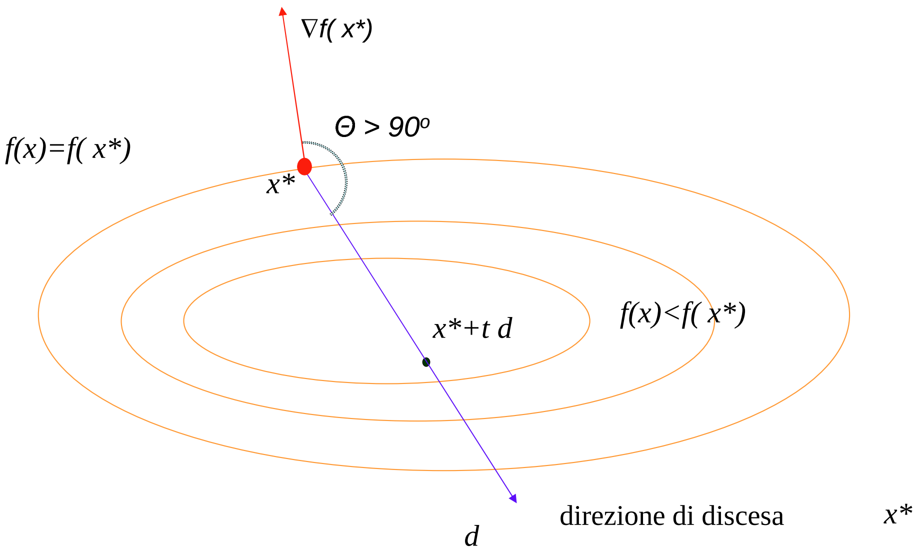

Figura 6.1: Esempio di direzione d di discesa in $\mathbb{R}^2$.

Naturalmente se $\nabla f(x) \neq 0$, la direzione del gradiente $d = \nabla f(x)$ è sempre una direzione di salita in x.

Osserviamo che nel caso di funzione lineare $f = c^T x$, risulta $\nabla f(x) = c$, quindi la condizione espressa dal Teorema 6.2.2 coincide con la caratterizzazione ottenuta in base della definizione di direzione di discesa. Nel caso lineare si tratta però di una condizione necessaria e sufficiente, mentre nel caso generale è solo una condizione sufficiente.
Nel caso di funzione lineare vale quindi la seguente caratterizzazione:

Sia f è una funzione lineare, ossia $f(x) \equiv c^T x$ allora

1. d è una direzione di discesa in x se e solo se $\nabla f(x)^T d = c^T d < 0$;
2. d è una direzione di salita in x se e solo se $\nabla f(x)^T d = c^T d > 0$;
3. d è una direzione lungo cui la funzione si mantiene costante rispetto al valore in x se e solo se $\nabla f(x)^T d = c^T d = 0$.

Nel caso generale, la condizione di discesa enunciata è solo sufficiente, in quanto possono esistere direzioni di discesa tali che $\nabla f(x)^T d = 0$. Illustriamo le varie possibilità nel seguente esempio.

Esempio 6.2.3 Consideriamo la funzione

$$f(x) = x_1^3 + x_2^2 - 3x_1.$$

sia dato il punto $\bar{x} = (-1,\ 0)^T$ e la direzione $d = (-1,\ 0)^T$. Il gradiente della funzione vale

$$\nabla f(x) = \begin{pmatrix} 3x_1^2 - 3 \\ 2x_2 \end{pmatrix}.$$

Si verifica facilmente che il gradiente della funzione si annulla nel punto $\bar{x}$, quindi la condizione sufficiente espressa dal teorema 6.2.2 non è soddisfatta in quanto $\nabla f(\bar{x})^T d = 0$. La direzione d è però di discesa in $\bar{x}$. Infatti risulta $\bar{x} + td = (-1-t,\ 0)^T$ e la funzione vale:

$$f(\bar{x} + td) = (-1-t)^3 - 3(-1-t) = -1 - t^3 - 3t - 3t^2 + 3 + 3t = -t^3 + 3t^2 + 2.$$

Mostriamo che esiste un valore di $\bar{t}$ tale che $-t^3 - 3t^2 + 2 < 2 = f(\bar{x})$ per ogni $0 < t \leq \bar{t}$. La condizione che si ottiene è $-t^3 - 3t^2 < 0$ che risulta verificata per qualunque valore di $t > 0$, quindi effettivamente la direzione è di discesa nel punto.
si consideri ora il punto $\tilde{x} = (1,\ 0)^T$ e la stessa direzione $d = (-1,\ 0)^T$. Si verifica facilmente che il gradiente della funzione si annulla nel punto $\tilde{x}$, e quindi la condizione sufficiente espressa dal teorema 6.2.2 anche in questo caso non è soddisfatta in quanto $\nabla f(\tilde{x})^T d = 0$. In questo caso però la direzione d non è di discesa, risulta infatti $\tilde{x} + td = (1-t,\ 0)^T$ e la funzione vale:

$$f(\bar{x} + td) = (1-t)^3 - 3(1-t) = 1 - t^3 - 3t + 3t^2 - 3 + 3t = -t^3 + 3t^2 - 2$$

Dobbiamo quindi verificare se esiste un $\bar{t} > 0$ tale che per ogni $t \in (0, \bar{t}]$ risulta $-t^3 + 3t^2 = t^2(-t+3) < 0$. Si verifica facilmente che la condizione è verificata solo per valori di t abbastanza grandi ($t > 3$) e quindi d non è di discesa. Si tratta invece di una direzione di salita, infatti scegliendo $\bar{t} < 3$, si ottiene che $f(\tilde{x} + td) > f(\tilde{x})$ per ogni $t \in (0, \bar{t}]$.
Consideriamo ora il punto $\hat{x} = (0,\ 0)^T$ in cui il gradiente di f vale $\nabla f(\hat{x}) = (-3,\ 0)^T$. Risulta che la direzione $d = (-1,\ 0)^T$ è di salita in $\hat{x}$, mentre la direzione opposta $-d = (1,\ 0)^T$ è di discesa.

Dall'esempio e dalle considerazioni precedenti risulta evidente che dato in un punto x ed una direzione d, sono possibili solo questi casi:

Sia f sia continuamente differenziabile nell'intorno di un punto $x \in R^n$ e sia $d \in R^n$ un vettore non nullo.

1. se $\nabla f(x)^T d < 0$ allora la direzione è di discesa in x;
2. se $\nabla f(x)^T d > 0$ allora la direzione è di salita in x;
3. se $\nabla f(x)^T d = 0$ allora non si può concludere se la direzione è di discesa o di salita in x o lungo la quale la funzione si mantiene costante.

La condizione del teorema 6.2.2 diventa anche necessaria nel caso di convessità della funzione obiettivo.

Teorema 6.2.4 *Se f è convessa allora una direzione d è di discesa in un punto x se e solo se*

$$\nabla f(x)^T d < 0.$$

Dimostrazione. Sia f convessa, dobbiamo solo dimostrare che la condizione $\nabla f(x)^T d < 0$ è necessaria (la sufficienza è data dal teorema 6.2.2). Sia $x \in \mathbb{R}^n$ e sia d una direzione di discesa per f in x. Possiamo scrivere per la convessità della funzione

$$f(x + \alpha d) \geq f(x) + \alpha \nabla f(x)^T d.$$

Se per assurdo fosse $\nabla f(x)^T d \geq 0$ allora per $\alpha > 0$ sufficientemente piccolo, si avrebbe $f(x + \alpha d) \geq f(x)$ contraddicendo l'ipotesi che d sia di discesa. $\square$

Se f è differenziabile due volte è possibile caratterizzare l'andamento di f lungo una direzione assegnata tenendo conto anche delle derivate seconde e di ciò si può tener conto, come si vedrà in seguito, per stabilire condizioni di ottimo del secondo ordine.

Introduciamo la definizione seguente.

Definizione 6.2.5 (Direzione a curvatura negativa)

Sia $f : R^n \to R$ due volte continuamente differenziabile nell'intorno di un punto $x \in R^n$. Si dice che un vettore $d \in R^n$, $d \neq 0$ è una direzione a curvatura negativa per f in x se risulta

$$d^T \nabla^2 f(x) d < 0. \qquad \square$$

Una direzione a curvatura negativa è quindi tale che la derivata direzionale seconda è negativa in x, per cui diminuisce localmente la derivata direzionale del primo ordine. In particolare vale il risultato seguente.

Teorema 6.2.6 (Condizione di discesa del secondo ordine)

Sia $f : R^n \to R$ due volte continuamente differenziabile nell'intorno di un punto $x \in R^n$ e sia $d \in R^n$ un vettore non nullo. Supponiamo che risulti

$$\nabla f(x)^T d = 0,$$

e che d sia una direzione a curvatura negativa in x, ossia che

$$d^T \nabla^2 f(x) d < 0.$$

Allora la direzione d è una direzione di discesa per f in x.

Dimostrazione. Poiché f è differenziabile due volte, si ha:

$$f(x + td) = f(x) + t\nabla f(x)^T d + \frac{1}{2} t^2 d^T \nabla^2 f(x) d + \beta(x, td)$$

in cui $\beta(x, td)/t^2 \to 0$. Essendo per ipotesi $\nabla f(x)^T d = 0$, si può scrivere:

$$\frac{f(x + td) - f(x)}{t^2} = \frac{1}{2} d^T \nabla^2 f(x) d + \frac{\beta(x, td)}{t^2}$$

e quindi, poiché

$$\lim_{t\to 0} \frac{\beta(x, td)}{t^2} = 0,$$

per valori sufficientemente piccoli di t si ha $f(x + td) - f(x) < 0$, per cui d è una direzione di discesa. $\square$

Esempio 6.2.7 Consideriamo la funzione dell'esempio 6.2.3,

$$f(x) = x_1^3 + x_2^2 - 3x_1.$$

nel punto $\bar{x} = (-1,\ 0)^T$. La direzione $d = (-1,\ 0)^T$ è tale che $\nabla f(\bar{x})^T d = 0$, quindi la condizione sufficiente espressa dal teorema 6.2.2 non è soddisfatta. Verifichiamo la condizione del Teorema 6.2.6. La matrice hessiana è

$$\nabla^2 f(x) = \begin{pmatrix} 6x_1 & 0 \\ 0 & 2 \end{pmatrix} \stackrel{\bar{x}}{=} \begin{pmatrix} -6 & 0 \\ 0 & 2 \end{pmatrix}.$$

Risulta allora $d^T\nabla^2 f(\bar{x})d = -6$ e quindi la direzione soddisfa la condizione sufficiente del Teorema 6.2.6, per cui si può concludere che è di discesa.

6.3 Condizioni di ottimalità

In questo paragrafo ricaviamo le prima condizioni di ottimalità per il problema

$$\begin{array}{ll} \min & f(x) \\ & x \in S \end{array} \tag{P}$$

Una conseguenza immediata della Definizione 6.2.1 di direzione di discesa e della Definizione 5.1.1 di direzione ammissibile è la condizione necessaria di minimo locale enunciata nel teorema successivo.

Teorema 6.3.1 (Condizione necessaria di minimo locale) *Sia $x^* \in S$ un punto di minimo locale del problema (P) allora non può esistere una direzione ammissibile in x^* che sia anche di discesa per f.*

Dimostrazione. Sia x^* un minimo locale. Per assurdo, se esistesse una direzione d al tempo stesso ammissibile e di discesa in x^*, allora in ogni intorno di x^* sarebbe possibile trovare, per $t > 0$ abbastanza piccolo, un punto $x^* + td \in S$ tale che $f(x^* + td) < f(x^*)$, il che contraddice l'ipotesi che x^* sia un punto di minimo locale. □

La condizione necessaria può consentire di selezionare tra tutti i punti ammissibili i potenziali candidati ad essere punti di minimo locale.
Il risultato espresso nel Teorema 6.3.1 può essere poco significativo qualora vi siano regioni (di frontiera) dell'insieme ammissibile, tali che nei punti di queste regioni non esistano affatto direzioni ammissibili. In tal caso, la condizione del Teorema 6.3.1 non è in grado di discriminare le soluzioni ottime dagli altri punti ammissibili. Una situazione del genere si verifica, tipicamente, quando siano presenti vincoli di eguaglianza non lineari.

Consideriamo, a titolo esemplificativo, un problema in due variabili.

Esempio 6.3.2 Sia dato il problema

$$\begin{array}{ll} \min & x_2 \\ & x_1^2 + x_2^2 = 1 \end{array}$$

L'unico punto di minimo è il punto $(0, \ -1)^T$. Se consideriamo un qualsiasi punto $(x_1, x_2) \in S$ l'insieme delle direzioni ammissibili è vuoto e quindi la condizione del Teorema 6.3.1 è soddisfatta in ogni punto di S e non dà alcuna informazione aggiuntiva per la selezione di punti candidati ad essere ottimi. Tale situazione è illustrata nella figura 6.2.

□

Per ricavare delle condizioni di ottimalità dalla condizione enunciata nel Teorema 6.3.1 occorre utilizzare una caratterizzazione analitica delle direzioni di discesa e delle direzioni ammissibili. In effetti la verifica della condizione espressa dal Teorema 6.3.1 è essenzialmente equivalente all'applicazione della definizione di minimo locale.
Si può però specificare ulteriormente la condizione necessaria utilizzando la caratterizzazione di direzione di discesa del teorema 6.2.2.
Si ottiene allora la seguente condizione necessaria.

Se x^* è un punto di minimo locale del del problema (P) non può esistere una direzione ammissibile $d \in R^n$ in x^* tale che $\nabla f(x^*)^T d < 0$.

Tale condizione può essere equivalentemente enunciata come segue.

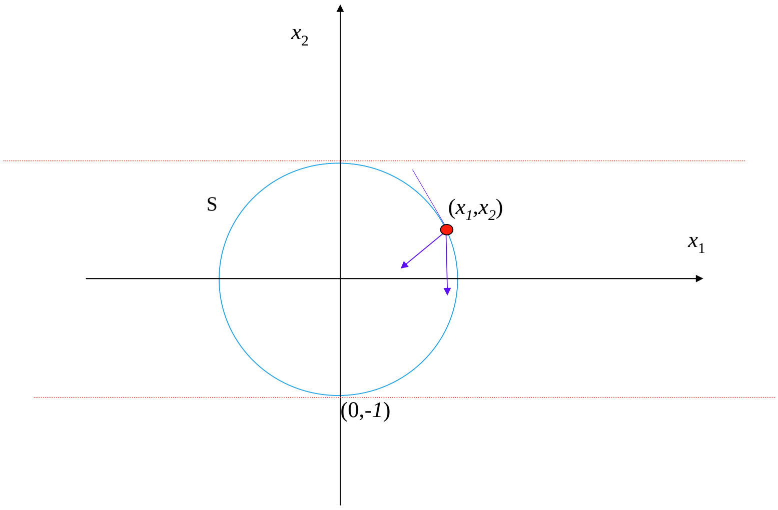

Figura 6.2: Esempio 6.3.2 in cui in qualunque punto ammissibile non esiste direzione ammissibile.

Teorema 6.3.3 (Condizione necessaria di minimo locale) *Se x^* è un punto di minimo locale del del problema (P), allora risulta*

$$\nabla f(x^*)^T d \geq 0 \qquad \textit{per ogni direzione ammissibile } d \in R^n \textit{ in } \; x^*.$$

Ricordando che $\nabla f(x^*)^T d = \|\nabla f(x^*)\| \|d\| \cos\theta$ dove θ è l'angolo compreso tra $\nabla f(x^*)$ e d, dal punto di vista geometrico la condizione $\nabla f(x^*)^T d \geq 0$ richiede che $0^o \leq \theta \leq 90^o$ per ogni d ammissibile.

Se x^* è un punto di minimo locale del del problema (P), allora per ogni d ammissibile l'angolo θ formato con $\nabla f(x^*)$ è

$$0^o \leq \theta \leq 90^o.$$

La condizione espressa dal Teorema 6.3.5 è solo necessaria di minimo locale. Illustriamo la condizione con un esempio.

Esempio 6.3.4 Sia dato il problema di ottimizzazione

$$\begin{array}{ll} \min & -x_1^2 - x_2^2 \\ & -1 \leq x_1 \leq 1 \\ & -1 \leq x_2 \leq 1 \end{array}$$

Si tratta della minimizzazione di una funzione quadratica concava su un poliedro (la regione ammissibile e le curve di livello della funzione obiettivo sono rappresentate in Figura 6.3.) Il gradiente della funzione obiettivo è

$$\nabla f(x) = \begin{pmatrix} -2x_1 \\ -2x_2 \end{pmatrix}$$

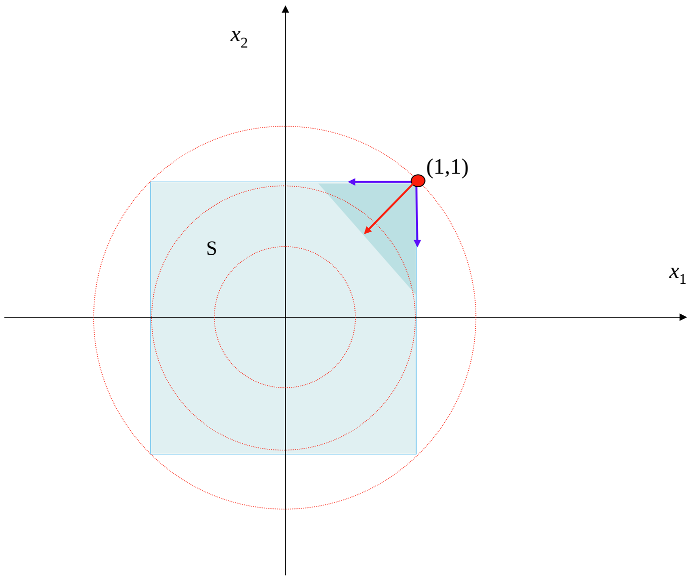

Figura 6.3: Illustrazione Esempio 6.3.4.

Sia dato il punto di minimo locale $x^* = (1,\ 1)^T$ in cui vale $\nabla f(x^*) = (-2,\ \ -2)^T$ (freccia rossa in Figura 6.3). Le direzioni ammissibili nel punto x^* sono individuate in base al Teorema 5.2.3 (in Figura 6.3 sono i vettori nella regione compresa tra le frecce blu). I vincoli possono essere riscritti nella forma

$$\begin{aligned} x_1 &\geq -1 \\ -x_1 &\geq -1 \\ x_2 &\geq -1 \\ -x_2 &\geq -1 \end{aligned}$$

I vincoli attivi nel punto x^* sono $I(x^*) = \{2, 4\}$ e quindi le direzioni ammissibile sono quelle che soddisfano

$$\begin{array}{lll} a_2^T d \geq 0 & (-1,\ 0)\begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \geq 0 & -d_1 \geq 0 \\ a_4^T d \geq 0 & (0,\ \ -1)\begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \geq 0 & -d_2 \geq 0 \end{array}$$

ovvero le direzioni $d = (d_1\ d_2)^T \leq 0$. Risulta $\nabla f(x^*)^T d = (-2\ \ -2)\begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = -2d_1 - 2d_2 \geq 0$ per ogni d ammissibile. Quindi la condizione necessaria è soddisfatta.

Sia dato ora il punto $\bar{x} = (0,\ 0)^T$ interno alla regione ammissibile (nessun vincolo attivo). Ogni vettore $d \in I\!R^2$ è ammissibile in $\bar{x}$. Inoltre risulta $\nabla f(\bar{x})^T d = 0$ per ogni d, in quanto $\nabla f(\bar{x}) = 0$, quindi è verificata la condizione necessaria. Ciononostante il punto $\bar{x}$ non è un minimo locale, anzi è un massimo. Infatti per ogni $d \in I\!R^2$ risulta $\bar{x} + td = (td_1,\ td_2)^T$ e la funzione vale $f(\bar{x} + td) = -t^2(d_1^2 + d_2^2)$. Per ogni $t \neq 0$ e $d \not\equiv 0$ risulta

$$f(\bar{x} + td) = -t^2(d_1^2 + d_2^2) < f(\bar{x}) = 0,$$

ovvero qualunque d non identicamente nulla è di discesa in $\bar{x}$. $\square$

La condizione espressa dal Teorema 6.3.5 diventa però anche sufficiente sotto opportune ipotesi sulla funzione obiettivo e sull'insieme S. Si ha in particolare il seguente risultato.

Possiamo dare delle condizioni che utilizzano la caratterizzazione del secondo ordine di direzione di discesa espressa al Teorema 6.2.6. In particolare quindi abbiamo:

Teorema 6.3.5 (Condizione necessaria del 2^o ordine di minimo locale) *Se x^* è un punto di minimo locale del del problema (P), allora per ogni d ammissibile in x^* tale che $\nabla f(x^*)^T d = 0$ risulta*

$$d^T \nabla^2 f(x^*)^T d \geq 0 \qquad \textit{per ogni direzione ammissibile } d \in R^n \textit{ in } x^*.$$

A partire dalle condizioni espresse dai teoremi 6.3.5 e 6.5.5 è possibile ricavare condizioni più specifiche in base alle caratterizzazioni delle direzioni ammissibili, e quindi in base alla struttura e alle proprietà delle funzioni che definiscono i vincoli del problema.
Considereremo nel seguito solo i casi di nostro interesse.

6.4 Ottimizzazione non vincolata

Abbiamo già osservato nel capitolo 2 che i problemi non vincolati sono una particolare classe di problemi non lineari, in cui $S = R^n$. Il problema in oggetto è quindi

$$\min_{x \in I\!R^n} f(x), \qquad (P-NV)$$

con $f : I\!R^n \to I\!R$ non lineare e continuamente differenziabile. Osserviamo preventivamente che in questo caso non si può applicare in modo diretto il teorema di Weierstrass per l'esistenza di un minimo. Condizioni di esistenza possono essere comunque ottenute con riferimento ad un insieme ammissibile "fittizio". In particolare, dato un qualunque punto $\hat{x}$ in cui la funzione vale $f(\hat{x})$, e si considera l'insieme non vuoto (contiene almeno $\hat{x}$)

$$\mathcal{L}(\hat{x}) = \{x \in I\!R^n : \ f(x) \leq f(\hat{x})\}.$$

Possiamo allora considerare il problema "vincolato"

$$\begin{array}{ll} \min & f(x) \\ & x \in \mathcal{L}(\hat{x}) \end{array} \qquad (P-NV\mathcal{L})$$

Se $\hat{x}$ non è il minimo globale di (P-NV), allora il minimo globale x^*, se esiste, sicuramente si trova in $\mathcal{L}(\hat{x})$. Infatti per definizione $f(x^*) \leq f(x)$ per ogni x e quindi in particolare vale anche $f(x^*) \leq f(\hat{x})$. Quindi risolvere (P-NV) é equivalente a risolvere (P-NV$\mathcal{L}$).
Per assicurare l'esistenza di un minimo globale di (P-NV$\mathcal{L}$) (e quindi di (P-NV), possiamo utilizzare il teorema di Weierstrass. È sufficiente richiedere che esista un $\hat{x}$ per cui l'insieme $\mathcal{L}(\hat{x})$ sia compatto.

Se esiste un $\hat{x} \in I\!R^n$ per cui l'insieme $\mathcal{L}(\hat{x})$ é compatto, allora il problema (P-NV) ammette un minimo globale.

Deriviamo ora le condizioni di ottimo per (P-NV). È ovvio che, nel caso non vincolato, dato un qualunque punto $x \in I\!R^n$ tutte le direzioni $d \in R^n$ sono ammissibili in x.

Se $S \equiv I\!R^n$ ogni vettore $d \in I\!R^n$ è una direzione ammissibile.

Dal teorema 6.3.5 si ottiene quindi la seguente condizione necessaria.

Teorema 6.4.1 (Condizione necessaria di minimo locale non vincolato) *Sia $f : R^n \to R$ continuamente differenziabile su R^n e sia $x^\star \in R^n$. Se x^* è un punto di minimo locale non vincolato di f in R^n allora si ha*

$$\nabla f(x^\star) = 0.$$

Dimostrazione. Se $x^* \in R^n$ è un punto di minimo locale non possono esistere direzioni di discesa in x^*. Se $\nabla f(x^*) \neq 0$, esisterebbe una direzione $d = -\nabla f(x^*) \neq 0$ di discesa in x^*. Ma questo è assurdo. □

Definizione 6.4.2 (Punto stazionario) *Sia $f : R^n \to R$ continuamente differenziabile su R^n. Un punto $\bar{x}$ tale che $\nabla f(\bar{x}) = 0$ si dice* **punto stazionario** *di f.*

I possibili candidati ad essere punti di minimo locale del problema (P-NV) si determinano determinando le soluzioni di $\nabla f(x) = 0$.

Esempio 6.4.3 Sia dato il problema

$$f(x_1, x_2) = {x_1}^4 + {x_2}^4 - 3x_1x_2,$$

sapendo che esiste un punto di minimo, determinarlo utilizzando le condizioni necessarie.
Il gradiente della funzione è dato da:

$$\nabla f = \begin{pmatrix} 4x_1^3 - 3x_2 \\ 4x_2^3 - 3x_1 \end{pmatrix}.$$

Dall'annullamento del gradiente si ottiene il sistema

$$\begin{cases} 4x_1^3 - 3x_2 = 0 \\ 4x_2^3 - 3x_1 = 0 \end{cases} \qquad \begin{cases} x_2\left(\frac{256}{27}x_2^8 - 3\right) = 0 \\ x_1 = \frac{4}{3}x_2^3 \end{cases}$$

Si ottengono le soluzioni: $A(0, 0)$, $B\left(\frac{\sqrt{3}}{2}, \frac{\sqrt{3}}{2}\right)$, $C\left(-\frac{\sqrt{3}}{2}, -\frac{\sqrt{3}}{2}\right)$.
Le curve di livello della funzione sono nella Figura 6.4 da cui si individuano i tre punti stazionari. Il punto $(0,\ 0)^T$ è un punto di sella con valore della funzione $f(0, 0) = 0$.
I punti $B, C = \left(\pm\frac{\sqrt{3}}{2}, \pm\frac{\sqrt{3}}{2}\right)$ abbiamo

$$\nabla^2 f = 3\begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix}$$

sono punti di minimo globale di valore della funzione obiettivo $f = -\frac{9}{8}$.

□

Senza ulteriori ipotesi sulla funzione obiettivo, la condizione è solo necessaria. Ricordando che se $\nabla f(x^*) \neq 0$ la direzione $d = \nabla f(x^*)$ è di salita nel punto x^*, si ottiene che la condizione $\nabla f(x^*) = 0$ è anche necessaria di massimo locale. In generale si hanno questi possibili casi.

Sia $f : R^n \to R$ continuamente differenziabile su R^n e sia $x^\star \in R^n$ tale che $\nabla f(x^\star) = 0$ allora è vera una delle seguenti affermazioni:

1. $x^\star$ è un minimo locale (non esistono in $x^\star$ direzioni di discesa);
2. $x^\star$ è un massimo locale (non esistono in $x^\star$ direzioni di salita);
3. $x^\star$ è un punto di sella (esistono in $x^\star$ sia direzioni di discesa che di salita).

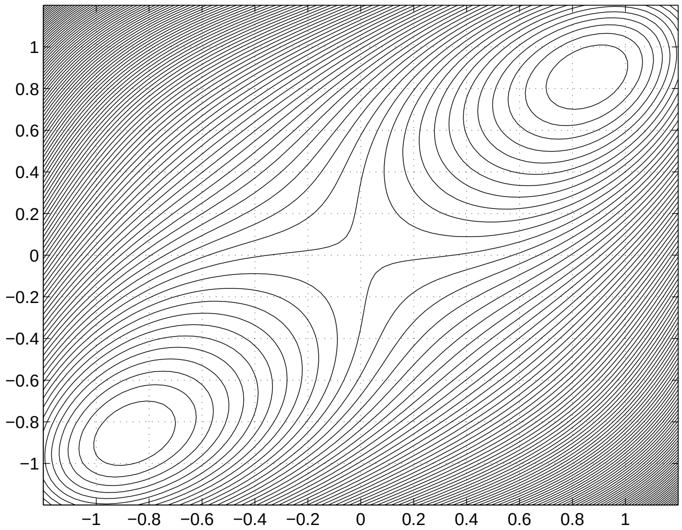

Figura 6.4: Curve di livello della funzione $x_1^4 + x_2^4 - 3x_1x_2$.

Un esempio di punto di sella è in Figura 6.5.

Osserviamo che nel caso non vincolato, dato un punto $x^* \in I\!R^n$, tutte le direzioni $d = x - x^*$ sono banalmente ammissibili. Quindi è possibile ottenere una condizione sufficiente dal teorema 6.5.3. In particolare si ha il risultato seguente.

Teorema 6.4.4 (Condizione sufficiente di minimo globale non vincolato) *Sia $f : R^n \to R$ continuamente differenziabile su R^n e sia $x^\star \in R^n$. Si supponga che f sia convessa. Se $\nabla f(x^\star) = 0$, allora $x^\star$ è un punto di minimo globale di f su R^n. Inoltre, se f è strettamente convessa su R^n, allora $x^\star$ è l'unico punto di minimo globale di f su R^n.*

Dai teoremi 6.4.1 e 6.4.4 si ottiene quindi una condizione necessaria e sufficiente di minimo globale non vincolato.

Teorema 6.4.5 (Condizione necessaria e sufficiente di minimo globale non vincolato) *Sia $f : R^n \to R$, con ∇f continuo su R^n e si supponga che f sia convessa. Allora $x^\star$ è un punto di minimo globale di f su R^n se e solo se $\nabla f(x^\star) = 0$. Inoltre, se f è strettamente convessa su R^n e se in $x^\star$ si ha $\nabla f(x^\star) = 0$, allora $x^\star$ è l'unico punto stazionario di f e costituisce anche l'unico punto di minimo globale della funzione.*

Se si considera la matrice Hessiana si ottiene una condizione necessaria del secondo ordine.

Teorema 6.4.6 (Condizione necessaria del secondo ordine)

Sia $f : R^n \to R$ due volte continuamente differenziabile in un intorno di $x^\star \in R^n$. Allora, se $x^\star$ è un punto di minimo locale non vincolato di f si ha:

(a) $\nabla f(x^\star) = 0$;

(b) $\nabla^2 f(x^*)$ *è semidefinita positiva, ossia* $y'\nabla^2 f(x^\star)y \geq 0$, per ogni $y \in R^n$.

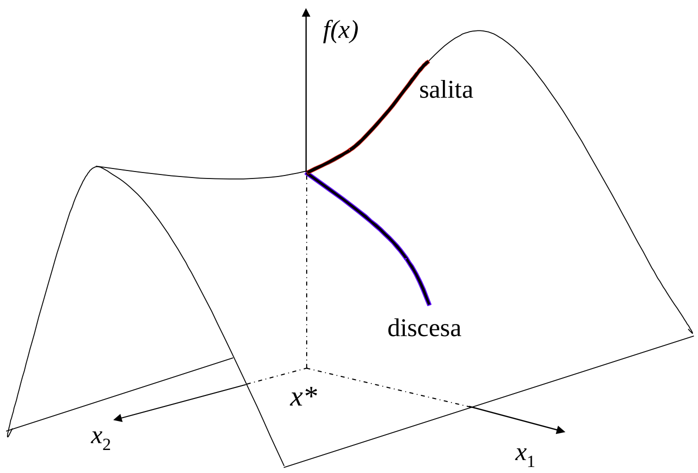

Figura 6.5: Esempio di punto di sella in $I\!R^2$.

Dim. La (a) segue dalla caratterizzazione di direzioni di discesa del secondo ordine data dal Teorema 6.2.6. □

Se tuttavia $\nabla^2 f(x^\star)$ è definita positiva, si può stabilire un risultato più forte, che riportiamo nel teorema successivo.

Teorema 6.4.7 (Condizione sufficiente del secondo ordine)

Sia $f : R^n \to R$ due volte continuamente differenziabile in un intorno di $x^* \in R^n$.

Supponiamo che valgano le condizioni:

(a) $\nabla f(x^\star) = 0$

(b) *la matrice Hessiana è definita positiva in $x^\star$, ossia:*

$$y'\nabla^2 f(x^\star) y > 0 \quad \text{per ogni} \quad y \in R^n, \quad y \neq 0.$$

Allora $x^\star$ è un punto di minimo locale stretto.

Dimostrazione. Utilizzando il Teorema di Taylor, e tenendo conto del fatto che $\nabla f(x^\star) = 0$, possiamo scrivere:

$$f(x^\star + \alpha d) = f(x^\star) + \frac{1}{2}\alpha^2 d'\nabla^2 f(x^\star) d + o(\alpha^2),$$

e quindi si ottiene:

$$\frac{f(x^\star + \alpha d) - f(x^\star)}{\alpha^2} = \frac{1}{2} d' \nabla^2 f(x^\star) d + \frac{o(\alpha^2)}{\alpha^2}$$

Quindi per valori di α piccoli, l'ultimo termine è trascurabile ($o(\alpha^2)/\alpha^2 \to 0$ per $\alpha \to 0$), per cui $f(x^\star + \alpha d) - f(x^\star) > 0$ per ogni d, il che prova l'enunciato. $\square$

Dai risultati precedenti si possono dedurre facilmente condizioni necessarie e condizioni sufficienti di massimo locale (basta infatti imporre condizioni di minimo su $-f$).

In particolare, la condizione che $x^\star$ sia *un punto stazionario*, ossia tale che $\nabla f(x^\star) = 0$ è condizione necessaria sia perché $x^\star$ sia un punto di minimo, sia perchè $x^\star$ sia un punto di massimo.

Una condizione necessaria del secondo ordine per un massimo locale è che $\nabla f(x^\star) = 0$ e $\nabla^2 f(x^\star)$ sia *semidefinita negativa*. Una condizione sufficiente di massimo locale è che $\nabla f(x^\star) = 0$ e $\nabla^2 f(x)$ sia *semidefinita negativa in un intorno di* $x^\star$. Condizione sufficiente perchè $x^\star$ sia un punto di massimo locale stretto è che $\nabla f(x^\star) = 0$ e $\nabla^2 f(x^\star)$ sia *definita negativa*.

Se risulta $\nabla f(x^\star) = 0$ e la matrice Hessiana $\nabla^2 f(x^\star)$ è *indefinita* (ossia, esistono vettori d per cui $d' \nabla^2 f(x^\star) d > 0$ e altri per cui $d' \nabla^2 f(x^\star) d < 0$) allora si può escludere che $x^\star$ sia un punto di minimo o di massimo locale e $x^\star$ viene usualmente denominato *punto di sella*.

Un punto di sella è un punto stazionario in corrispondenza al quale esistono sia direzioni *di discesa* (quando $d' \nabla^2 f(x^\star) d < 0$) sia direzioni *di salita* (quando $d' \nabla^2 f(x^\star) d > 0$).

o Se $\nabla f(x^\star) = 0$ e $\nabla^2 f(x^\star)$ è *semidefinita* (negativa o positiva) non si può determinare la natura di $x^\star$ in assenza di altre informazioni. Tuttavia se $\nabla^2 f(x^\star)$ *non* è semidefinita positiva (negativa) si può escludere che x^* sia un punto di minimo (massimo).

Una classe di particolare interesse di funzioni non lineari è quella delle funzioni quadratiche, ovvero del tipo

$$f(x) = \frac{1}{2} x^T Q x + c^T x$$

con Q matrice simmetrica $n \times n$ e $c \in I\!R^n$. Poiché la matrice hessiana di una funzione quadratica è costante è facile verificare la convessità della funzione (vedi paragrafo 3.2 del Capitolo 3). Nel caso quadratico è anche possibile formalizzare i risultati di esistenza di un minimo globale.

Vale in particolare la seguente caratterizzazione.

Teorema 6.4.8 (Minimizzazione di una funzione quadratica) *Sia* $q(x) = \frac{1}{2} x^T Q x + c^T x$, *con* Q *simmetrica e* $c \in R^n$. *Allora:*

1. $q(x)$ *ammette un punto di minimo se e solo se* Q *è semidefinita positiva ed esiste* $x^\star$ *tale che* $Qx^\star + c = 0$*;*
2. $q(x)$ *ammette un unico punto di minimo globale se e solo se* Q *è definita positiva;*
3. *se* Q *è semidefinita positiva ogni punto* $x^\star$ *tale che* $Qx^\star + c = 0$ *è un punto di minimo globale di* $q(x)$*.*

Dimostrazione. (Non in programma) Ricordiamo che risulta $\nabla q(x) = Qx + c$ e $\nabla^2 f(x) = Q$. Inoltre se risulta Q semidefinita positiva, $q(x)$ è convessa, se risulta Q definita positiva, $q(x)$ è strettamente convessa.
Assegnati $x^\star, x \in R^n$ e posto $x = x^\star + s$ si può scrivere:

$$q(x) = q(x^\star + s) = q(x^\star) + (Qx^\star + c)^T s + \frac{1}{2} s^T Q s. \tag{6.1}$$

Supponiamo ora che $Qx^\star + c = 0$ e che Q sia semidefinita positiva; in tal caso dalla (6.1) segue $q(x) \geq q(x^\star)$ per ogni $x \in R^n$.

Inversamente, se q ammette un punto di minimo $x^\star$, per il Teorema 6.4.1 deve essere $\nabla q(x^\star) = 0$ e risulta $q(x) \geq q(x^\star)$ per ogni $x \in I\!R^n$. Dalla (6.1) segue $0 \leq q(x) - q(x^\star) = \frac{1}{2}s^T Qs$ con $s = x - x^\star$, ovvero $Q \succeq 0$. Ciò prova la (1).
La (2) segue dalla (1) e dal fatto che, $s^T Qs > 0$ per ogni s se e solo se Q è definita positiva.
Infine, la (3) segue ancora dalla (6.1) perchè per ogni $x^\star$ tale che $\nabla q(x^\star) = 0$ si ha $q(x) \geq q(x^\star)$. $\square$

Illustriamo questo teorema con un esempio.

Esempio 6.4.9 Sia data la funzione

$$q(x_1, x_2) = \frac{1}{2}(\alpha x_1^2 + \beta x_2^2) - x_1.$$

Studiamo le esistenza e la natura dei punti estremali al variare dei parametri α e β.

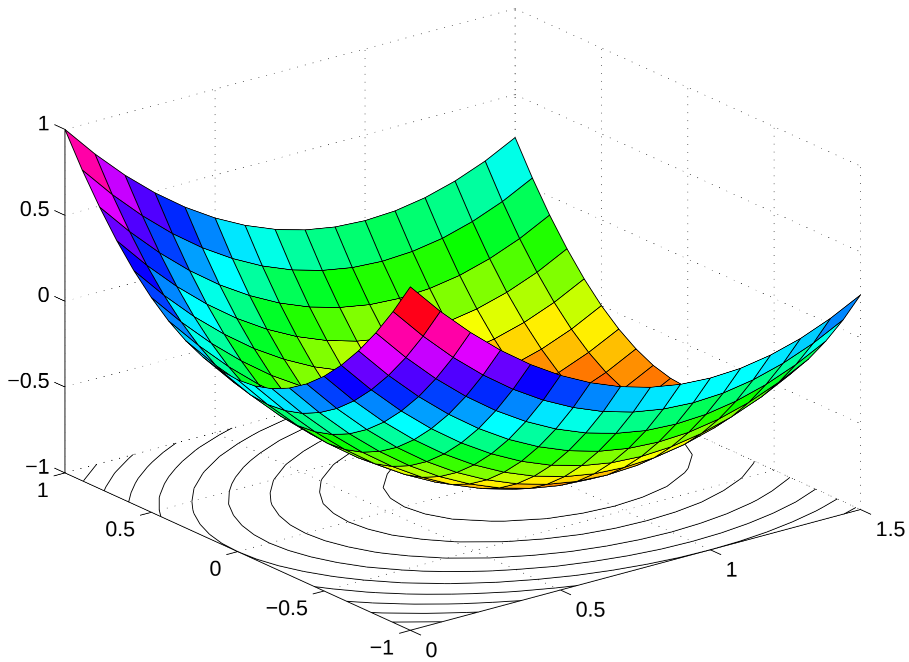

Figura 6.6: Grafico di $q(x_1, x_2)$ per $\alpha = \beta = 1$.

Scriviamo il gradiente e la matrice hessiana di q. Si ha

$$\nabla q = \begin{pmatrix} \alpha x_1 - 1 \\ \beta x_2 \end{pmatrix} \qquad \nabla^2 q = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}$$

Se $\alpha > 0$ e $\beta > 0$, allora esiste un unica soluzione al sistema $\nabla q = 0$ ed è $\left(\frac{1}{\alpha}, 0\right)^T$. Inoltre la matrice $\nabla^2 q$ è definita positiva; si tratta quindi dell'unico punto di minimo globale.
Se $\alpha = 0$ e β è qualsiasi, non esiste soluzione al sistema $\nabla q = 0$. Notare che se $\beta \geq 0$ la matrice è semidefinita positiva, ma questo non assicura l'esistenza del minimo globale.
Se $\alpha > 0$ e $\beta = 0$, esistono infinite soluzioni al sistema $\nabla q = 0$ ed è $\left(\frac{1}{\alpha}, \xi\right)^T$ con ξ qualsiasi. Inoltre la matrice $\nabla^2 q$ è semidefinita positiva; si tratta quindi di infiniti punti di minimo globale.

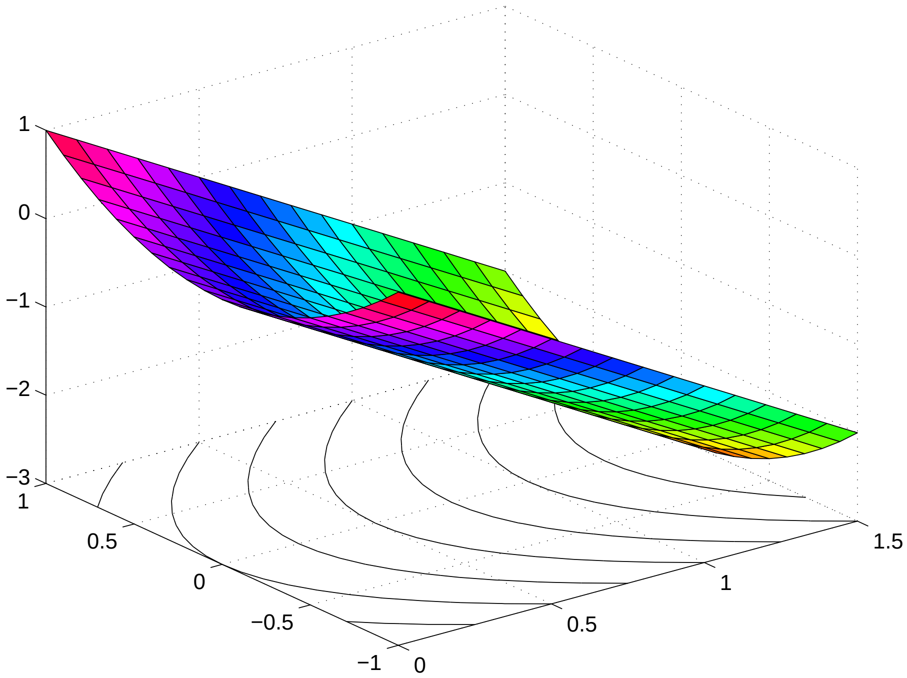

Figura 6.7: Grafico di $q(x_1, x_2)$ per $\alpha = 0$ $\beta = 1$.

Se $\alpha < 0$ e $\beta > 0$ si ha un'unica soluzione $(\frac{1}{\alpha}, 0)$. Ma la matrice hessiana è indefinita; si tratta quindi di un punto di sella.

Nel caso di $\alpha < 0$ e $\beta < 0$, allora esiste un unica soluzione al sistema $\nabla q = 0$ ed è $\left(\frac{1}{\alpha}, 0\right)^T$. Inoltre la matrice $\nabla^2 q$ è definita negativa; si tratta quindi dell'unico punto di massimo globale. $\square$

6.5 Ottimizzazione su insieme convesso generico

Consideriamo un problema di ottimo in cui in cui supponiamo che l'insieme ammissibile S sia un *insieme convesso*, ovvero

$$\begin{array}{ll} min & f(x) \\ & x \in S. \end{array} \qquad (P-CONV)$$

Possiamo caratterizzare le direzioni ammissibili di questo insieme.

Teorema 6.5.1 (Direzioni ammissibili insieme convesso) *Sia $S \subseteq \mathbb{R}^n$ un insieme convesso e sia $\bar{x}$ un qualsiasi punto di S. Allora comunque si fissi $x \in S$ tale che $x \neq \bar{x}$, la direzione $d = x - \bar{x}$ è una direzione ammissibile per S in $\bar{x}$.*

Dimostrazione. Sia $\bar{x} \in S$. Comunque preso $x \in S$, con $x \neq \bar{x}$, per la convessità di S, si ha che $(1-\beta)\bar{x} + \beta x \in S$ per ogni $\beta \in [0, 1]$ e quindi $\bar{x} + \beta(\bar{x} - x) \in S$ per ogni $\beta \in [0, 1]$. Da cui segue che la direzione $d = x - \bar{x}$ è una direzione ammissibile per S in $\bar{x}$. $\square$

È facile verificare che, inversamente, se $d \neq 0$ è una direzione ammissibile per S in $\bar{x}$ allora esiste un punto $x \in S$ e uno scalare t tale che $d = t(x - \bar{x})$.

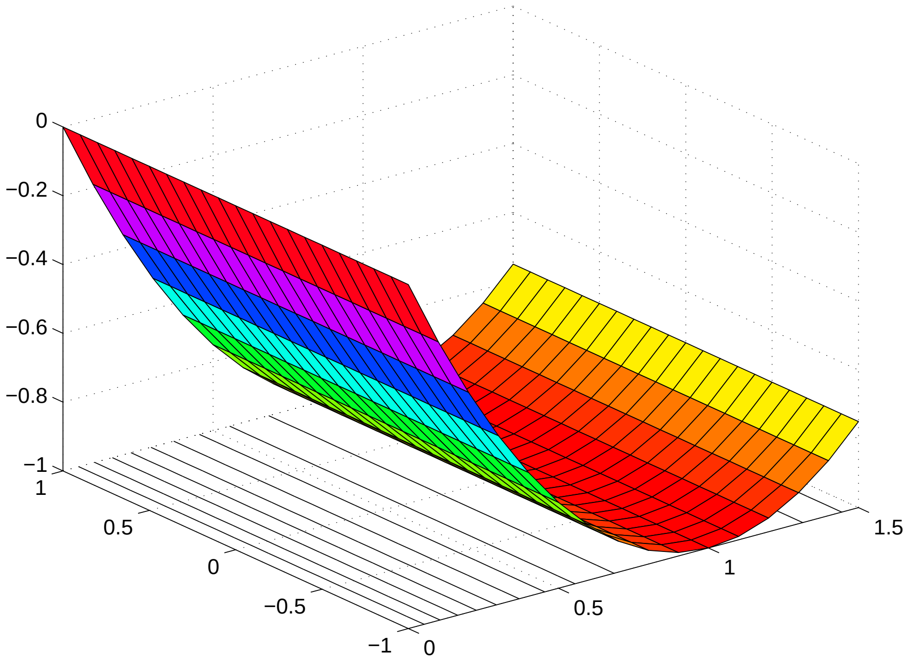

Figura 6.8: Grafico di $q(x_1, x_2)$ per $\alpha = 1$ $\beta = 0$.

Quindi utilizzando direttamente il teorema 6.3.5, otteniamo la seguente condizione necessaria.

Teorema 6.5.2 (Condizioni necessarie di minimo locale con insieme ammissibile convesso) *Sia $x^* \in S$ un punto di minimo locale del problema (P-CONV) e supponiamo che f sia continuamente differenziabile su $\mathbb{R}^n$. Allora si ha:*

$$\nabla f(x^*)^T (x - x^*) \geq 0, \quad \text{per ogni} \quad x \in S. \tag{6.2}$$

Dim. Osserviamo innanzitutto che in base al teorema 6.5.1 la direzione $d = x - x^*$ è una direzione ammissibile per f in S per ogni $x \in S$. Se esistesse $d = x - x^*$ tale che $\nabla f(x^*)^T d < 0$ per il Teorema 6.2.2 la direzione ammissibile d sarebbe di discesa ma questo è assurdo perché in base al Teorema 6.3.1 sappiamo che non può esistere una direzione ammissibile di discesa in x^*. $\square$

Nel caso che f sia convessa, si può dimostrare anche il seguente risultato

Teorema 6.5.3 (Condizione sufficiente di minimo globale) *Sia $x^* \in \mathbb{R}^n$ e supponiamo che S sia convesso. Supponiamo che f sia convessa su S. Se risulta*

$$\nabla f(x^*)^T d \geq 0 \qquad \textit{per ogni direzione ammissibile } d \in R^n \textit{ in } \ x^*, \tag{6.3}$$

allora x^* è un minimo globale del problema (P).
Se inoltre f è strettamente convessa su S, e vale (6.3) allora x^* è anche l'unico punto di minimo globale del problema (P).

Dimostrazione. La dimostrazione è per assurdo. Supponiamo che x^* non sia minimo globale, allora deve esistere un $\hat{x} \in S$ tale che $f(\hat{x}) < f(x^*)$. Poiché $\hat{x} \in S$ allora la direzione $d = \hat{x} - x^*$ è ammissibile e per l'ipotesi del Teorema deve risultare

$$\nabla f(x^*)^T (\hat{x} - x^*) \geq 0.$$

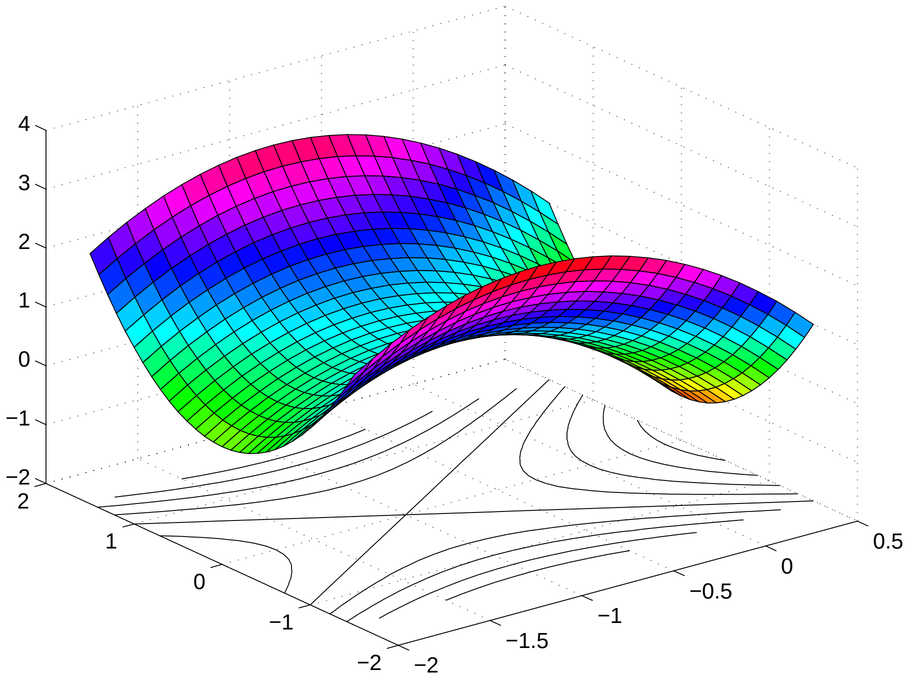

Figura 6.9: Grafico di $q(x_1, x_2)$ per $\alpha = -1$ $\beta = 1$.

Poiché la funzione è convessa si ha

$$f(\hat{x}) \geq f(x^*) + \nabla f(x^*)^T(\hat{x} - x^*).$$

Mettendo insieme le due formule precedenti (in rosso) si ottiene $f(\hat{x}) \geq f(x^*)$, che è assurdo perché abbiamo supposto che fosse $f(\hat{x}) < f(x^*)$.
Supponiamo ora che f sia strettamente convessa e che per assurdo esista un altro punto $\tilde{x}$ di minimo globale, tale che $f(\tilde{x}) = f(x^*)$. Poiché $\tilde{x} \in S$ allora la direzione $d = \tilde{x} - x^*$ è ammissibile e risulta $\nabla f(x^*)^T(\tilde{x} - x^*) \geq 0$. Poiché la funzione è strettamente convessa possiamo scrivere

$$f(\tilde{x}) > f(x^*) + \nabla f(x^*)^T(\tilde{x} - x^*) \geq f(x^*),$$

che è assurdo. □

Illustriamo la condizione espressa dal teorema con un esempio.

Esempio 6.5.4 Sia dato il problema di ottimizzazione

$$\begin{array}{ll} \min & -x_2 \\ & x_1 - x_2^2 \leq 0 \\ & x_1 + x_2 \leq 1 \\ & x_1 \geq 0 \\ & x_2 \geq 0 \end{array}$$

Si tratta della minimizzazione di una funzione lineare (quindi convessa) su un insieme non convesso (la regione ammissibile e le curve di livello della funzione obiettivo sono rappresentate in Figura 6.11.) Il gradiente della funzione obiettivo è

$$\nabla f(x) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

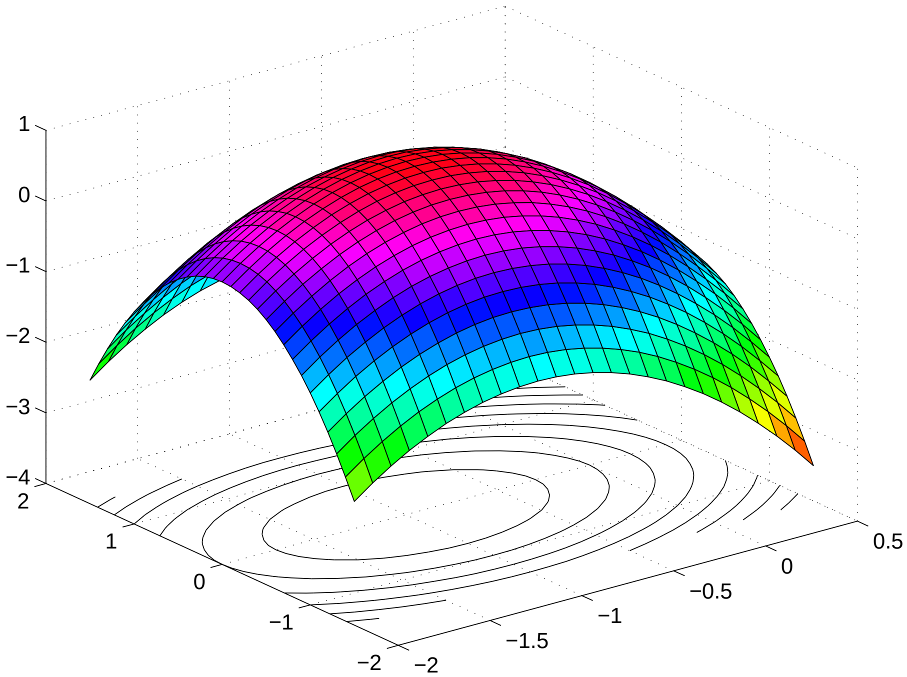

Figura 6.10: Grafico di $q(x_1, x_2)$ per $\alpha = \beta = -1$.

Consideriamo il punto $x^* = (0,\ 1)^T$. In figura 6.11 sono rappresentate le direzione ammissibili (frecce in azzurro) e il gradiente della funzione obiettivo (freccia in rosso) nel punto x^*. Il gradiente forma un angolo inferiore a 90^o con tutte le direzioni ammissibili, quindi risulta $\nabla f(x^*)^T d \geq 0$ e la condizione necessaria del teorema 6.3.5 è verificata. Inoltre si osserva che per ogni $x \in S$, le direzioni $x - x^*$ sono ammissibili, e la funzione è convessa. Quindi anche la condizione sufficiente del teorema 6.5.5 è verificata. Si può dunque concludere che x^* è minimo globale. $\square$

È banale che se f è convessa, la condizione (6.2) diviene quindi una condizione necessaria e sufficiente di minimo globale.

Teorema 6.5.5 (Condizioni necessarie e sufficienti di minimo globale nel caso convesso) *Sia S un sottoinsieme convesso di R^n e supponiamo che f sia una funzione convessa continuamente differenziabile su $I\!R^n$. Allora $x^* \in S$ è un punto di minimo globale del problema (P-CONV) se e solo se*

$$\nabla f(x^*)^T (x - x^*) \geq 0, \quad \text{per ogni} \quad x \in S. \tag{6.4}$$

Teorema 6.5.6 (Condizioni necessarie del secondo ordine di minimo locale con insieme ammissibile convesso)

Sia $x^* \in S$ un punto di minimo locale del problema (P-CONV). Se f è due volte continuamente differenziabile su un insieme aperto contenente S, risulta

$$(x - x^*)' \nabla^2 f(x^*)(x - x^*) \geq 0, \quad \text{per ogni} \quad x \in S \quad \text{tale che} \quad \nabla f(x^*)'(x - x^*) = 0. \tag{6.5}$$

Dimostrazione. In base al Teorema 6.3.1 sappiamo che non può esistere una direzione ammissibile di discesa. Se f è due volte continuamente differenziabile ed esiste $x \in S$ tale che la direzione ammissibile

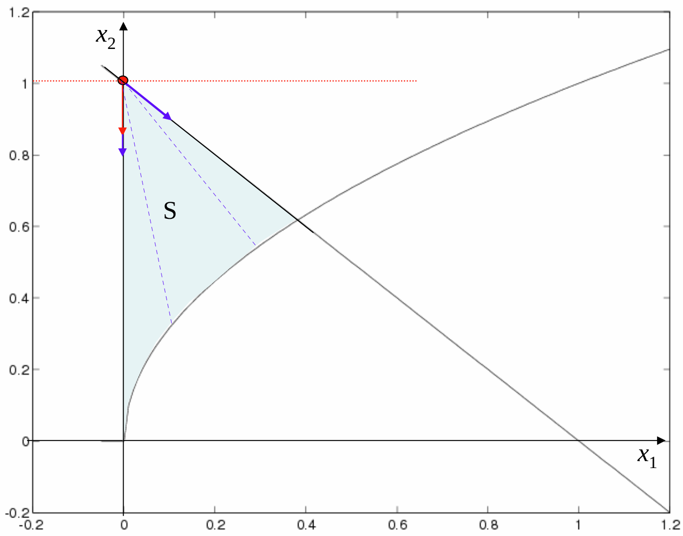

Figura 6.11: Poliedro Esempio 6.5.4.

$d = x - x^*$ soddisfi $\nabla f(x^*)'d = 0$. In base all'ipotesi di differenziabilità si ha:

$$f(x^* + td) = f(x^*) + \nabla f(x^*)'d + \frac{t^2}{2} d'\nabla^2 f(x^*)d + \beta(x^*, td), \tag{6.6}$$

dove

$$\lim_{t \to 0} \frac{\beta(x^*, td)}{t^2} = 0. \tag{6.7}$$

Poiché $x^* + td \in S$ per $t \in [0, 1]$ e poiché x^* è un punto di minimo locale deve essere

$$f(x^* + td) \geq f(x^*) \tag{6.8}$$

per ogni $t > 0$ sufficientemente piccolo e quindi, per tali valori di $t > 0$, sostituendo nella (6.8) l'espressione di $f(x^* + td)$ data dalla (6.6), dividendo per t^2 e tenendo conto dell'ipotesi $\nabla f(x^*)'d = 0$, si può scrivere:

$$\frac{1}{2} d'\nabla^2 f(x^*)d + \frac{\beta(x^*, td)}{t^2} \geq 0.$$

Passando al limite per $t \to 0^+$ e tenendo conto della (6.7) si ottiene allora la (6.5). $\square$

Consideriamo il seguente esempio in cui i vincoli sono solo limitazioni superiori ed inferiori sulle variabili.

Esempio 6.5.7 Consideriamo l'esempio 6.3.4

$$\begin{array}{ll} \min & -x_1^2 - x_2^2 \\ & -1 \leq x_1 \leq 1 \\ & -1 \leq x_2 \leq 1 \end{array}$$

L'insieme ammissibile è convesso, quindi le condizioni necessarie richiedono che in un punto x^* risulti $\nabla f(x^*)^T(x - x^*) \geq 0$ per ogni $x \in S$. In particolare in questo caso

$$\nabla f = \begin{pmatrix} -2x_1 \\ -2x_2 \end{pmatrix}$$

e deve risultare quindi

$$-2\bar{x}_1(x_1 - \bar{x}_1) - 2\bar{x}_2(x_2 - \bar{x}_2) \geq 0 \quad \text{per ogni} \quad -1 \leq x_1 \leq 1 \quad -1 \leq x_2 \leq 1.$$

Verifichiamo che le condizioni sono soddisfatte in $\bar{x} = (1,\ 1)^T$. Deve risultare

$$-2(x_1 - 1) - 2(x_2 - 1) \geq 0 \ \text{ per ogni } \ -1 \leq x_1 \leq 1 \quad -1 \leq x_2 \leq 1.$$

Il termine $x_1 - 1 \leq 0$ per punti $x \in S$ e analogamente $x_2 - 1 \leq 0$ per $x \in S$, quindi la condizione è effettivamente soddisfatta. $\square$

6.6 Ottimizzazione su un poliedro

Una classe significativa di problemi di programmazione matematica è quella in cui l'insieme ammissibile è un poliedro. Consideriamo, senza perdita di generalità, un poliedro definito come

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

in cui A è una matrice reale $m \times n$ e $b \in \mathbb{R}^m$. Il problema in esame è quindi

$$\begin{array}{ll} \min & f(x) \\ & Ax \geq b \end{array} \qquad (P-POL)$$

in cui f è una funzione continuamente differenziabile. Ci proponiamo nel seguito di particolarizzare le condizioni di ottimo già enunciate nei paragrafi precedenti per ricavare condizioni di ottimo (locale e globale) che tengano conto della struttura dell'insieme ammissibile.

Nel Capitolo 5 abbiamo caratterizzato le direzioni ammissibili per un poliedro. In particolare dato $x^* \in S$, le direzioni ammissibili sono tutti e soli i vettori che soddisfano $a_i^T d \geq 0$, per ogni $i \in I(x^*)$, dove $I(x^*) = \{i : \ a_i^T x^* = b_i\}$ è l'insieme degli indici dei vincoli attivi in x^*.
Introduciamo la seguente notazione matriciale: in x^* indichiamo con A_I la matrice $|I(x^*)| \times n$ costituita dalle righe di A con indice in $I(x^*)$, ovvero, se $A = (a_i^T)_{i=1,\dots,m}$, si ha

$$A_I = (a_i^T)_{i \in I(x^*)}.$$

Sia $x^* \in S$, le direzioni ammissibili in x^* sono tutte e sole le soluzioni del sistema lineare

$$A_I d \geq 0$$

dove A_I è la matrice $A_I = (a_i^T)_{i \in I(x^*)}$ e $I(x^*) = \{i : \ a_i^T x^* = b_i\}$.

Possiamo applicare il teorema 6.3.1 ed ottenere la seguente condizione necessaria.

Teorema 6.6.1 (Condizione necessaria di minimo locale su un poliedro) *Se x^* è un punto di minimo locale del problema (P-POL) allora risulta*

$$\nabla f(x^*)^T d \geq 0 \quad \textit{per ogni direzione } d \in R^n : \ a_i^T d \geq 0, \quad \forall i \in I(x^*) = \{i : \ a_i^T x^* = b_i\}.$$

Questa condizione può essere formulata come *non esistenza* di soluzione di una sistema lineare. Utilizzando la notazione matriciale introdotta precedentemente, possiamo scrivere la condizione necessaria espressa dal teorema 6.6.1 come segue.

Se x^* è un punto di minimo locale di (P-POL) allora *non esiste* una soluzione $d \in R^n$ al sistema di disequazioni lineare

$$\begin{aligned} &A_I d \geq 0, \\ &\nabla f(x^*)^T d < 0. \end{aligned} \tag{6.9}$$

Illustriamo questa condizione necessaria con un esempio.

Esempio 6.6.2 Consideriamo il problema

$$\begin{array}{ll} \min & (x_1 + 10)^2 + (x_2 - 12)^2 \\ & 3x_1 - 2x_2 \geq -30 \\ & 2x_1 - x_2 \geq -12 \\ & x_1 \geq 0, \\ & x_2 \geq 0 \end{array}$$

il cui poliedro è stato descritto nell'Esempio 5.2.4. Consideriamo il punto ammissibile $\bar{x} = (0,\ 12)^T$ che risulta essere minimo (vedi Figura 6.12 per le curve di livello della funzione obiettivo e la regione ammissibile).

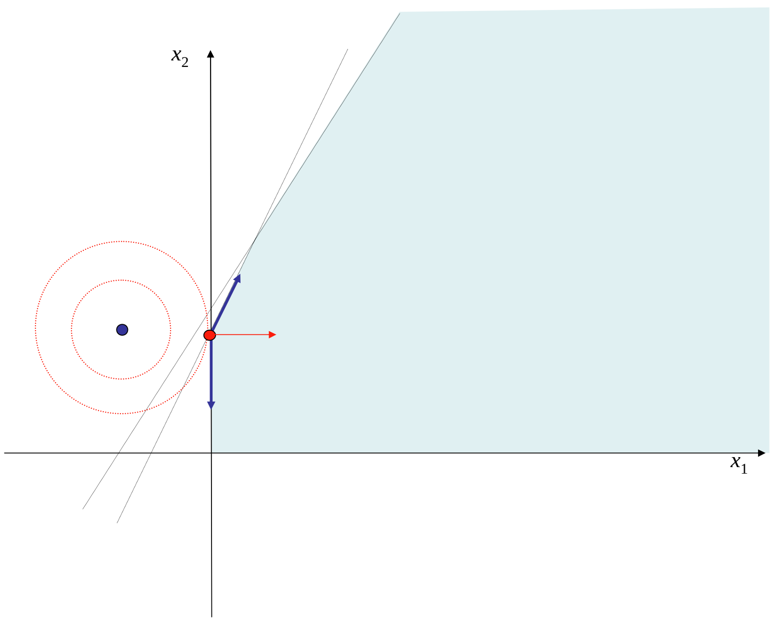

Figura 6.12: Poliedro Esempio 6.6.2.

Il sistema delle direzione ammissibili è già stato definito nell'Esempio 5.2.4. Il gradiente della funzione obiettivo

$$\nabla f(x) = \begin{pmatrix} 2(x_1 + 10) \\ 2(x_2 - 12) \end{pmatrix}$$

che in $\bar{x}$ vale $\nabla f(\bar{x}) = (20,\ 0)^T$. Il sistema (6.9) diventa

$$\begin{array}{l} 2d_1 - d_2 \geq 0 \\ d_1 \geq 0 \\ 20d_1 < 0 \end{array} \tag{6.10}$$

che non ha ovviamente soluzione (le ultime due disequazioni incompatibili). □

Ricordando che un poliedro è un insieme convesso e quindi vale la caratterizzazione delle direzione ammissibili data nel teorema 6.5.1 è possibile applicare il Teorema 6.5.5 e ottenere il seguente risultato.

Teorema 6.6.3 (Condizione necessaria e sufficiente di minimo globale su un poliedro) *Sia $f(x)$ convessa. Un punto $x^* \in S$ è un minimo globale per problema (P-POL) se e solo se*

$$\nabla f(x^*)^T d \geq 0 \quad \textit{per ogni direzione } d \in R^n: \ a_i^T d \geq 0, \quad \forall i \in I(x^*) = \{i: \ a_i^T x^* = b_i\}.$$

Anche in questo caso si può riscrivere la condizione come:

Sia f una funzione convessa. Un punto $x^* \in S$ è minimo globale di (P-POL) se e solo se *non esiste* una soluzione $d \in R^n$ al sistema di disequazioni lineare

$$\begin{array}{l} A_I d \geq 0, \\ \nabla f(x^*)^T d < 0. \end{array} \tag{6.11}$$

Consideriamo nuovamente un esempio.

Esempio 6.6.4 Consideriamo nuovamente il problema dell'Esempio 6.6.2. La funzione obiettivo è strettamente convessa. Quindi la condizione (6.11) è necessaria e sufficiente di minimo globale. Quindi il punto $(0,\ 12)^T$ che la verifica è un minimo globale.
Verifichiamo cosa succede in un punto diverso, ad esempio $\hat{x} = (0,\ 0)^T$. In $\hat{x}$ risulta $I(\hat{x}) = \{3, 4\}$ e le direzioni ammissibili soddisfano: $d_1 \geq 0,\ d_2 \geq 0$. Inoltre si ha $\nabla f(\hat{x}) = (20,\ -24)^T$, quindi il sistema (6.11) è

$$\begin{array}{l} 20d_1 - 24d_2 < 0 \\ d_1 \geq 0 \\ d_2 \geq 0 \end{array}$$

In Figura 6.13 sono rappresentate le direzioni ammissibili (tutti i vettori tra le due frecce blu) e il gradiente della funzione obiettivo (freccia rossa). È facile verificare che esistono soluzioni al sistema; in particolare sono soluzione tutti i vettori $d = (d_1,\ d_2)^T$ tali che $0 \leq d_1 < \frac{6}{5} d_2$. Quindi, come già sapevamo, il punto $\hat{x}$ non è minimo globale in quanto non è soddisfatta la condizione necessaria e sufficiente. □

Possiamo inoltre dare una condizione necessaria del secondo ordine utilizzando il teorema 6.3.5.

Teorema 6.6.5 (Condizioni necessarie del 2^o ordine di minimo locale sul poliedro) *Sia x^* un punto di minimo del Problema (P-POL). Allora risulta*

$$d^T \nabla_x^2 f(x^*) d \geq 0 \quad \textit{per ogni} \ \ d \in I\!R^n: \ a_i^T d \geq 0, \ \ i \in I(x^*) \ \ \textit{e} \ \nabla f(x^*)^T d = 0.$$

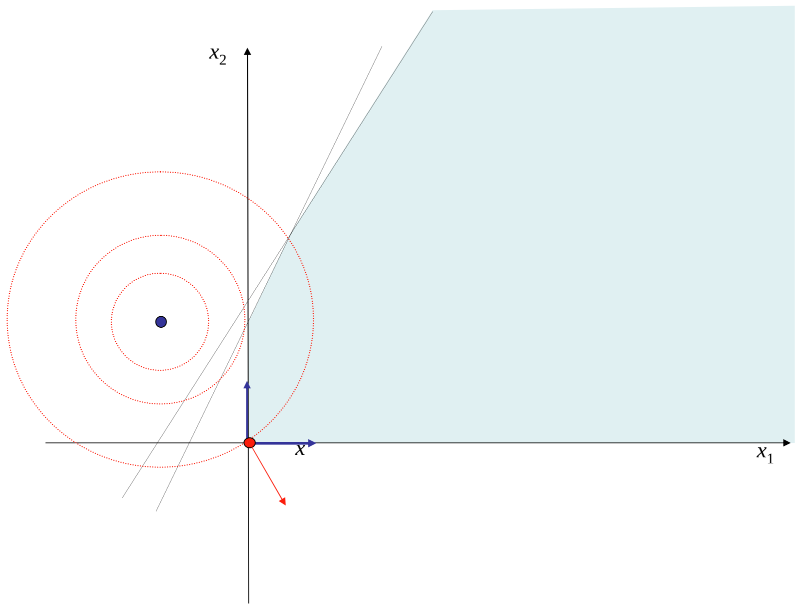

Figura 6.13: Poliedro Esempio 6.6.4.

Le considerazioni svolte nel caso di poliedro descritto da soli vincoli di disuguaglianza $Ax \geq b$, possono essere applicate anche a poliedri definiti da vincoli di uguaglianza e disuguaglianza.

Possiamo infatti utilizzare la caratterizzazione delle direzioni ammissibili data nel paragrafo 5.2. Consideriamo allora il caso piú generale di problema con vincoli in forma standard del tipo:

$$\begin{aligned} \min \quad & f(x) \\ & Ax = b \\ & x \geq 0 \end{aligned} \tag{$P-POL-ST$}$$

In questo caso, utilizzando la condizione espressa dal Teorema 5.2.6, possiamo affermare:

Se x^* è minimo globale del problema (P-POL-ST) allora *non esiste* una soluzione $d \in R^n$ al sistema di disequazioni lineare

$$\begin{aligned} & Ad = 0, \\ & d_J \geq 0 \\ & c^T d < 0. \end{aligned} \tag{6.12}$$

con $J = \{j \in \{1, \ldots, n\} : \ x_j^* = 0\}$.

Naturalmente anche in questo caso le condizioni diventano necessarie e sufficienti nel caso in cui la funzione obiettivo sia convessa.

La Programmazione Lineare Tra i problemi di minimizzazione su poliedro particolare interesse rivestono i problemi di Programmazione Lineare. Utilizzando il teorema 6.6.3, possiamo enunciare la seguente condizione che caratterizza ulteriormente (rispetto ai risultati del Capitolo 5) le soluzioni ottime di un problema di Programmazione Lineare.

Sia dato il problema di Programmazione Lineare

$$\begin{aligned} \min \quad & c^T x \\ & Ax \geq b \end{aligned} \tag{PL}$$

Un punto x^* è minimo globale se e solo se *non esiste* una soluzione $d \in R^n$ al sistema di disequazioni lineare

$$\begin{aligned} & A_I d \geq 0, \\ & c^T d < 0. \end{aligned} \tag{6.13}$$

Analogamente nel caso di problemi in forma standard (P-POL-ST), possiamo affermare:

Un punto x^* è minimo globale del problema di Programmazione Lineare

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

se e solo se *non esiste* una soluzione $d \in R^n$ al sistema di equazioni e disequazioni lineari

$$\begin{aligned} & Ad = 0, \\ & d_J \geq 0 \\ & c^T d < 0. \end{aligned} \tag{6.14}$$

con $J = \{j \in \{1, \ldots, n\} : \; x_j^* = 0\}$.

Le condizioni di "non esistenza" di soluzioni di un sistema lineare ottenute in questo paragrafo, possono essere riformulate in modo equivalente come condizioni di esistenza di un diverso sistema di equazioni. Questa riformulazione costituisce la forma classica delle condizioni di ottimo per problemi di programmazione matematica. È più facilmente utilizzabile a fine algoritmico e consente anche un'analisi di sensibilità alla variazione dei dati. Sarà descritta nel prossimo Capitolo.

Capitolo 7

Le condizioni di Karush-Kuhn-Tucker

7.1 Introduzione

In questo capitolo deriveremo le condizioni necessarie di Karush-Kuhn-Tucker (KKT) per problemi vincolati in cui S è descritto da vincoli di disuguaglianza e/o di uguaglianza lineari. Si tratta di formulare in modo alternativo le condizioni necessarie del paragrafo 6.6. In particolare, le condizioni di "non esistenza" di soluzioni di un sistema lineare ottenute nel paragrafo 6.6, possono essere riformulate in modo equivalente come condizioni di esistenza di un diverso sistema di equazioni.
A questo scopo si devono utilizzare i cosidetti *teoremi dell'alternativa*.

7.2 Teoremi dell'alternativa

I teoremi dell'alternativa consentono di ridurre il problema della *non esistenza* di soluzioni di un sistema assegnato a quello dell'*esistenza* di soluzioni di un altro sistema.

Un risultato di alternativa relativo a due assegnati sistemi (I) e (II), consiste nel dimostrare che:

Il sistema (I) ha soluzione se e solo se il sistema (II) non ha soluzione.

Tra i teoremi di alternativa per i sistemi di disequazioni lineari, uno dei più noti, e anche quello che utilizzeremo nel seguito del capitolo, è il *Lemma di Farkas* che si può enunciare nella forma seguente.

Teorema 7.2.1 (Lemma di Farkas) *Sia B matrice $p \times n$ e $c \in I\!R^n$. Il sistema*

$$Bd \geq 0 \qquad c^T d < 0 \tag{I}$$

non ha soluzione $d \in I\!R^n$ *se e solo se il sistema*

$$B^T u = c \qquad u \geq 0 \tag{II}$$

ha soluzione $u \in I\!R^p$.

Dimostrazione. [**(II) ha soluzione $\longrightarrow$ (I) non ha soluzione.**] Sia $\bar{u}$ soluzione del sistema (I) e supponiamo per assurdo che esista una soluzione $\bar{d}$ del sistema (II), ovvero che la coppia $\bar{u}, \bar{d}$ soddisfi:

$$\begin{array}{ll} B\bar{d} \geq 0 & c^T \bar{d} < 0, \\ B^T \bar{u} = c & \bar{u} \geq 0. \end{array}$$

Allora si può scrivere:

$$B\bar d \geq 0 \overset{\bar u \geq 0}{\Longrightarrow} \bar u^T B \bar d \geq 0 \overset{B^T \bar u = c}{\Longrightarrow} c^T \bar d \geq 0,$$

che contraddice l'ipotesi che $\bar d$ soddisfi $c^T \bar d < 0$.

[(I) non ha soluzione $\longrightarrow$ (II) ha soluzione.] La dimostrazione di questa implicazione è in due parti. Si dimostra preliminarmente che se (I) non ha soluzione allora esiste un vettore $u \in R^p$ tale ch $B^T u = c$. Successivamente dimostreremo che $u \geq 0$.

Se (I) non ha soluzione, allora in particolare non esiste una soluzione nemmeno al sistema di equazioni lineari

$$Bd = 0 \qquad c^T d = -1 \qquad [(I)eq].$$

Il sistema [(I)eq] si può scrivere in forma matriciale come segue:

$$\begin{pmatrix} B \\ c^T \end{pmatrix} d = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

È noto che tale sistema non ha soluzione se e solo se

$$\text{rango}\begin{pmatrix} B \\ c^T \end{pmatrix} \neq \text{rango}\begin{pmatrix} B & 0 \\ c^T & -1 \end{pmatrix}$$

e quindi se:

$$\text{rango}\begin{pmatrix} B & 0 \\ c^T & -1 \end{pmatrix} = \text{rango}\begin{pmatrix} B \\ c^T \end{pmatrix} + 1. \tag{7.1}$$

D'altra parte, l'ultima riga $(c^T \ \ -1)$ è linearmente indipendente dalle righe di $(B \ 0)$ e quindi:

$$\text{rango}\begin{pmatrix} B & 0 \\ c^T & 1 \end{pmatrix} = \text{rango}\begin{pmatrix} B & 0 \end{pmatrix} + 1 = \text{rango}(B) + 1. \tag{7.2}$$

Dalle (7.1) (7.2), tenendo conto del fatto che il rango di una matrice è eguale al rango della trasposta, si ottiene

$$\text{rango}\begin{pmatrix} B^T & c \end{pmatrix} = \text{rango}\begin{pmatrix} B \\ c^T \end{pmatrix} = \text{rango}(B)$$

e ciò implica che il sistema $B^T u = c$ ha soluzione, ovvero esiste una rappresentazione di c del tipo:

$$c = \sum_{i=1}^{p} u_i b_i. \tag{7.3}$$

Dimostriamo ora che $u \geq 0$.

La dimostrazione è per induzione sul numero di disequazioni che compongono il sistema, ovvero sul numero di righe p della matrice B.

Dimostriamo innanzitutto che la tesi è vera per $p = 1$ e quindi supponiamo che

$$\nexists\, d \in R: \quad c^T d < 0,\ b_1^T d \geq 0.$$

Quindi per la (7.3) risulta $c = ub_1$ con $u \in R$. Se $c = 0$, il risultato è ovvio in quanto si può assumere $u = 0$. Se $c \neq 0$ deve anche essere $b_1 \neq 0$ e possiamo considerare il vettore $\bar d = b_1 \neq 0$. Ne segue che $b_1^T \bar d = \|b_1\|^2 > 0$ e quindi, per ipotesi, deve risultare $c^T \bar d \geq 0$. Si ottiene quindi

$$c^T \bar d = (ub_1)^T (b_1) = u\|b_1\|^2 \geq 0,$$

che implica $u \geq 0$.

Supponiamo ora che il risultato sia vero per una matrice con $p-1$ righe e dimostriamo che vale per una matrice con p righe. Quindi supponiamo che:

$$\text{non esista } d \in R^n \text{ tale che } Bd \geq 0,\ c^T d < 0 \tag{7.4}$$

con B matrice $p \times n$.

Sappiamo che esiste $u \in R^p$ tale che vale la (7.3). Tra tutti i possibili u per cui vale la (7.3) determiniamo un vettore $\bar{u}$ con il massimo numero di componenti non negative e indichiamo con s il numero di componenti non negative di $\bar{u}$. Riordiniamo le componenti di $\bar{u}$ e conseguentemente le colonne di B in modo che le componenti non negative siano le prime s, ovvero risulti

$$\bar{u}_1, \ldots, \bar{u}_s \geq 0, \qquad \bar{u}_{s+1}, \ldots, \bar{u}_p < 0.$$

Allora possiamo scrivere

$$c = \sum_{i=1}^{s} \bar{u}_i b_i + \sum_{i=s+1}^{p-1} \bar{u}_i b_i + \bar{u}_p b_p.$$

Si tratta di dimostrare che deve essere $s = p$. Per assurdo supponiamo che $s < p$.
Possiamo scrivere

$$c = \hat{c} + \sum_{i=s+1}^{m-1} \bar{u}_i b_i, \tag{7.5}$$

avendo posto

$$\hat{c} = \sum_{i=1}^{s} \bar{u}_i b_i + \bar{u}_p b_p \tag{7.6}$$

La dimostrazione procede in due passi principali. Si dimostra inizialmente che

[Affermazione 1] Se $\nexists\ d \in R^n : \begin{array}{l} Bd \geq 0 \\ c^T d < 0 \end{array} \implies \nexists\ d \in R^n : \begin{array}{l} Bd \geq 0 \\ \hat{c}^T d < 0 \end{array}$

Successivamente che

[Affermazione 2]

Se $\nexists\ d \in R^n : \begin{array}{l} Bd \geq 0 \\ \hat{c}^T d < 0 \end{array} \implies \nexists\ d \in R^n : \begin{array}{ll} b_i^T d \geq 0, & i = 1, \ldots, p-1, \\ \hat{c}^T d < 0. & \end{array}$

Se [Affermazione 1] e [Affermazione 2] sono vere, come conseguenza dell'ipotesi 7.4, si ha che

$$\nexists\ d \in R^n \text{ tale che } b_i^T d \geq 0 \quad i = 1, \ldots, p-1 \quad \hat{c}^T d < 0. \tag{7.7}$$

Ma poiché il sistema (7.7) è di dimensione $p-1$, soddisfa l'ipotesi induttiva e quindi esiste un $\hat{u} \in R^{p-1}$ tale che

$$\hat{c} = \sum_{i=1}^{p-1} \hat{u}_i b_i, \quad \hat{u}_i \geq 0,\ i = 1, \ldots, p-1.$$

Sostituendo questa espressione di $\hat{c}$ nella definizione (7.6) di c, si ottiene:

$$\begin{aligned} c = \quad & \hat{c} + \sum_{i=s+1}^{p-1} \bar{u}_i b_i = \sum_{i=1}^{p-1} \hat{u}_i b_i + \sum_{i=s+1}^{m-1} \bar{u}_i b_i \\ = \quad & \sum_{i=1}^{s} \hat{u}_i b_i + \sum_{i=s+1}^{p-1} (\bar{u}_i + \hat{u}_i) b_i \\ = \quad & \sum_{i=1}^{s} \hat{u}_i b_i + \sum_{i=s+1}^{p-1} (\bar{u}_i + \hat{u}_i) b_i + 0 \cdot b_p \end{aligned}$$

Quindi c risulta essere la combinazione lineare dei vettori b_i con $s+1$ coefficienti non negativi $\hat{u}_1, \ldots, \hat{u}_s, \hat{u}_p = 0$. Ma questo è assurdo perché per ipotesi avevamo supposto che s fosse il massimo numero di coefficienti non negativi nella definizione di c.

Si tratta quindi di dimostrare che valgono le [Affermazione 1] e [Affermazione 2].

Dimostrazione [Affermazione 1]. Sia d tale che $Bd \geq 0$, e quindi, per ipotesi $c^T d \geq 0$; moltiplicando scalarmente la (7.5) per d si ottiene

$$0 \leq c^T d = \hat{c}^T d + \sum_{i=s+1}^{p-1} \bar{u}_i \overbrace{b_i^T d}^{\geq 0} \quad \text{per ogni } d: \ b_i' d \geq 0 \quad i = 1, \ldots, p.$$

Poiché $\bar{u}_{s+1}, \ldots, \bar{u}_{p-1} < 0$, la $\displaystyle\sum_{i=s+1}^{p-1} \bar{u}_i b_i^T d \leq 0$ e quindi deve essere necessariamente:

$$\hat{c}^T d \geq 0 \quad \text{per ogni } d: \ Bd \geq 0.$$

L'[Affermazione 1] è cosí dimostrata.

Dimostrazione [Affermazione 2]. Per definizione (7.6), $\hat{c}$ è la combinazione lineare dei vettori b_i con coefficienti $\bar{u}_1, \ldots, \bar{u}_s, 0, \ldots, 0, \bar{u}_p$ tutti non negativi tranne l'ultimo $\bar{u}_p$. La dimostrazione è per assurdo, quindi supponiamo che non sia vera, cioè che esista $\bar{d} \in R^n$ tale che $b_i^T \bar{d} \geq 0$ per $i = 1, \ldots, p-1$ e $\hat{c}^T \bar{d} < 0$. Per definizione di $\hat{c}$ possiamo scrivere

$$\hat{c}^T \bar{d} = \sum_{i=1}^{s} \bar{u}_i b_i^T \bar{d} + \bar{u}_p b_p^T \bar{d}.$$

Se $b_i^T \bar{d} \geq 0$ per $i = 1, \ldots, p-1$, allora $\displaystyle\sum_{i=1}^{s} \bar{u}_i b_i^T d \geq 0$. D'altra parte $\bar{u}_p < 0$ e quindi $\hat{c}^T \bar{d} < 0$ se e solo se risulta $b_p^T \bar{d} > 0$. Quindi $\bar{d}$ è un vettore tale che

$$b_i^T \bar{d} \geq 0 \quad i = 1, \ldots, p-1 \qquad b_p^T \bar{d} > 0,$$

ma allora per ipotesi deve essere $\hat{c}^T \bar{d} \geq 0$ che contraddice $\hat{c}^T \bar{d} < 0$. Quindi anche l'[Affermazione 2] è dimostrata. □

7.3 Le condizioni di Karush-Kuhn-Tucker

Consideriamo il problema di ottimizzazione su un poliedro e senza perdita di generalità ci riferiamo inizialmente ad un poliedro con vincoli di disuguaglianza. Il problema in considerazione è quindi:

$$\begin{aligned} \min \quad & f(x) \\ & Ax \geq b \end{aligned} \qquad (P-POL)$$

Nel paragrafo 6.6 abbiamo derivato le condizioni necessarie di ottimo:

Se x^* è un punto di minimo locale allora *non esiste* una soluzione $d \in R^n$ al sistema di disequazioni lineare

$$\begin{aligned} &A_I d \geq 0, \\ &\nabla f(x^*)^T d < 0. \end{aligned} \tag{7.8}$$

dove A_I è la matrice $A_I = (a_i^T)_{i \in I(x^*)}$ di dimensione $|I(x^*)| \times n$ e $I(x^*) = \{i : \; a_i^T x^* = b_i\}$.

Identificando il sistema (7.8) con il sistema (I) del Lemma di Farkas, e utilizzando il Lemma di Farkas, possiamo formulare le condizioni 7.8 come segue.

Non esiste una soluzione $d \in R^n$ al sistema di disequazioni lineare (7.8) se e solo se *esiste* una soluzione del sistema

$$\begin{aligned} &A_I^T u = \nabla f(x^*), \\ &u \geq 0. \end{aligned} \tag{7.9}$$

dove u è un vettore di dimensione $|I(x^*)|$.

Esempio 7.3.1 Consideriamo il sistema (6.10) ottenuto nell'esempio 6.6.2:

$$\begin{aligned} &2d_1 - d_2 \geq 0 \\ &d_1 \geq 0 \\ &20d_1 < 0 \end{aligned}$$

A cui corrispondono

$$A_I = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix} \qquad \nabla f(x^*) = (20, \; 0)^T.$$

Il sistema (6.10) non ammette soluzione. Applicando il Lemma di Farkas deve esistere un vettore $u \in I\!R^2$ tale che $u \geq 0$ e

$$\begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}^T \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 20 \\ 0 \end{pmatrix}$$

Il sistema si scrive

$$\begin{aligned} &2u_1 + u_2 = 20 \\ &-u_1 = 0 \\ &u_1 \geq 0, \; u_2 \geq 0 \end{aligned}$$

che ammette la soluzione $u_1 = 0$, $u_2 = 20$. $\square$

Utilizzando il Lemma di Farkas, possiamo stabilire le condizioni di ottimo per (P-POL), note come *condizioni di Karush-Kuhn-Tucker* (KKT), nella forma seguente.

Teorema 7.3.2 (Condizioni di Karush-Kuhn-Tucker per (P-POL)) *Sia x^* un punto di minimo locale del problema (P-POL). Allora esiste un vettore $\lambda^* \in R^m$ tale che risultino soddisfatte le condizioni seguenti:*

(i) $Ax^* \geq b$,

(ii) $\nabla f(x^*) - A^T \lambda^* = 0$,

(iii) $\lambda^* \geq 0$,

(iv) $\lambda_i^*(b_i - a_i^T x^*) = 0$ *per* $i = 1, \ldots, m$.

Dimostrazione. Sia x^* un punto di minimo locale; poiché x^* deve essere ammissibile vale la (i). Inoltre non deve ammettere soluzione il sistema (7.8). Abbiamo già osservato che, identificando il sistema (7.8) con il sistema (I) del Lemma di Farkas, possiamo affermare che esiste una soluzione u del sistema (7.9). Sia $u_i^* \geq 0$, $i \in I(x^*)$ una soluzione del sistema precedente e definiamo un vettore $\lambda^* \in R^m$ ponendo:

$$\lambda_i^* = \begin{cases} u_i^* & \text{per } i \in I(x^*) \\ 0 & \text{per } i \notin I(x^*). \end{cases} \tag{7.10}$$

Risulta ovviamente $\lambda^* \geq 0$ ed è soddisfatta la (iii). Inoltre, la (iv) è una conseguenza immediata della definizione (7.10) di λ^*. Abbiamo infatti che

$$\lambda_i^*(b_i - a_i^T x^*) = \begin{cases} u_i^*(b_i - a_i^T x^*) = 0 & \text{per } i \in I(x^*) \\ 0(b_i - a_i^T x^*) = 0 & \text{per } i \notin I(x^*). \end{cases}$$

Infine, ricordando che la (7.9) può essere scritta $\nabla f(x^*) = \displaystyle\sum_{i \in I(x^*)} u_i^* a_i$ possiamo scrivere

$$\nabla f(x^*) = \sum_{i \in I(x^*)} u_i^* a_i + \sum_{i \notin I(x^*)} 0 \cdot a_i = A^T \lambda^*.$$

È quindi soddisfatta la (ii). $\square$

Il vettore $\lambda \in I\!R^m$ si dice *moltiplicatore di Lagrange (generalizzato)* relativo ai vincoli $Ax \geq b$.

La condizione (iv) è nota come *condizione di complementarità* ed esprime il fatto che in un punto di minimo locale deve essere nullo, per ciascun vincolo, il prodotto $\lambda_i^*(b_i - {a_i^*}^T x^*)$, per cui, se il vincolo non è attivo, si deve annullare il corrispondente moltiplicatore.
Tenuto conto che $\lambda^* \geq 0$ e $Ax^* - b \geq 0$ la condizione (iv) si può scrivere equivalentemente come segue.

La condizione di complementarità è

$$\lambda^{*T}(b - Ax^*) = \sum_{i=1}^{m} \lambda_i^*(b_i - a_i^T x^*) = 0.$$

Le condizioni enunciate si possono esprimere in una forma equivalente facendo riferimento alla *funzione Lagrangiana.*

La funzione Lagrangiana per il problema (P-POL) è definita come

$$L(x, \lambda) = f(x) + \lambda^T (b - Ax)$$

Indicando con $\nabla_x L(x, \lambda) = \nabla f(x) - A^T \lambda$ il gradiente di L rispetto a x, la condizione (ii) esprime il fatto che nel punto (x^*, λ^*) deve annullarsi il gradiente della funzione Lagrangiana rispetto a x, per cui la coppia (x^*, λ^*) costituisce un *punto stazionario* della funzione Lagrangiana. Possiamo allora riscrivere le condizioni di KKT come segue

(Condizioni di Karush-Kuhn-Tucker)

Sia x^* un punto di minimo locale del problema (P-POL). Allora esiste un vettore $\lambda^* \in I\!R^m$ tale che risultino soddisfatte le condizioni seguenti:

(i) $Ax^* \geq b$, (ammissibilità)

(ii) $\nabla_x L(x^*, \lambda^*) = 0$ (stazionarietà),

(iii) $\lambda^* \geq 0$,

(iv) ${\lambda^*}^T(b - Ax^*) = 0$ (complementarità).

I punti candidati ad essere minimo di un problema del tipo (P-POL) possono essere determinati risolvendo le condizioni di KKT. Introduciamo allora la seguente definizione.

Un punto $\bar{x}$ è detto punto di KKT del problema (P-POL) se esiste $\bar{\lambda} \in I\!R^m$ tale che valgano le condizioni:

(i) $A\bar{x} \geq b$,

(ii) $\nabla L(\bar{x}, \bar{\lambda}) = 0$,

(iii) $\bar{\lambda} \geq 0$,

(iv) $\bar{\lambda}^T(b - A\bar{x}) = 0$.

I possibili candidati ad essere minimo locale sono quindi i punti di KKT.

Esempio 7.3.3 Consideriamo il problema

$$\begin{array}{ll} \max & x_2^2 + \dfrac{1}{2}x_1 \\ & x_1 - x_2 \leq 1 \\ & \frac{1}{2}x_1 + x_2 \leq 1 \\ & x_1 \geq 0 \end{array}$$

Determiniamo tutti punti di KKT. Mettiamo innanzitutto il problema nella forma (P-POL).

$$\begin{array}{ll} \min & -x_2^2 - \dfrac{1}{2}x_1 \\ & -x_1 + x_2 \geq -1 \\ & -\frac{1}{2}x_1 - x_2 \geq -1 \\ & x_1 \geq 0 \end{array}$$

La funzione Lagrangiana per questo problema è:

$$L(x, \lambda) = -x_2^2 - \frac{1}{2}x_1 + \lambda_1(-1 + x_1 - x_2) + \lambda_2(-1 + \frac{1}{2}x_1 + x_2) - \lambda_3 x_1$$

Le condizioni di KKT sono:

(i) (ammissibilità)

$$\begin{array}{l} -x_1 + x_2 \geq -1 \\ -\dfrac{1}{2}x_1 - x_2 \geq -1 \\ x_1 \geq 0 \end{array}$$

(ii) (stazionarietà)

$$-\frac{1}{2}+\lambda_1+\frac{1}{2}\lambda_2-\lambda_3=0$$
$$-2x_2-\lambda_1+\lambda_2=0$$

(iii) (non negatività)

$$\lambda_1,\lambda_1,\lambda_3\geq 0$$

(iv) (complementarità)

$$\lambda_1(-1+x_1-x_2)=0$$
$$\lambda_2(-1+\frac{1}{2}x_1+x_2)=0$$
$$\lambda_3 x_1=0$$

Determiniamo i punti di KKT esaminando tutti i casi possibili. Dalla condizione $\lambda_3 x_1=0$ si ottiene che

$$0=\lambda_3 x_1=\begin{cases} x_1=0 \xmapsto{(i)} -1\leq x_2\leq 1 \begin{cases} x_2=-1 & \xmapsto{(\mathrm{iv})_2} \lambda_2=0, \xmapsto{(\mathrm{ii})_2} \lambda_1=2 \xmapsto{(\mathrm{ii})_3} \lambda_3=\frac{1}{2} \\ x_2=1 & \xmapsto{(\mathrm{iv})_1} \lambda_1=0, \xmapsto{(\mathrm{ii})_2} \lambda_2=2, \xmapsto{(\mathrm{ii})_3} \lambda_3=\frac{3}{2} \\ -1<x_2<1 & \xmapsto{(\mathrm{iv})} \lambda_1=\lambda_2=0, \xmapsto{(\mathrm{ii})_2} x_2=0, \xmapsto{(\mathrm{ii})_3} \lambda_3=-\frac{1}{2} \end{cases} \\ \lambda_3=0 \quad \begin{cases} \lambda_2=0 \xmapsto{(\mathrm{ii})_1} \lambda_1=\frac{1}{2}, \xmapsto{(\mathrm{ii})_2} x_2=-\frac{1}{4}, \xmapsto{(\mathrm{iv})_1} x_1=\frac{3}{4} \\ \lambda_2>0 \begin{cases} \lambda_1>0 \xmapsto{(\mathrm{iv})_{1,2}+(\mathrm{i})} x_1=\frac{4}{3}, x_2=\frac{1}{3}, \xmapsto{(\mathrm{ii})_{1,2}} \lambda_2=\frac{7}{9}\ \lambda_1=\frac{1}{9} \\ \lambda_1=0 \xmapsto{(\mathrm{ii})_1} \lambda_2=1, \xmapsto{(\mathrm{ii})_2} x_2=\frac{1}{2}, \xmapsto{(\mathrm{iv})_2} x_1=1. \end{cases} \end{cases} \end{cases}$$

Quindi abbiamo trovato 5 punti di KKT

1. $(0,\ -1)^T$ con moltiplicatori $\lambda=(2\ ,0,\ 3/2)^T$ con valore della funz. obiettivo $f=-1$;
2. $(0,\ 1)^T$ con moltiplicatori $\lambda=(0\ ,2,\ 1/2)^T$ con valore della funz. obiettivo $f=-1$;
3. $(3/4,\ -1/4)^T$ con moltiplicatori $\lambda=(1/2\ ,0,\ 0)^T$ con valore della funz. obiettivo $f=-7/16$;
4. $(4/3,\ 1/3)^T$ con moltiplicatori $\lambda=(1/9\ ,7/9,\ 0)^T$ con valore della funz. obiettivo $f=-7/9$;
5. $(1,\ 1/2)^T$ con moltiplicatori $\lambda=(0\ ,1,\ 0)^T$ con valore della funz. obiettivo $f=-3/4$;

Poiché l'insieme ammissibile è compatto, esiste un punto di minimo globale e si trova tra i punti di KKT determinati so pra. Si tratta di scegliere quello con valore della funzione obiettivo minore: si tratta dei due punti $(0,\ 1)$ e $(0,\ -1)$.

□

Ricordando che la condizione (7.8) diventa necessaria e sufficiente di ottimo globale nell'ipotesi che f sia convessa, otteniamo che anche le condizioni di KKT sono ecessaria e sufficiente di ottimo globale nell'ipotesi che f sia convessa. Vale infatti il teorema seguente.

Teorema 7.3.4 (Condizioni necessarie e sufficienti di ottimo globale)

Sia f una funzione convessa con ∇f continuo in $\mathbb{R}^n$. Un punto x^* è minimo globale di (P-POL) se e solo se esiste $\lambda^*\in\mathbb{R}^m$ tale che valgano le condizioni:

(i) $Ax^* \geq b$,

(ii) $\nabla f(x^*) - A^T\lambda^* = 0$,

(iii) $\lambda^* \geq 0$,

(iv) ${\lambda^*}^T(b - Ax^*) = 0$.

Se inoltre f è strettamente convessa e valgono le condizioni precedenti, allora il punto x^* è l'unico punto di minimo globale di f su S.

Esempio 7.3.5 Consideriamo il problema dell'Esempio 6.6.2

$$\begin{array}{ll} \min & (x_1+10)^2 + (x_2-12)^2 \\ & 3x_1 - 2x_2 \geq -30 \\ & 2x_1 - x_2 \geq -12 \\ & x_1 \geq 0, \\ & x_2 \geq 0 \end{array}$$

Si tratta di un problema convesso. La funzione Lagrangiana per questo problema è la funzione $L : I\!R^2 \times I\!R^4 \to I\!R$ definita come:

$$L(x,\lambda) = (x_1+10)^2 + (x_2-12)^2 + \lambda_1(-30 - 3x_1 + 2x_2) + \lambda_2(-12 - 2x_1 + x_2) - \lambda_3 x_1 - \lambda_4 x_2$$

Le condizioni di KKT in un punto x ammissibile richiedono che valga

$$\nabla_x L(x,\lambda) = \begin{pmatrix} 2(x_1+10) - 3\lambda_1 - 2\lambda_2 - \lambda_3 \\ 2(x_2-12) + 2\lambda_1 + \lambda_2 - \lambda_4 \end{pmatrix} = 0$$

con

$$\begin{aligned} \lambda &\geq 0 \\ \lambda_1(-30 - 3x_1 + 2x_2) &= 0 \\ \lambda_2(-12 - 2x_1 + x_2) &= 0 \\ \lambda_3 x_1 &= 0 \\ \lambda_4 x_2 &= 0 \end{aligned}$$

Possiamo facilmente verificare che il punto ammissibile $x = (0,0)^T$ non soddisfa le condizioni di KKT. Infatti dalle condizioni di complementaritá si ottiene che $\lambda_2 = \lambda_1 = 0$. Sostituendo in $\nabla_x L(x,\lambda) = 0$ si ha

$$\begin{aligned} 20 - \lambda_3 &= 0 \\ -24 - \lambda_4 &= 0 \end{aligned}$$

da cui si ottiene il valore $\lambda_4 = -24 < 0$ che non è accettabile.

Consideriamo invece il punto $(0,12)^T$. In questo caso dalle condizioni di complementaritá si ottiene che $\lambda_4 = \lambda_1 = 0$. Sostituendo in $\nabla_x L(x,\lambda) = 0$ si ha

$$\begin{aligned} 20 - 2\lambda_2 - \lambda_3 &= 0 \\ \lambda_2 &= 0 \end{aligned}$$

da cui si ottiene $\lambda_2 = 0$ e $\lambda_3 = 20 > 0$. Quindi il punto $(0,12)^T$ soddisfa le condizioni di KKT. □

Possiamo inoltre enunciare anche le condizioni necessarie del secondo ordine

Teorema 7.3.6 (Condizioni necessarie del 2^o ordine di minimo locale sul poliedro) *Sia x^* un punto di minimo del Problema (P-POL). Allora esiste un vettore $\lambda^* \in R^m$ tale che risultino soddisfatte le condizioni di KKT ed inoltre risulta*

$$d^T \nabla_x^2 L(x^*,\lambda^*) d \geq 0 \quad \textit{per ogni} \;\; d \in I\!R^n : \;\; a_i^T d \geq 0, \;\; i \in I(x^*) \quad e \;\; \nabla f(x^*)^T d = 0.$$

La dimostrazione segue banalmente osservando che $\nabla_x^2 L(x^*, \lambda^*) = \nabla^2 f(x^*)$ e quindi si tratta di una semplice "riscrittura" del Teorema 6.6.5.

Dalle condizioni di KKT per problemi nella forma (P-POL) si possono anche derivare direttamente le condizioni di KKT per un problema con poliedro in forma standard

$$\begin{array}{ll} \min & f(x) \\ & Ax = b \\ & x \geq 0 \end{array} \qquad (P - ST)$$

Infatti, il sistema di equazioni $Ax = b$ si può equivalentemente scrivere come $Ax \geq b$ e $-Ax \geq -b$, ovvero (P-ST) si scrive:

$$\begin{array}{ll} \min & f(x) \\ & Ax \geq b \\ & -Ax \geq -b \\ & x \geq 0 \end{array} \qquad (P - ST)_2$$

La funzione Lagrangiana per (P-ST)$_2$ è:

$$L(x, \lambda, \mu) = f(x) + \mu_1^T(b - Ax) + \mu_2^T(-b + Ax) - \lambda^T x$$

con $\mu_1, \mu_2 \in \mathbb{R}^m$ e $\lambda \in \mathbb{R}^n$. Ovvero si ottiene

$$L(x, \lambda, \mu) = f(x) + (\mu_1 - \mu_2)^T(b - Ax) - \lambda^T x.$$

Le condizioni di KKT richiedono che sia

$$\nabla_x L = \nabla f(x) - A^T(\mu_1 - \mu_2) - \lambda,$$

con $\lambda \geq 0$, e $\mu_1, \mu_2 \geq 0$. Se si pone $\mu = \mu_1 - \mu_2$, si ottiene che μ può assumere qualunque valore (positivo, negativo, nullo) in quanto differenza di due quantità non negative.
Introduciamo allora la funzione Lagrangiana per il problema (P-ST)

La funzione Lagrangiana per il problema (P-ST) è definita come

$$L(x, \lambda, \mu) = f(x) + \mu^T(Ax - b) - \lambda^T x$$

Si derivano facilmente le condizioni di ottimo. Abbiamo allora ottenuto le condizioni di KKT per problemi in forma standard.

Teorema 7.3.7 (Condizioni di Karush-Kuhn-Tucker per (P-ST)) *Sia x^* un punto di minimo locale del problema (P-ST). Allora esistono un vettore $\lambda^* \in \mathbb{R}^n$ e $\mu^* \in \mathbb{R}^m$ tale che risultino soddisfatte le condizioni seguenti:*

(i) $Ax^* = b$, $x^* \geq 0$ *(ammissibilità)*,

(ii) $\nabla_x L(x^*, \lambda^*, \mu^*) = \nabla f(x^*) + A^T \mu^* - \lambda^* = 0$ *(stazionarietà)*,

(iii) $\lambda^* \geq 0$,

(iv) ${\lambda^*}^T x^* = 0$ *(complementarità)*.

Possiamo anche derivare le condizioni necessarie del secondo ordine ricordando la definizione (**??**) di direzione ammissibile per un poliedro in forma standard.
In particolare si ha:

Teorema 7.3.8 (Condizioni necessarie del 2^o ordine per (P-ST)) *Sia x^* un punto di minimo locale del problema (P-ST). Allora esistono un vettore $\lambda^* \in I\!R^n$ e $\mu^* \in I\!R^m$ tali che risultino soddisfatte le condizioni di KKt e inoltre risulti:*

$$z^T \nabla^2_{xx} L(x^*, \lambda^*, \mu^*) z \geq 0 \quad \text{per ogni } z: \ Az = 0, \quad \text{con } z_j \geq 0 \ \text{ per } j: \ x^*_j = 0 \quad \text{e} \quad \nabla f(x^*)^T z = 0.$$

Naturalmente le condizioni di KKT possono essere scritte per un qualunque problema di minimizzazione su un poliedro[1]. Nel caso generale supponiamo di avere un problema definito da vincoli di uguaglianza e disuguaglianza del tipo.

$$\begin{array}{ll} \min & f(x) \\ & Dx = h \\ & Ax \geq b \end{array} \qquad \text{(P-GEN)}$$

in cui D è una matrice $p \times n$, A è una matrice $q \times n$, $h \in I\!R^p$, $b \in I\!R^q$. Allora

La funzione Lagrangiana per (P-GEN) è

$$L(x, \lambda, \mu) = f(x) + \lambda^T (b - Ax) + \mu^T (Dx - h).$$

Inoltre

Le condizioni di KKT per il problema (P-GEN) sono soddisfatte in un punto $(x^*, \lambda^*, \mu^*) \in I\!R^n \times I\!R^p \times I\!R^q$ se:

(i) $Ax^* \geq b$, $Dx^* = h$ (ammissibilità),

(ii) $\nabla_x L(x^*, \lambda^*, \mu^*) = \nabla f(x^*) - A^T \lambda^* + D^T \mu^* = 0$ (stazionarietà),

(iii) $\lambda^* \geq 0$,

(iv) ${\lambda^*}^T (b - Ax^*) = 0$ (complementarità).

Esempio 7.3.9 Consideriamo la seguente modifica del problema dell'Esempio 7.3.3.

$$\begin{array}{ll} \max & x_2^2 + \dfrac{1}{2} x_1 \\ & x_1 - x_2 = 1 \\ & \frac{1}{2} x_1 + x_2 \leq 1 \\ & x_1 \geq 0 \end{array}$$

Determiniamo tutti punti di KKT. Mettiamo innanzitutto il problema nella forma (P-GEN).

$$\begin{array}{ll} \min & -x_2^2 - \dfrac{1}{2} x_1 \\ & x_1 - x_2 = 1 \\ & -\frac{1}{2} x_1 - x_2 \geq -1 \\ & x_1 \geq 0 \end{array}$$

La funzione Lagrangiana per questo problema è:

$$L(x, \lambda) = -x_2^2 - \frac{1}{2} x_1 + \mu(-1 + x_1 - x_2) + \lambda_1(-1 + \frac{1}{2} x_1 + x_2) - \lambda_2 x_1$$

Le condizioni di KKT sono:

[1] Nel caso in cui l'insieme ammissibile non sia un poliedro, è ancora possibile definire le condizioni necessarie di KKT purché i vincoli soddisfino opportune ipotesi dette *condizioni di qualificazione dei vincoli.* In questo corso non si trattano vincoli non lineari generici.

(i) (ammissibilità)

$$\begin{aligned}&x_1 - x_2 = 1\\ &-\frac{1}{2}x_1 - x_2 \geq -1\\ &x_1 \geq 0\end{aligned}$$

(ii) (stazionarietà)

$$\begin{aligned}&-\frac{1}{2} + \mu + \frac{1}{2}\lambda_1 - \lambda_2 = 0\\ &-2x_2 - \mu + \lambda_1 = 0\end{aligned}$$

(iii) (non negatività)

$$\lambda_1, \lambda_2 \geq 0$$

(iv) (complementarità)

$$\begin{aligned}&\lambda_1(-1 + \frac{1}{2}x_1 + x_2) = 0\\ &\lambda_2 x_1 = 0\end{aligned}$$

Determiniamo i punti di KKT esaminando tutti i casi possibili. Dalla condizione $\lambda_3 x_1 = 0$ si ottiene che

$$0 = \lambda_2 x_1 = \begin{cases} x_1 = 0 \overset{(i)}{\longmapsto} \quad x_2 = -1 \overset{(\text{iv})_2}{\longmapsto} \lambda_1 = 0, \overset{(\text{ii})_2}{\longmapsto} \mu = 2 \overset{(\text{ii})_3}{\longmapsto} \lambda_2 = 3/2 \\ \\ \lambda_2 = 0 \text{ imponiamo } \begin{cases} \lambda_1 = 0 \overset{(\text{ii})_1}{\longmapsto} \mu = 1/2, \overset{(ii)_2}{\longmapsto} x_2 = -1/4, \overset{(\text{iv})_1}{\longmapsto} x_1 = 3/4 \\ \lambda_1 > 0 \overset{(\text{iv})_2+(\text{i})}{\longmapsto} x_1 = 4/3, x_2 = 1/3, \overset{(\text{ii})_1+(\text{ii})_2}{\longmapsto} \lambda_1 = \frac{7}{9}\ \mu = \frac{1}{9}. \end{cases} \end{cases}$$

Quindi abbiamo trovato 3 punti di KKT

1. $x^1 = (0,\ -1)^T$ con moltiplicatori $\mu = 2$ e $\lambda = (0,\ 3/2)^T$ con valore della funz. obiettivo $f = -1$;
2. $x^2 = (3/4,\ -1/4)^T$ con moltiplicatori $\mu = 1/2$ e $\lambda = (0,\ 0)^T$ con valore della funz. obiettivo $f = -7/16$;
3. $x^3 = (4/3,\ 1/3)^T$ con moltiplicatori $\mu = 1/9$ e $\lambda = (7/9,\ 0)^T$ con valore della funz. obiettivo $f = -7/9$;

Poiché l'insieme ammissibile è compatto, esiste un punto di minimo globale e si trova tra i punti di KKT determinati sopra. Si tratta di scegliere quello con valore della funzione obiettivo minore: si tratta del punto $x^1 = (0,\ -1)$.

□

Possiamo inoltre scrivere le condizioni necessarie del secondo ordine.

Se x^* è un minimo del problema (P-GEN) allora esistono λ^*, μ^* che soddisfano le condizioni di KKT ed inoltre risulta

$$z^T \nabla^2_{xx} L(x^*, \lambda^*, \mu^*) z \geq 0 \quad \text{per ogni } z: \ Dz = 0,\ a_i^T z \geq 0 \ \ i \in I(x^*) \ \text{ e } \ \nabla f(x^*)^T z = 0$$

Consideriamo nuovamente l'esempio 7.3.9.

Esempio 7.3.10 *Consideriamo i 3 punti di KKT precedentemente determinati e verifichiamo se soddisfano le condizioni necessarie del secondo ordine. La funzione Lagrangiana per questo problema è:*

$$L(x,\lambda) = -x_2^2 - \frac{1}{2}x_1 + \mu(-1 + x_1 - x_2) + \lambda_1(-1 + \frac{1}{2}x_1 + x_2) - \lambda_2 x_1$$

il cui hessiano è

$$\nabla^2_{xx} L(x,\lambda,\mu) = \begin{pmatrix} 0 & 0 \\ 0 & -2 \end{pmatrix} = \nabla^2 f(x).$$

Quindi per ogni $z = (z_1,\ z_2)^T$ *risulta* $z^T \nabla^2_{xx} L(x^*,\lambda^*,\mu^*) z = -2z_2^2$. *Inoltre* $\nabla f = \begin{pmatrix} -1/2 \\ -2x_2 \end{pmatrix}$.

Consideriamo il punto $x^1 = (0,\ -1)^T$. *Dobbiamo determinare i vettori z che definiscono le direzioni ammissibili. In* x^1 *è attivo il vincolo di disuguaglianza* $x_1 \geq 0$, *quindi le direzioni ammissibili sono i vettori* $z \in R^2$ *che risolvono il sistema*

$$\begin{array}{l} z_1 - z_2 = 0 \\ z_1 \geq 0, \end{array}$$

ovvero tutti i vettori del tipo $z = (t,\ t)^T$ *con* $t \geq 0$. *Inoltre deve risultare*

$$\nabla f(x^1)^T z = (t,\ t)^T \begin{pmatrix} -\frac{1}{2} \\ 2 \end{pmatrix} = -\frac{1}{2}t + 2t = 0,$$

da cui si ottiene $t = 0$ *e* $z = (0,\ 0)^T$. *La condizione necessaria del secondo ordine quindi è soddisfatta in quanto* $-2t^2 \equiv 0$.

Consideriamo il punto $x^2 = (3/4,\ -1/4)^T$. *In* x^2 *non è attivo nessun vincolo di disuguaglianza, quindi le direzioni ammissibili sono i vettori* $z \in R^2$ *che risolvono il sistema*

$$z_1 - z_2 = 0$$

ovvero tutti i vettori del tipo $z = (t,\ t)^T$. *Inoltre deve risultare*

$$\nabla f(x^2)^T z = (t,\ t)^T \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix} = -\frac{1}{2}t + \frac{1}{2}t = 0.$$

La condizione necessaria del secondo ordine quindi NON è soddisfatta in quanto $-2t^2 \geq 0$ *non è verificata qualunque sia* t. *Il punto* x^2 *non può essere un minimo locale.*

Consideriamo il punto $x^3 = (4/3,\ 1/3)^T$. *In* x^3 *è attivo il primo vincolo di disuguaglianza, quindi le direzioni ammissibili sono i vettori* $z \in R^2$ *che risolvono il sistema*

$$\begin{array}{l} z_1 - z_2 = 0 \\ -\frac{1}{2}z_1 - z_2 \geq 0 \end{array}$$

ovvero tutti i vettori del tipo $z = (t,\ t)^T$ *con* $t \leq 0$. *Inoltre deve risultare*

$$\nabla f(x^3)^T z = (t,\ t)^T \begin{pmatrix} -\frac{1}{2} \\ -\frac{2}{3} \end{pmatrix} = -\frac{1}{2}t - \frac{2}{3}t = 0,$$

che è verificata per $t = 0$. *Quindi l'unico vettore* $z = (0,\ 0)^T$ *e la condizione necessaria del secondo ordine quindi è soddisfatta in quanto* $-2t^2 \equiv 0$ *è verificata.* □

Capitolo 8

Le condizioni di ottimo e la dualità per la PL

8.1 Introduzione

In questo capitolo deriveremo le condizioni necessarie e sufficienti di ottimo per la programmazione lineare. Utilizzeremo le condizioni di KKT derivate nel capitolo 7. A partire dalle condizioni di ottimo si derivano i principali risultati della teoria della dualità per la PL, con cenni all'interpretazione economica e all'analisi di sensibilità.

8.2 Le condizioni di ottimalità nella Programmazione Lineare

Nel caso in cui la funzione obiettivo sia lineare ovvero del tipo $f(x) = c^T x$, i problemi (P-POL) e (P-ST) sono problemi di Programmazione Lineare. Abbiamo già caratterizzato le soluzioni ottime nel capitolo 5. Possiamo ora formulare le condizioni di ottimo per la PL. Poiché si tratta di problemi convessi le condizioni di KKT sono necessarie e sufficienti.

Consideriamo inizialmente il problema di PL

$$\begin{aligned} \min \quad & c^T x \\ & Ax \geq b \end{aligned} \qquad (PL)$$

La funzione Lagrangiana si scrive

$$L(x, \lambda) = c^T x + \lambda^T (b - Ax)$$

e le condizioni di KKT sono le seguenti.

Teorema 8.2.1 (Condizioni di KKT per il problema (PL)) *Un punto x^* è minimo globale del problema di Programmazione Lineare (PL) se e solo se esiste $\lambda^* \in \mathbb{R}^m$ tale che valgano le condizioni:*

(i) $Ax^* \geq b$,

(ii) $c = A^T \lambda^*$,

(iii) $\lambda^* \geq 0$,

(iv) ${\lambda^*}^T (b - Ax^*) = 0$.

La condizione di complementaritá (iv) del Teorema 8.2.1 può essere scritta in modo equivalente utilizzando la condizione (ii). In particolare possiamo scrivere

$$0 \stackrel{(iv)}{=} \lambda^{*T}(b - Ax^*) = \lambda^{*T}b - \lambda^{*T}Ax^* = \lambda^{*T}b - x^{*T}A^T\lambda^* \stackrel{(ii)}{=} \lambda^{*T}b - x^{*T}c$$

Possiamo quindi scrivere le condizioni necessarie e sufficienti di KKT in una forma equivalente e più diffusamente utilizzata come condizioni di ottimo per la Programmazione Lineare.

(Condizioni necessarie e sufficienti di ottimo per (PL))

Un punto x^* è minimo globale del problema di Programmazione Lineare

$$\begin{aligned} \min \quad & c^T x \\ & Ax \geq b \end{aligned} \tag{PL}$$

se e solo se esiste $\lambda^* \in I\!R^m$ tale che valgano le condizioni:

(i) $Ax^* \geq b$,

(ii) $c = A^T\lambda^*$,

(iii) $\lambda^* \geq 0$,

(iv) $b^T\lambda^* = c^T x^*$.

Possiamo fare qualche ulteriore considerazione. Le condizioni (ii) e (iii) esprimono il fatto che $\lambda^* \in I\!R^m$ deve essere ammissibile rispetto ai vincoli lineari

$$\begin{aligned} & A^T\lambda = c \\ & \lambda \geq 0 \end{aligned} \tag{8.1}$$

Il sistema (8.1) rappresenta un poliedro in forma standard nello spazio $I\!R^m$.

Consideriamo allora un $x \in I\!R^n$ soluzione ammissibile del problema (PL), e un $\lambda \in I\!R^m$ appartenente al poliedro definito da (8.1), ovvero una coppia (x, λ) che soddisfano

$$Ax \geq b, \qquad \lambda \geq 0, \quad A^T\lambda = c. \tag{8.2}$$

Possiamo allora facilmente scrivere

$$\begin{aligned} c^T x \quad & \stackrel{(\lambda \geq 0,\ b \leq Ax)}{\geq} \quad c^T x + \lambda^T(b - Ax) \\ & = \quad b^T\lambda + (c^T - \lambda^T A)x \\ & \stackrel{(A^T\lambda = c)}{=} \quad b^T\lambda. \end{aligned}$$

Per ogni coppia (x, λ) che soddisfa $Ax \geq b,\ \lambda \geq 0,\ A^T\lambda = c$ risulta

$$c^T x \geq b^T\lambda.$$

Si tratta di un risultato molto significativo, che riprenderemo più avanti. Notiamo che, in particolare, poiché il minimo globale x^* del problema (PL) è ammissibile, otteniamo immediatamente che vale

$$c^T x^\star \geq b^T\lambda \tag{8.3}$$

per ogni λ che soddisfa (8.1). Possiamo quindi affermare che una qualunque soluzione $\lambda \in I\!R^m$ che soddisfa (8.1) consente di determinare una limitazione inferiore al valore ottimo del problema (PL).

Ricordando inoltre che x^* è un minimo del problema (PL), e quindi risulta $c^T x^* \leq c^T x$ per ogni x tale che $Ax \geq b$, possiamo affermare:

(Limitazione superiore ed inferiore per il valore ottimo del problema (PL))

Sia x^* un minimo globale del problema di Programmazione Lineare (PL). Per ogni $(\tilde{x}, \tilde{\lambda})$ che soddisfa $Ax \geq b,\ \lambda \geq 0,\ A^T\lambda = c$, risulta

$$b^T\tilde{\lambda} \leq c^T x^* \leq c^T\tilde{x}.$$

Per migliorare la limitazione inferiore espressa dalle disuguaglianza (8.3), si può rendere quanto più possibile grande il termine di destra $b^T\lambda$ della disuguaglianza (8.3), cioè si può massimizzare la quantità $b^T\lambda$ al variare del vettore $\lambda \in \mathbb{R}^m$, tra tutti i vettori che soddisfano $A^T\lambda = c$, $\lambda \geq 0$. Piú formalmente si può cercare $\lambda^\star$ soluzione del problema

$$\begin{array}{ll} \max & b^T\lambda \\ & A^T\lambda = c \\ & \lambda \geq 0 \end{array}$$

Si tratta di un problema di PL in forma standard (è sufficiente cambiare verso all'ottimizzazione da max a min).

Possiamo dimostrare che le condizioni necessarie e sufficienti di minimo globale per questo problema sono le stesse condizioni per il problema (PL). Abbiamo infatti il seguente teorema.

Teorema 8.2.2 (Condizioni di ottimalità per la (PL-ST)) *Un punto $\lambda^* \in I\!R^m$ è minimo globale del problema di Programmazione Lineare*

$$\begin{array}{ll} \max & b^T\lambda \\ & A^T\lambda = c \\ & \lambda \geq 0 \end{array} \qquad (PL-ST)$$

se e solo se esiste $x^* \in I\!R^n$ tale che valgano le condizioni:

(i) $A^T\lambda^* = c$,

(ii) $\lambda^* \geq 0$,

(iii) $Ax^* \geq b$,

(iv) $b^T\lambda^* = c^T x^*$.

Dimostrazione. Utilizziamo il teorema 7.3.7 applicato a (PL-ST). Osserviamo che (PL-ST) è in forma di massimizzazione e che è definito nello spazio $I\!R^m$. La funzione Lagrangiana per (PL-ST) è

$$L(\lambda, x, v) = -b^T\lambda + x^T(A^T\lambda - c) - v^T\lambda$$

dove $x \in I\!R^n$ (righe di A^T) e $v \in I\!R^m$.
Per il teorema 7.3.7 esistono dei vettori $x^* \in I\!R^n$ e $v^* \in I\!R^m$ (dimensione λ) tali che

(a) $A^T\lambda^* = c$, $\lambda^* \geq 0$,

(b) $-b + Ax^* - v^* = 0$,

(c) $v^* \geq 0$,

(d) ${v^*}^T\lambda^* = 0$.

La (a) è esattamente la (i) e (ii). Dalla condizione (b) si può ricavare v^* ed eliminarlo dalle altre condizioni in cui è presente. Si ha

$$-b + Ax^* = v^*.$$

Imponendo la (c) $v^* \geq 0$ si ottiene

$$-b + Ax^* \geq 0$$

che è la (iii). Sostituendo in (d) si ha poi

$$0 = {v^*}^T \lambda^* = (-b + Ax^*)^T \lambda^* = -b^T \lambda^* + {x^*}^T A^T \lambda^* = -b^T \lambda^* + {x^*}^T c$$

che è la (iv). □

Possiamo quindi concludere che le condizioni necessarie e sufficienti di ottimo per i problemi di Programmazione Lineare (PL) e (PL-ST) sono le stesse.

Anche per il problema di Programmazione Lineare (PL-ST) possiamo dare una limitazione inferiore e superiore al valore ottimo $b^T \lambda^*$. Infatti, presa una qualunque coppia (x, λ) che soddisfa (8.2), si può scrivere

$$\begin{array}{rcl} b^T \lambda & \overset{(A^T\lambda = c)}{=} & b^T \lambda + x^T (c - A^T \lambda) \\ & = & c^T x + \lambda^T (b - Ax) \\ & \overset{(\lambda \geq 0,\ b \leq Ax)}{\leq} & c^T x. \end{array}$$

Possiamo allora affermare:

(Limitazione superiore ed inferiore per il valore ottimo del problema (PL-ST))

Sia λ^* un minimo globale del problema di Programmazione Lineare (PL-ST). Per ogni $(\tilde{x}, \tilde{\lambda})$ che soddisfano (8.2), si ha

$$b^T \tilde{\lambda} \leq b^T \lambda^* \leq c^T \tilde{x}.$$

È evidente la forte relazione che esiste tra i due problemi di Programmazione Lineare (PL) e (PL-ST). Questa coppia di problemi viene detta *primale-duale*.

(Problema primale e duale)

Il problema

$$\begin{array}{ll} \min & c^T x \\ & Ax \geq b \end{array} \qquad (PL)$$

e il problema

$$\begin{array}{ll} \max & b^T \lambda \\ & A^T \lambda = c \\ & \lambda \geq 0 \end{array} \qquad (PL - ST)$$

costituiscono una coppia *primale-duale* di Programmazione Lineare.

In maniera del tutto simmetrica, il problema (PL) risulterà il problema duale del problema (PL-ST). La coppia primale (P)-duale (D) costituita dai problemi (PL) e (PL-ST) è solo una possibile coppia primale-duale. La costruzione del problema duale a partire da un qualsiasi problema di PL è rimandata nel prossimo paragrafo.

La teoria della dualità costituisce uno degli aspetti fondamentali nello studio della Programmazione Lineare, che merita una particolare attenzione.

I risultati che abbiamo ricavato per la coppia primale-duale (PL) e (PL-ST) possono essere generalizzati ad una qualunque coppia primale-duale e costituiscono i risultati fondamentali della teoria della dualità.

In particolare la limitazione sul valore della funzione obiettivo costituisce il teorema della dualità debole.

Teorema 8.2.3 (Teorema della Dualità debole) *Per ogni soluzione ammissibile $\bar{x}$ del problema primale (P) ed ogni soluzione ammissibile $\bar{\lambda}$ del problema duale (D) si ha*

$$b^T \bar{\lambda} \leq c^T \bar{x}$$

cioè il valore della funzione obiettivo duale in $\bar{\lambda}$ è minore o uguale del valore della fuzione obiettivo primale in $\bar{x}$.

Da cui si ottiene il seguente corollario.

Corollario 8.2.4 *Se $\bar{x}$ è una soluzione ammissibile del problema primale (P) e $\bar{u}$ una soluzione ammissibile del problema duale (D) tali che*

$$c^T \bar{x} = b^T \bar{\lambda} \tag{8.4}$$

allora $\bar{x}$ e $\bar{u}$ sono soluzioni ottime rispettivamente per il problema primale (P) e per il problema duale (D).

Utilizzando il teorema della dualità debole possiamo anche derivare delle condizioni di illimitatezza di un problema di PL. Vale in particolare questo risultato.

Teorema 8.2.5 (Condizione di illimitatezza del primale) *Se il problema primale (P) è illimitato (inferiormente) allora il problema duale (D) è inammissibile. Viceversa se il problema duale (D) è illimitato (superiormente) allora il problema primale (P) è inammissibile.*

Dim.: Supponiamo che (P) sia illimitato e che, per assurdo, il problema duale (D) sia ammissibile, cioè che esista una soluzione ammissibile $\bar{\lambda}$ del problema duale (D). Allora, per il Teorema 8.2.3 (Teorema della Dualità debole), risulta $c^T x \geq b^T \bar{\lambda}$ per ogni soluzione ammissibile x del problema primale (P). Ma questo contraddice l'ipotesi che (P) sia illimitato inferiormente.
Viceversa □

Le condizioni necessarie e sufficienti di ottimo per la Programmazione Lineare ricavate nel paragrafo precedente costituiscono invece il cosiddetto teorema della *dualità forte*.

Teorema 8.2.6 (Teorema della Dualità Forte) *Se il problema primale (P) ammette una soluzione ottima $x^\star$ allora anche il problema duale (D) ammette una soluzione ottima $\mu^\star$. Simmetricamente, se il problema duale (D) ammette una soluzione ottima $\mu^\star$ allora anche il problema primale (P) ammette una soluzione ottima $x^\star$. Inoltre i valori delle funzioni obiettivo dei due problemi all'ottimo sono uguali cioè risulta*

$$c^T x^\star = b^T u^\star.$$

Siamo ora in grado di riformulare delle *condizioni di ottimalità* in riferimento alla coppia primale–duale (PL) e (PL-ST).

Condizioni di Ottimalità
Siano dati $\bar{x} \in \mathbb{R}^n$ e $\bar{\lambda} \in \mathbb{R}^m$. Allora $\bar{x}$ e $\bar{\lambda}$ sono soluzioni ottime rispettivamente per la coppia

$$(PL) \qquad \begin{array}{ll} \min & c^T x \\ & Ax \geq b \end{array} \qquad\qquad \begin{array}{ll} \max & b^T \lambda \\ & A^T \lambda = c \\ & \lambda \geq 0 \end{array} \qquad (PL-ST)$$

se e solo se valgono le seguenti condizioni

(i) $A\bar{x} \geq b$, (ammissibilità primale)

(ii) $A^T\bar{\lambda} = c, \quad \bar{\lambda} \geq 0$ (ammissibilità duale);

(iii) $c^T\bar{x} = b^T\bar{\lambda}$ (coincidenza dei valori delle funzioni obiettivo).

Sulla base dei risultati fino ad ora esaminati si evince che data un coppia primale–duale di problemi di Programmazione Lineare possono verificarsi le seguenti situazioni: o entrambi ammettono soluzione ottima, oppure se uno è illimitato l'altro è inammissibile, oppure sono entrambi inammissibili. Queste possibilità sono riportate schematicamente nella tabella che segue.

		DUALE		
		OTTIMO FINITO	ILLIMITATO SUPERIOR.	INAMMISSIBILE
PRIMALE	OTTIMO FINITO	SI	NO	NO
	ILLIMITATO INFERIOR.	NO	NO	SI
	INAMMISSIBILE	NO	SI	SI

8.3 Costruzione del duale di un problema di Programmazione Lineare

La coppia di problemi primale-duale definita nel paragrafo precedente è solo una delle possibile coppie primali-duali di problemi di Programmazione Lineare.

Si consideri un problema Programmazione Lineare scritto nella forma più generale possibile cioè nella forma

$$\begin{array}{ll} \min & d^T z + e^T y \\ & Cz + Dy = h \\ & Ez + Fy \geq g \\ & z \geq 0 \end{array} \tag{8.5}$$

in cui le variabili sono $x = (z,\ y)^T \in I\!R^n$ suddivise $z \in \mathbb{R}^{n_1}$, $y \in \mathbb{R}^{n_2}$ con $z \geq 0$ e y non vincolata in segno. Corrispondentemente $c \in \mathbb{R}^{n_1}$, $d \in \mathbb{R}^{n_2}$, C matrice $p \times n_1$, D matrice $p \times n_2$ e $h \in \mathbb{R}^p$; E matrice $q \times n_1$, F matrice $q \times n_2$ e $g \in \mathbb{R}^q$.

La notazione in cui è scritto questo generico problema di Programmazione Lineare (8.5) è tale da evidenziare separatamente gli elementi che intervengono nella formulazione: le variabili sono partizionate nella variabili x vincolate in segno e y non vincolate in segno e corrispondentemente anche i coefficienti di costo della funzione obiettivo sono partizionati in c e d; i vincoli sono scritti suddividendo quelli di uguaglianza (=) e quelli di disuguaglianza (≥ 0).

Per costruire il problema duale del problema (8.5) è sufficiente riportarsi alla forma del problema (PL) con soli vincoli di disuguaglianza . A tale scopo riscriviamo il problema (8.5) nella forma equivalente

$$
\begin{array}{ll}
\min & d^T z + e^T y \\
& Cz + Dy \geq h \\
& -Cz - Dy \geq -h \\
& Ez + Fy \geq g \\
& z \geq 0
\end{array}
$$

ovvero in forma matriciale

$$
\begin{array}{ll}
\min & \begin{pmatrix} d \\ e \end{pmatrix}^T \begin{pmatrix} z \\ y \end{pmatrix} \\
& \begin{pmatrix} C & D \\ -C & -D \\ E & F \\ I_{n_1} & 0 \end{pmatrix} \begin{pmatrix} z \\ y \end{pmatrix} \geq \begin{pmatrix} h \\ -h \\ g \\ 0 \end{pmatrix},
\end{array}
$$

dove I_{n_1} è la matrice identità di ordine n_1. Quindi il problema (8.5) è stato ricondotto nella forma $\min\ c^T x,\ Ax \geq b$ in cui

$$
A = \begin{pmatrix} C & D \\ -C & -D \\ E & F \\ I_{n_1} & 0 \end{pmatrix} \qquad b = \begin{pmatrix} h \\ -h \\ g \\ 0 \end{pmatrix}
$$

Siamo quindi in grado di scrivere il duale di questo problema nella forma (PL-ST)

$$
\begin{array}{ll}
\max & b^T \lambda \\
& A^T \lambda = c \\
& \lambda \geq 0
\end{array}
$$

in cui il vettore $\lambda = (u,\ v,\ w,\ t)^T \in I\!R^m = I\!R^{2p+q+n_1}$ Scriviamo quindi

$$
A^T = \begin{pmatrix} C & -C & E & I_{n_1} \\ D & -D & F & 0 \end{pmatrix}
$$

Il problema duale si scrive in forma matriciale

$$
\begin{array}{ll}
\max & \begin{pmatrix} h \\ -h \\ g \\ 0 \end{pmatrix}^T \begin{pmatrix} u \\ v \\ w \\ t \end{pmatrix} \\
& \begin{pmatrix} C & -C & E & I_{n_1} \\ D & -D & F & 0 \end{pmatrix} \begin{pmatrix} u \\ v \\ w \\ t \end{pmatrix} = \begin{pmatrix} d \\ e \end{pmatrix} \\
& u,\ v,\ w,\ t \geq 0.
\end{array}
$$

Sviluppando si ottiene

$$\begin{array}{ll} \max & h^T u - h^T v + g^T w \\ & C^T u - C^T v + E^T w + I_{n_1} t = d \\ & D^T u - D^T v + F^T w = e \\ & u \geq 0, v \geq 0, w \geq 0, t \geq 0 \end{array} \quad \text{che è equivalente a} \quad \begin{array}{ll} \max & h^T (u - v) + g^T w \\ & C^T (u - v) + E^T w + t = d \\ & D^T (u - v) + F^T w = e \\ & u \geq 0, v \geq 0, w \geq 0, t \geq 0. \end{array}$$

Effettuando il cambio di variabili $u - v = \mu$ si ottiene il seguente problema

$$\begin{array}{ll} \max & h^T \mu + g^T w \\ & C^T \mu + E^T w + t = d \\ & D^T \mu + F^T w = e \\ & w \geq 0, z \geq 0. \end{array}$$

Eliminando la variabile t utilizzando il primo vincolo $t = d - C^T \mu - E^T w$ possiamo ancora scrivere

$$\begin{array}{ll} \max & h^T \mu + g^T w \\ & C^T \mu + E^T w \leq d \\ & D^T \mu + F^T w = e \\ & w \geq 0 \end{array} \tag{8.6}$$

nelle variabili $(\mu, w) \in \mathbb{R}^q \times \mathbb{R}^p$, con μ non vincolata in segno e $w \geq 0$.
Il problema (8.6) è il *problema duale* del problema (8.5) che viene detto *problema primale*.

Le variabili (x, y) sono dette *variabili primali*; le variabili (μ, w) sono dette *variabili duali*. I due problemi costituiscono la generica coppia *coppia primale–duale*.

$$(P) \quad \begin{array}{ll} \min & c^T z + d^T y \\ & Cz + Dy = h \\ & Ez + Fy \geq g \\ & z \geq 0 \end{array} \qquad \begin{array}{ll} \max & h^T \mu + g^T w \\ & C^T \mu + E^T w \leq d \\ & D^T \mu + F^T w = e \\ & w \geq 0. \end{array} \quad (D)$$

Dall'osservazione dei due problemi si deducono facilmente le proprietà fondamentali di una coppia primale–duale; innanzitutto un problema è di minimizzazione mentre l'altro è di massimizzazione. Inoltre poiché la matrice dei coefficienti dei vincoli di un problema si ottiene trasponendo quella dell'altro, si ha che ad ogni variabile di un problema corrisponde un vincolo nell'altro. Si osserva inoltre uno scambio tra i termini noti di un problema e i coefficienti della funzione obiettivo dell'altro.

Queste proprietà possono essere così schematicamente riassunte:

- il problema duale di un problema di minimizzazione è un problema di massimizzazione e simmetricamente, il problema duale di un problema di massimizzazione è un problema di minimizzazione;
- ad ogni vincolo di uguaglianza del problema primale è associata una variabile nel problema duale non vincolata in segno;
- ad ogni vincolo di disuguaglianza (di maggiore o uguale) del problema primale è associata una variabile nel problema duale vincolata in segno;
- ad ogni variabile vincolata in segno del problema primale è associato un vincolo di disuguaglianza ($\leq$ se il problema è di massimizzazione, $\geq$ se il problema è di minimizzazione) del problema duale;
- ad ogni variabile non vincolata in segno del problema primale è associato un vincolo di uguaglianza del problema duale.

Queste corrispondenze possono essere riassunte nella tabella che segue dove gli insieme I, J, M e N sono insiemi di indici:

	PRIMALE	**DUALE**	
	$\min c^T x$	$\max b^T u$	
VINCOLI	$= b_i, \quad i \in I$ $\geq b_i, \quad i \in J$	$u_i, \quad i \in I$, libere $u_i, \quad i \in J, u_i \geq 0$	VARIABILI
VARIABILI	$x_j \geq 0, \quad j \in M$ $x_j, \quad j \in N$ libere	$\leq c_j, \quad j \in M$ $= c_j, \quad j \in N$	VINCOLI

I risultati che abbiamo riportato nel paragrfao precedente con riferimento alla coppia primale-duale (PL) e (PL-ST) possono essere generalizzati ad una qualunque coppia primale-duale e costituiscono i risultati fondamentali della teoria della dualità.

Esempio 8.3.1 *Si consideri il seguente problema di Programmazione Lineare*

$$\begin{array}{ll} \max & 4x_1 + 3x_2 + 2x_3 \\ & x_1 + 2x_2 + 3x_3 \leq 8 \\ & 2x_1 - x_3 \leq 7 \\ & 3x_1 + 4x_2 - x_3 \leq 5 \\ & x_2 + x_3 \leq 6 \\ & x_2 \geq 0 \end{array}$$

Il problema duale è il seguente problema di minimizzazione

$$\begin{array}{ll} \min & 8u_1 + 7u_2 + 5u_3 + 6u_4 \\ & u_1 + 2u_2 + 3u_3 = 4 \\ & 2u_1 + 4u_3 + u_4 \geq 3 \\ & 3u_1 - u_2 - u_3 - u_4 = 2 \\ & u_1 \geq 0, u_2 \geq 0, u_3 \geq 0, u_4 \geq 0. \end{array}$$

□

Esempio 8.3.2 *Si consideri il seguente problema di Programmazione Lineare*

$$\begin{array}{ll} \min & 2x_1 - 3x_2 + x_3 \\ & 3x_1 + x_2 + 5x_3 \geq 7 \\ & x_1 + x_2 - 6x_3 \leq 9 \\ & 4x_1 - x_2 - 2x_3 = 8 \\ & x_1 \geq 0, x_2 \geq 0. \end{array}$$

Dopo aver riscritto il secondo vincolo come $-x_1 - x_2 + 6x_3 \geq -9$ *si può formulare facilmente il problema duale associato*

$$\begin{array}{ll} \max & 7u_1 - 9u_2 + 8u_3 \\ & 3u_1 - u_2 + 4u_3 \leq 2 \\ & u_1 - u_2 - u_3 \leq -3v5u_1 + 6u_2 - 2u_3 = 1 \\ & u_1 \geq 0, u_2 \geq 0. \end{array}$$

□

Esempio 8.3.3 *Si consideri il problema di Programmazione Lineare*

$$\begin{array}{ll} \min & x_1 + 3x_2 \\ & x_1 + 4x_2 \geq 24 \\ & 5x_1 + x_2 \geq 25 \\ & x_1 \geq 0, x_2 \geq 0 \end{array}$$

Geometricamente si ricava facilmente che questo problema ammette soluzione ottima nel punto $(x_1, x_2) = (4, 5)$ *e il valore ottimo della funzione obiettivo è pari a* 19*. Se si considera il problema duale*

$$\begin{array}{ll} \max & 24u_1 + 25u_2 \\ & u_1 + 5u_2 \leq 1 \\ & 4u_1 + u_2 \leq 3 \\ & u_1 \geq 0, u_2 \geq 0; \end{array}$$

si ricava facilmente (geometricamente) che, in accordo con quanto previsto dal Teorema della Dualità Forte, anche questo problema ammette soluzione ottima nel punto $(u_1, u_2) = \left(\frac{14}{19}, \frac{1}{19}\right)$ *e il valore ottimo della funzione obiettivo vale* 19*.*

Esempio 8.3.4 *Si consideri il problema di Programmazione Lineare*

$$\begin{array}{ll} \max & 2x_1 + 3x_2 \\ & -2x_1 + x_2 \leq 3 \\ & -\frac{1}{2}x_1 + x_2 \leq 6 \\ & x_1 \geq 0, x_2 \geq 0. \end{array}$$

Geometricamente si ricava che il problema è illimitato superiormente. Quindi, per l'analisi teorica vista deve risultare che il suo duale è inammissibile. E infatti se si considera il problema duale associato

$$\begin{array}{ll} \min & 3u_1 + 6u_2 \\ & -2u_1 - \frac{1}{2}u_2 \geq 2 \\ & u_1 + u_2 \geq 3 \\ & u_1 \geq 0, u_2 \geq 0 \end{array}$$

si vede facilmente che questo problema non ammette soluzioni ammissibili.

8.4 Interpretazione della Dualità

Nei modelli reali le variabili (primali) possono rappresentare, ad esempio, livelli di produzione e i coefficienti di costo possono essere associati ai profitti ricavati dalla vendita dei prodotti. Quindi la funzione obiettivo di un problema primale indica direttamente come un aumento della produzione può influenzare il profitto. Sempre in relazione, ad esempio, ad un modello per la pianificazione della produzione, i vincoli di un problema (primale) possono rappresentare una limitazione dovuta alla limitata disponibilità delle risorse; ora, un aumento della disponibilità delle risorse può consentire un aumento della produzione e quindi anche del profitto, ma questa relazione tra aumento della disponibilità delle risorse e aumento del profitto non si deduce facilmente dal problema formulato (il problema primale). Uno dei possibili usi della dualità è quello di rendere esplicito l'effetto dei cambiamenti nei vincoli (ad esempio in quelli di disponibilità di risorse) sul valore della funzione obiettivo.

L'analisi della sensitività si occupa proprio di stabilire come si modifica la soluzione ottima di un modello lineare in conseguenza di variazioni dei dati. Un ruolo fondamentale in questa analisi è giocato dai cosidetti *prezzi ombra*, che, misurano i "costi" impliciti associati ai vincoli. Come vedremo, i prezzi ombra sono nient'altro che le variabili duali.

8.4.1 Interpretazione geometrica della variazione dei dati sui problemi primale duale

Consideriamo la seguente coppia di problemi primale-duale di programmazione lineare:

$$(P)\quad \begin{array}{ll} \max & 2x_1 + 6x_2 \\ & x_1 + 2x_2 \leq 5 \\ & x_1 + x_2 \leq 4 \\ & x_1, x_2 \geq 0 \end{array} \qquad\qquad \begin{array}{ll} \min & 5u_1 + 4u_2 \\ & u_1 + u_2 \geq 2 \\ & 2u_1 + u_2 \geq 6 \\ & u_1, u_2 \geq 0 \end{array}\quad (D)$$

Supponiamo che il problema primale rappresenti un semplice problema di produzione in cui x_1, x_2 rappresentano livelli di attività produttiva e in vincoli corrispondano a due risorse limitate.
La soluzione grafica dei problemi (P) e (D) è riportata in Figura 8.1.

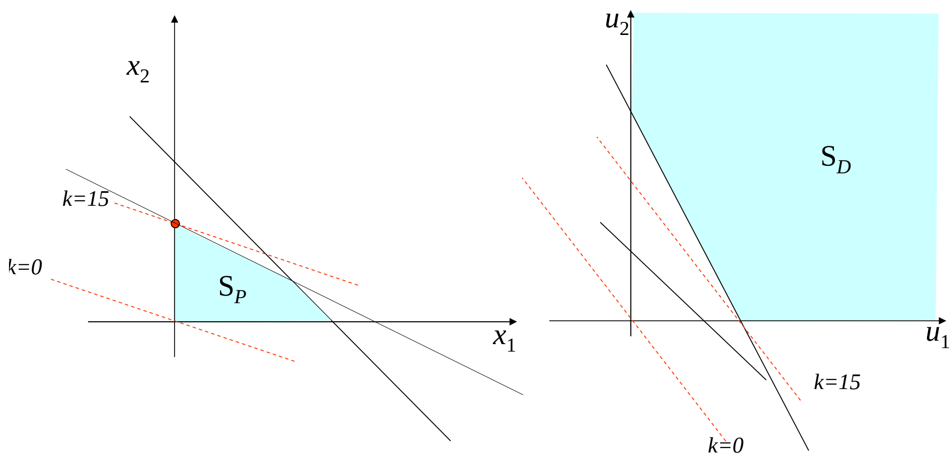

Figura 8.1: Soluzione grafica della coppia primale duale

Il problema primale ha unica soluzione $x^* = (0,\ 5/2)^T$, il duale ammette l'unica soluzione ottima $u^* = (3,\ 0)^T$.
Si osserva geometricamente (ed è ovvio anche dal valore delle variabili duali u^*) che in x^* è attivo un solo vincolo del problema primale, il primo. Questo significa che il livello di produzione ottima corrisponde ad utilizzare completamente la prima risorsa ($x_1^* + 2x_2^* = 5$), mentre la seconda risorsa è in eccesso ($x_1^* + x_2^* = 2,5 < 4$). Inoltre il livello di produzione ottima pone $x_1^* = 0$ che significa che il prodotto rappresentato da x_1 non è prodotto.

Ci si possono porre allora due diverse domande

1. conviene acquisire una quantità maggiore di risorsa relativa al primo vincolo ? Quanta e a quale prezzo ?

2. quando diventa conveniente produrre anche il prodotto x_1 ?

Possiamo rispondere a queste domande facendo delle analisi di tipo geometrico sul problema primale e duale.

Cambiamenti nei termini noti dei vincoli. Acquisire una quantità maggiore di risorsa relativa al primo vincolo significa incrementare il r.h.s del vincolo da 5 a $5 + \delta$ con $\delta \geq 0$. Da un punto di vista analitico i due problemi (P) e (D) diventano

$$(P)_\delta \quad \begin{array}{ll} \max & 2x_1 + 6x_2 \\ & x_1 + 2x_2 \leq 5 + \delta \\ & x_1 + x_2 \leq 4 \\ & x_1, x_2 \geq 0 \end{array} \qquad \begin{array}{ll} \min & (5+\delta)u_1 + 4u_2 \\ & u_1 + u_2 \geq 2 \\ & 2u_1 + u_2 \geq 6 \\ & u_1, u_2 \geq 0 \end{array} \quad (D)_\delta$$

Geometricamente un aumento di δ del r.h.s. del primo vincolo, corrisponde a spostare il vincolo "verso l'alto". La situazione è rappresentata geometricamente in Figura 8.2, in rosso le variazioni. È evidente che la soluzione ottima si sposta nel punto $x^*_\delta = (0, \frac{5+\delta}{2})^T$ di valore $c^T_\delta x^*_\delta = 6\frac{(5+\delta)}{2} = c^T x^* + 3\delta$. L'aumento nella funzione obiettivo è quindi pari a 3δ. Il massimo valore che può assumere δ corrisponde al valore per cui il vincolo $x_1 + 2x_2 \leq 5 + \delta$ passa per il punto $(0,\ 4)^T$, ovvero $\delta = 3$.

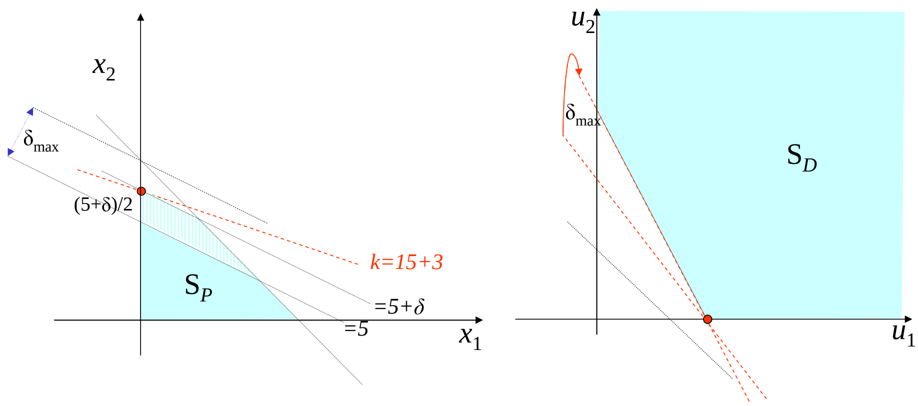

Figura 8.2: Soluzione grafica della coppia primale duale al variare di δ

Per $\delta > 3$, il primo vincolo non gioca più alcun ruolo nella definizione della regione ammissibile e la risorsa che diventa vincolante è la seconda.

Nel problema duale, un variazione di δ corrisponde ad una variazione del coefficiente angolare della funzione obiettivo. La regione ammissibile del problema duale è rimasta invariata. Il valore $\delta = 3$ corrisponde alla funzione obiettivo $8u_1 + 4u_2$ rappresentata in rosso in Figura 8.2 per valore $= 0$. Le curve di livello di questa funzione sono parallele al secondo vincolo del problema duale. La soluzione ottima del problema duale $(D)_\delta$ rimane la soluzione $u^* = (3,\ 0)^T$ per $\delta \leq 3$. Poiché la soluzione ottima $u^*_\delta = u^*$ il valore $c^T_\delta u^*_\delta = (5+\delta)3 = 5u^*_1 + \delta u^*_1$. La variazione del valore ottimo è 3δ pari a quanto determinato nel caso primale.
Il valore $3 = u^*_1$ rappresenta anche quanto sono disposto a pagare al massimo un'unità in più di tale risorsa.

Per valori di $\delta > 3$, la soluzione ottima del problema duale cambia e diventa il punto $(0,\ 6)^T$ di valore 24 indipendente dal valore di δ.

Possiamo riassumere le considerazioni fatte fin qui nella seguente affermazione

> L'aumento della prima risorsa (r.h.s. primo vincolo) di un valore δ produce un aumento della funzione obiettivo pari a $u_1^*\delta = 3\delta$ purché $\delta \leq 3$.
> Il valore ottimo della variabile duale u_1^* rappresenta il massimo costo affrontabile per l'aumento di una unità di risorsa.

Il valore $u_1^* = 3$ è detto *prezzo ombra* relativo alla prima risorsa (al primo vincolo). L'intervallo $0 \leq \delta \leq 3$ è l'intervallo entro cui la previsione sulla variazione della funzione obiettivo è valida. (In realtà si può prevedere anche un decremento $-\delta$).

Variazioni coefficienti di costo. Se il valore ottimo di una variabile di decisione è nullo ($x_1^* = 0$) significa che produrre il prodotto corrispondente non è vantaggioso economicamente. In questo caso ci si può chiedere se e quando la produzione di x_1 può diventare vantaggiosa. Questo corrisponde a chiedere quanto deve variare il coefficiente della funzione obiettivo (cioè il profitto per unità di x_1 prodotte) perché diventi ottimo. Consideriamo allora la coppia primale-duale

$$(P)_\alpha \quad \begin{array}{ll} \max & (2+\alpha)x_1 + 6x_2 \\ & x_1 + 2x_2 \leq 5 \\ & x_1 + x_2 \leq 4 \\ & x_1, x_2 \geq 0 \end{array} \qquad \begin{array}{ll} \min & 5u_1 + 4u_2 \\ & u_1 + u_2 \geq 2 + \alpha \\ & 2u_1 + u_2 \geq 6 \\ & u_1, u_2 \geq 0 \end{array} \quad (D)_\alpha$$

dove $\alpha \geq 0$. Questa volta nel problema duale sta cambiando la regione ammissibile e in particolare il r.h.s. del primo vincolo duale, mentre la funzione obiettivo rimane invariata.
Quindi se vogliamo che il valore della funzione obiettivo rimanga invariato, significa che $u_\alpha^* = u^* = (3,\ 0)^T$.
Chiedere che sia $x_1^* > 0$ corrisponde a chiedere che il primo vincolo duale sia soddisfatto all'uguaglianza (per le condizioni di complementarità deve valere $x_1^*(2 + \alpha - u_1^* - 2u_2^*) = 0$). Geometricamente questo significa muovere il primo vincolo duale fino a quando non diventa attivo. In particolare il primo valore utile è quello in cui il primo vincolo passa per il punto $(3,\ 0)^T$. Quindi abbiamo $\alpha = 1$.
Questo corrisponde a modificare la funzione obiettivo del primale in $3x_1 + 6x_2$, le cui curve di livello sono parallele al primo vincolo primale. In effetti le soluzioni ottime del primale diventano tutte quelle sul segmento $(0,\ 5/2)^T$ e $(3, 1)^T$ di valore 15.

Il valore $\alpha = 1$ è detto *costo ridotto* relativo alla variabile x_1.

> La produzione di un prodotto diventa vantaggiosa se il coefficiente la funzione obiettivo viene aumentato del costo ridotto.
> Il valore ottimo della funzione obiettivo è invariato.

8.4.2 Interpretazione economica della dualità e prezzi ombra

Come visto nell'esempio, in generale, le variabili duali (i prezzi ombra) rappresentano l'effetto di cambiamenti nel termine noto dei vincoli. Si consideri, infatti un generico problema di Programmazione Lineare (in forma standard) (P), il suo duale (D) ed inoltre si consideri il problema (P_Δ) ottenuto modificando il termine noto da b a $b + \Delta$ (con $\Delta = (\delta_1 \ldots, \delta_m)^T \in \mathbb{R}^m$) e il corrispondente problema duale (D_Δ):

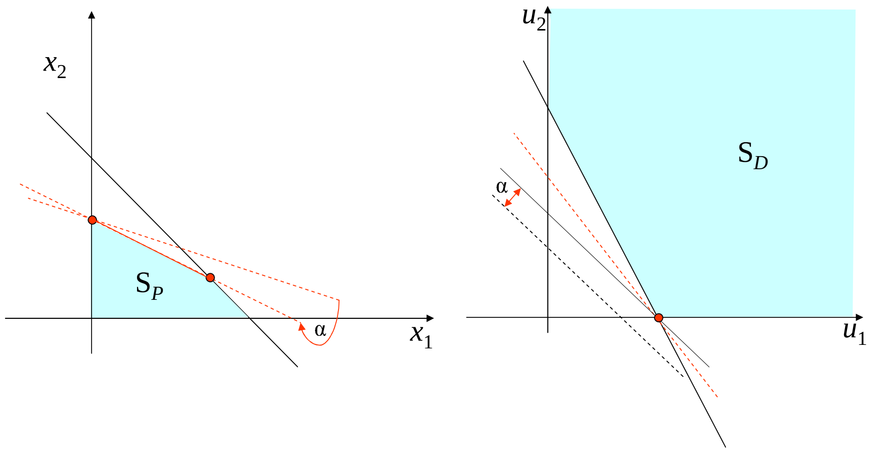

Figura 8.3: Soluzione grafica della coppia primale duale al variare di α

$$(\mathrm{P}) \qquad \begin{cases} \min c^T x \\ Ax = b \\ x \geq 0 \end{cases} \qquad\qquad (\mathrm{D}) \qquad \begin{cases} \max b^T u \\ A^T u \leq c \end{cases}$$

$$(\mathrm{P}_\Delta) \qquad \begin{cases} \min c^T x \\ Ax = b + \Delta \\ x \geq 0 \end{cases} \qquad\qquad (\mathrm{D}_\Delta) \qquad \begin{cases} \max (b+\Delta)^T u \\ A^T u \leq c \end{cases}$$

Siano $x^\star$ e $u^\star$ rispettivamente la soluzione ottima del problema (P) e del problema (D). Siano inoltre $x^\star(\Delta)$ e $u^\star(\Delta)$ rispettivamente la soluzione del problema (P_Δ) e del problema (D_Δ)

Dalle formulazioni di questi problemi si possono facilmente dedurre due osservazioni:

- la variazione del termine noto b nel problema primale si riflette in un cambiamento dei coefficienti della funzione obiettivo del problema duale;
- la regione ammissibile del problema (D) e del problema (D_Δ) sono uguali; da questo segue che se $u^\star \in \mathbb{R}^m$ è soluzione ottima del problema (D) allora $u^\star$ è ammissibile per il problema (D_Δ), ma non necessariamente è ottima per (D_Δ).

Inoltre per il Teorema della dualità forte applicato alla coppia primale–duale (P)–(D) deve essere

$$c^T x^\star = b^T u^\star, \tag{8.7}$$

mentre, sempre per il Teorema della dualità forte ma applicato alla coppia primale–duale (P_Δ)–(D_Δ) deve essere

$$c^T x^\star(\Delta) = (b + \Delta)^T u^\star(\Delta). \tag{8.8}$$

Se la soluzione ottima $x^\star$ soddisfa un'opportuna ipotesi (cioè che in $x^\star$ non ci siano più di n vincoli attivi) e se il vettore Δ ha componenti "sufficientemente" piccole allora si può dimostrare che:

$$u^\star(\Delta) = u^\star. \tag{8.9}$$

Utilizzando la (8.7), la (8.8) e la (8.9) si ha:

$$c^T x^\star(\Delta) = (b+\Delta)^T u^*(\Delta) = b^T u^\star + \Delta^T u^\star = c^T x^\star + \Delta^T u^\star, \tag{8.10}$$

che può essere riscritta nella seguente forma:

$$c^T x^\star(\Delta) - c^T x^\star = \delta_1 u_1^\star + \delta_2 u_2^\star + \ldots + \delta_m u_m^\star, \tag{8.11}$$

dove $\Delta = (\delta_1, \ldots, \delta_m)^T$.

Dalla precedente relazione segue che una possibile interpretazione della variabile duale $u_i^\star$ è quella di essere un prezzo associato ad un incremento unitario del termine noto b_i. Per questa ragione le variabili duali $u_i^\star$, $i = 1, \ldots, m$, vengono denominate *prezzi ombra*.

Il *prezzo ombra* di un vincolo rappresenta la variazione del valore della funzione obiettivo quando il r.h.s. del vincolo è aumentato di una unità, mentre tutti gli altri dati del problema sono fissati.

Sebbene la (8.9) (e di conseguenza la (8.11)) valga solamente sotto opportune ipotesi, in molte situazioni pratiche, le variabili duali $u_i^\star$, $i = 1, \ldots, m$, forniscono delle utili indicazioni su quale componente b_i variare per migliorare il valore ottimo della funzione obiettivo.

Ricordiamo i principi generali che regolano l'uso dei prezzi ombre.

1. I prezzi ombra sono tanti quanti i vincoli lineari generali (esclusi i vincoli di non negatività sulle variabili).
2. l'unità di misura dei prezzi ombra è espressa in (unità di misura funzione obiettivo)/(unità misura vincolo).
3. i costi ridotti corrispondono al valore delle variabili duali (all'ottimo), ovvero ai moltiplicatori di KKT relativi alla soluzione ottima primale.

Il duale del problema di allocazione ottima di risorse Si consideri nuovamente il semplice problema di allocazione ottima dell'Esempio 4.2.1 che è rappresentato dal seguente problema di Programazione Lineare:

$$\begin{array}{ll} \max & (7x_1 + 10x_2) \\ & x_1 + x_2 \leq 750 \\ & x_1 + 2x_2 \leq 1000 \\ & x_2 \leq 400 \\ & x_1 \geq 0, x_2 \geq 0. \end{array} \tag{8.12}$$

Ricordiamo che le variabili x_1 e x_2 sono associate rispettivamente ai quantitativi di colorante **C1** e **C2** da produrre e che la produzione avviene utilizzando tre preparati base **P1**, **P2** e **P3** dei quali si ha una disponibilità massima rispettivamente pari a 750, 1000 e 400 ettogrammi. Supponiamo, ora di voler sottrarre preparati base dalla produzione dei coloranti per venderli direttamente. Indichiamo con u_1, u_2 e u_3 i prezzi associati rispettivamente alla vendita diretta di un ettogrammo di preparato base **P1**, **P2** e **P3**. Supponendo di destinare tutti i preparati alla vendita diretta, il profitto che si otterrebbe sarebbe

$$750u_1 + 1000u_2 + 400u_3. \tag{8.13}$$

Naturalmente si vorrà fare in modo che questa operazione di sottrazione dei preparati base dalla produzione dei coloranti e vendita diretta risulti economicamente conveniente e quindi mentre si vuole minimizzare l'espressione (8.13) affinché i prezzi di vendita risultino competitivi sul mercato, si imporrà che il profitto ottenuto vendendo direttamente i quantitativi di preparato base necessario per ottenere un litro di colorante sia maggiore o uguale del profitto associato alla vendita di un litro di colorante stesso; quindi, utilizzando i dati del problema riportati nella tabella dell'Esempio 4.2.1, si deve imporre che risulti

$$u_1 + u_2 \geq 7$$

per quanto riguarda il colorante **C1** e

$$u_1 + 2u_2 + u_3 \geq 10$$

per quanto riguarda il colorante **C2** e naturalmente deve essere $u_1 \geq 0$, $u_2 \geq 0$ e $u_3 \geq 0$. Quindi il modello lineare che rappresenta l'operazione sopra descritta è il seguente:

$$\begin{array}{ll} \min & (750u_1 + 1000u_2 + 400u_3) \\ & u_1 + u_2 \geq 7 \\ & u_1 + 2u_2 + u_3 \geq 10 \\ & u_1 \geq 0, u_2 \geq 0, u_3 \geq 0. \end{array}$$

Esaminando questo problema si vede immediatamente che esso rappresenta il *problema duale* del problema dato (8.12).

In generale, se si considera un generico problema di allocazione ottima di m risorse $\mathbf{R_i}$, $i = 1, \ldots, m$ con la possibilità di fabbricare n prodotti $\mathbf{P_j}$, $j = 1, \ldots, n$, come abbiamo già esaminato nel capitolo precedente si può formulare questo problema come

$$\begin{array}{ll} \max & c^T x \\ & Ax \leq b \\ & x \geq 0 \end{array} \tag{8.14}$$

dove ricordiamo $x \in \mathbb{R}^n$ è il vettore avente per componenti i livelli di produzione di ciascuno dei prodotti, $c \in \mathbb{R}^n$ il vettore dei profitti netti e $b \in \mathbb{R}^m$ il vettore delle disponibilità massima di ciascuna delle risorse.

Supponiamo ora di voler sottrarre risorse alla produzione per venderle direttamente e siano u_i, $i = 1, \ldots, m$ i prezzi unitari associati alla vendita dell'i-esima risorsa. Supponendo che per ciascuna risorsa si voglia destinare alla vendita una quantità pari alla disponibilità massima di quella risorsa, si ottiene un profitto pari a

$$b_1u_1 + b_2u_2 + \cdots + b_mu_m.$$

Per rendere competitivi sul mercato i prezzi unitari u_i da assegnare alle risorse vendute direttamente, si vogliono scegliere i valori piú bassi possibile per le u_i, ma naturalmente, affinché questa operazione di vendita diretta in luogo della fabbricazione dei prodotti risulti conveniente si deve imporre che il profitto ottenuto vendendo direttamente le risorse necessarie per fabbricare un prodotto sia maggiore o uguale al profitto che si ricaverebbe dalla vendita del prodotto finito. Quindi per ogni prodotto, si deve imporre che valga

$$\begin{array}{cccc} a_{11}u_1 + & \ldots & +a_{m1}u_m & \geq c_1 \\ a_{12}u_1 + & \ldots & +a_{m2}u_m & \geq c_2 \\ \vdots & \ldots & \vdots & \vdots \\ a_{1n}u_1 + & \ldots & +a_{mn}u_m & \geq c_n \end{array}$$

con $u_i \geq 0$, $i = 1, \ldots, m$ e dove le quantità a_{ij} rappresentano la quantità di risorsa $\mathbf{R_i}$ necessaria per fabbricare una unità di prodotto $\mathbf{P_j}$.

Quindi il problema da risolvere può essere scritto nella forma

$$\begin{array}{ll} \min & b^T u \\ & A^T u \geq c \\ & u \geq 0 \end{array}$$

che è il problema duale del problema (8.14).

Il duale del problema di miscelazione. Si consideri il problema di miscelazione dell'Esempio 4.2.2 che è rappresentato dal seguente problema di Programmazione Lineare:

$$\begin{array}{ll} \min & 400x_1 + 600x_2 \\ & 140x_1 \geq 70 \\ & 20x_1 + 10x_2 \geq 30 \\ & 25x_1 + 50x_2 \geq 75 \\ & x_1 \geq 0, x_2 \geq 0. \end{array}$$

Ricordiamo che le variabili x_1 e x_2 rappresentano le quantità di polpa di frutta e di dolcificante da utilizzare nella produzione del succo di frutta che deve avere come requisito un contenuto minimo di 70 mg di vitamina C, 30 mg di sali minerali e 75 grammi di zucchero. Supponiamo ora che un'industria farmaceutica venda compresse di nutrimenti puri, cioè compresse di vitamina C, di sali minerali e di zucchero e che vuole immettere queste compresse su un ipotetico mercato come offerta sostitutiva al succo di frutta per l'acquisizione di vitamina C, di sali minerali e di zucchero. Naturalmente questa industria farmaceutica vuole massimizzare il profitto ricavato dalla vendita delle compresse, ma al tempo stesso deve dare un prezzo alle compresse tale da essere competitiva. Siano allora u_1, u_2 e u_3 i prezzi di vendita rispettivamente di 1 mg di vitamina C, di 1 mg di sali minerali e di 1 grammo di zucchero; supponendo che la vendita di questi nutrimenti puri sia pari ai fabbisogni minimi (cioè a 70 mg di vitamina C, a 30 mg di sali minerali e a 75 grammi di zucchero), l'espressione del profitto dell'industria farmaceutica che dovrà essere massimizzata è

$$70u_1 + 30u_2 + 75u_3.$$

Affinché i prezzi di vendita dei componenti puri in compresse fissati dall'industria siano concorrenziali, si deve imporre che il costo unitario dei nutrimenti puri sia minore o uguale al prezzo che si dovrebbe pagare per avere la stessa quantità di componente attraverso gli ingredienti del succo di frutta, cioè dalla polpa di frutta e dal dolcificante. Quindi si devono imporre i seguenti vincoli

$$\begin{array}{rcl} 140u_1 + 20u_2 + 25u_3 & \leq & 400 \\ 10u_2 + 50u_3 & \leq & 600. \end{array}$$

Inoltre dovrà essere $u_1 \geq 0$, $u_2 \geq 0$, $u_3 \geq 0$.

Quindi il problema complessivo formulato dall'industria farmaceutica è

$$\begin{array}{ll} \max & 70u_1 + 30u_2 + 75u_3 \\ & 140u_1 + 20u_2 + 25u_3 \leq 400 \\ & 10u_2 + 50u_3 \leq 600 \\ & u_1 \geq 0, u_2 \geq 0, u_3 \geq 0 \end{array}$$

che è il problema duale del problema di miscelazione considerato.

In generale, consideriamo un generico problema di miscelazione in cui si hanno n sostanze $\mathbf{S_j}$, $j = i, \ldots, n$ ciascuna delle quali contiene una quantità a_{ij} di componente utile $\mathbf{C_i}$, $i = 1, \ldots, m$. Come abbiamo già esaminato nel capitolo precedente un problema di miscelazione di questo tipo si può formulare come

$$\begin{array}{ll} \min & c^T x \\ & Ax \geq b \\ & x \geq 0 \end{array} \tag{8.15}$$

dove ricordiamo che $x \in \mathbb{R}^n$ è il vettore avente per componenti le quantità di ciascuna sostanza da introdurre nella miscela, $c \in \mathbb{R}^n$ il vettore dei costi unitari delle sostanze, $b \in \mathbb{R}^m$ il vettore dei requisiti minimi di componenti utili da introdurre nella miscela, e $A \in \mathbb{R}^{m\times n}$ la matrice i cui elementi sono le a_{ij}.

Supponiamo ora che un'industria sia in grado di fornire componenti utili allo stato puro e che voglia immettere sul mercato questi componenti utili e siano u_i, $i = 1,\dots,m$ i prezzi unitari da assegnare a ciascuno di essi. Supponendo che la richiesta del mercato sia pari ai fabbisogni minimi della miscela, cioè per ciascun componente pari a b_i, il profitto totale dell'industria che vende i componenti utili allo stato puro è

$$b_1u_1 + b_2u_2 + \cdots + b_mu_m.$$

Inoltre, affinché i prezzi u_i assegnati dall'industria ai componenti puri siano concorrenziali, si deve imporre che il costo dei componenti puri sia minore o uguale al prezzo che dovrebbe pagare per avere la stessa quantità di componente ottenuto attraverso le sostanze e quindi deve valere

$$\sum_{i=1}^{m} a_{ij}u_i \leq c_j, \quad j = 1,\dots,n.$$

Inoltre si deve imporre $u_i \geq 0$, $i = 1,\dots,m$.

Quindi il problema formulato si può scrivere nella forma

$$\begin{array}{ll} \max & b^T u \\ & A^T u \leq c \\ & u \geq 0 \end{array}$$

che è immediato verificare essere il problema duale del problema di miscelazione assegnato (8.15).

Il duale del problema dei trasporti Si consideri un problema di trasporto che è rappresentato dal seguente problema di Programmazione Lineare:

$$\min(250x_1 + 100x_2 + 85x_3 \;+\; 120y_1 + 80y_2 \;+\; 150y_3)$$

$$\begin{array}{ccccccccccccc} x_1 & + & x_2 & + & x_3 & & & & & & & = & 50 \\ & & & & & & y_1 & + & y_2 & + & y_3 & = & 55 \\ x_1 & + & & & & & y_1 & & & & & = & 30 \\ & & x_2 & + & & & & & y_2 & & & = & 40 \\ & & & & x_3 & + & & & & & y_3 & = & 35 \end{array}$$

$$x_1 \geq 0, \;\; x_2 \geq 0, \; x_3 \geq 0, \; y_1 \geq 0, \; y_2 \geq 0, \; y_3 \geq 0.$$

dove x_1, x_2, x_3 rappresentano rispettivamente quantità di acqua da trasportare dal primo stabilimento a tre impianti e y_1, y_2, y_3 rappresentano rispettivamente la quantità di acqua da trasportare dal secondo stabilimento a tre impianti. Ricordiamo inoltre che la disponibilità giornaliera di acqua presso i due stabilimenti è rispettivamente di 50 e 55 ettolitri di acqua, mentre le richieste giornaliere di acqua presso i tre impianti sono rispettivamente di 30, 40 e 35 ettolitri.

Supponiamo ora che una compagnia specializzata in trasporto di acqua (esterna all'industria) voglia proporsi all'industria di acque minerali per effettuare il trasporto dell'acqua dagli stabilimenti agli impianti. Naturalmente la compagnia di trasporti, per convincere l'industria di acque minerali ad avvalersi del servizio di trasporto esterno, dovrà proporre dei prezzi di trasporto vantaggiosi. A tale scopo la compagnia dei trasporti propone all'industria di prelevare un ettolitro di acqua da ciascuno dei due stabilimenti per un prezzo unitario (in migliaia di lire) rispettivamente pari a u_1 e u_2 e di consegnare un ettolitro di acqua a ciascuno dei tre impianti per un prezzo unitario (in migliaia di lire) rispettivamente pari a v_1, v_2 e v_3. Quindi la compagnia dei trasporti vorrà massimizzare la funzione che fornisce il suo profitto complessivo che è data da

$$50u_1 + 55u_2 + 30v_1 + 40v_2 + 35v_3.$$

Tuttavia affinché l'offerta della compagnia dei trasporti risulti vantaggiosa per l'industria delle acque minerali i prezzi del trasporto proposti dovranno risultare non superiori a quelli che l'industria avrebbe effettuando in proprio i trasporti stessi. Quindi dovrà risultare

$$\begin{array}{rcl}
u_1 + v_1 & \leq & 250 \\
u_1 + v_2 & \leq & 100 \\
u_1 + v_3 & \leq & 85 \\
u_2 + v_1 & \leq & 120 \\
u_2 + v_2 & \leq & 80 \\
u_2 + v_3 & \leq & 150.
\end{array}$$

Quindi, la compagnia dei trasporti dovrà risolvere il problema

$$\begin{array}{lccccccccccc}
\max & (50u_1 & + & 55u_2 & + & 30v_1 & + & 40v_2 & + & 35v_3) & & \\
& u_1 & & & & +v_1 & & & & & \leq & 250 \\
& u_1 & & & & & & +v_2 & & & \leq & 100 \\
& u_1 & & & & & & & & +v_3 & \leq & 85 \\
& & & u_2 & & +v_1 & & & & & \leq & 120 \\
& & & u_2 & & & & +v_2 & & & \leq & 80 \\
& & & u_2 & & & & & & +v_3 & \leq & 150
\end{array}$$

che si verifica immediatamente essere il problema duale del problema dei trasporti assegnato.

In generale, consideriamo ora un generico problema dei trasporti già esaminato nel capitolo precedente. Supponiamo che un'azienda voglia provvedere in proprio ad effettuare il trasporto di materiali e che quindi cerchi di risolvere il problema dei trasporti

$$\begin{array}{ll}
\min & \left(\displaystyle\sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \right) \\
& \displaystyle\sum_{j=1}^{n} x_{ij} = a_i \qquad i = 1, \ldots, m \\
& \displaystyle\sum_{i=1}^{m} x_{ij} = b_j \qquad j = 1, \ldots, n \\
& x_{ij} \geq 0 \qquad i = 1, \ldots, n; \quad j = 1, \ldots, m
\end{array} \tag{8.16}$$

dove, ricordiamo, che le c_{ij} rappresentano il costo del trasporto dall'origine i alla destinazione j, le a_i le disponibilità all'i-esima origine e le b_j le richieste alla j-esima destinazione.

Supponiamo, ora che una compagnia che esegue trasporti voglia proporsi a questa azienda, come alternativa vantaggiosa all'effettuazione dei trasporti in proprio; a tale scopo questa compagnia propone all'azienda di prelevare un'unità di prodotto dall'origine i per un prezzo unitario u_i e di consegnare una unità di prodotto alla destinazione j per un prezzo unitario v_j. Per assicurare che i suoi prezzi siano competitivi rispetto a quelli che l'azienda avrebbe effettuando i trasporti in proprio, la compagnia di trasporti deve fare sì che risulti

$$u_i + v_j \leq c_{ij}$$

per ogni $i = 1, \ldots, m$ e $j = 1, \ldots, n$. D'altra parte la compagnia di trasporti vuole scegliere i prezzi da proporre $u_1, \ldots, u_m$ e $v_1, \ldots, v_n$ in modo da massimizzare il suo profitto complessivo. Poiché le quantità a_i e b_j di prodotto rispettivamente disponibili all'origine i e richieste alla destinazione j sono note alla compagnia di trasporti, questa cercherà di massimizzare la funzione

$$\max \left(\sum_{i=1}^{m} a_i u_i + \sum_{j=1}^{n} b_j v_j \right).$$

Quindi il problema che la compagnia di trasporti formula per determinare quali prezzi u_i e v_j proporre all'azienda è il seguente

$$\begin{array}{ll} \max & \sum_{i=1}^{m} a_i u_i + \sum_{j=1}^{n} b_j v_j \\ & u_i + v_j \leq c_{ij} \qquad i = 1, \ldots, m, \qquad j = 1, \ldots, n \end{array} \tag{8.17}$$

che è il problema duale del problema dei trasporti (8.16).