Dispensa del Corso di Complementi di Ricerca Operativa

Marco Trubian*

Milano, Marzo 2008

*Dipartimento di Scienze dell'Informazione, Università degli Studi di Milano, Via Comelico 39, 20135 Milano, e-mail: trubian@dsi.unimi.it

Indice

Elenco delle figure

1 Notazione

$\mathbb{R}^n$ = insieme dei vettori reali n-dimensionali
$\mathbb{R}^n_+$ = insieme dei vettori reali n-dimensionali non negativi
$\mathbb{Z}^n$ = insieme dei vettori interi n-dimensionali
$\mathbb{Z}^n_+$ = insieme dei vettori interi n-dimensionali interi non negativi
$[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$ = intervallo chiuso
$(a, b) = \{x \in \mathbb{R} : a < x < b\}$ = intervallo aperto
$\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ = vettore colonna ad n componenti ($\boldsymbol{x} \in \mathbb{R}^n$)
$\boldsymbol{x}^T = [x_1, x_2, \ldots, x_n]$ = vettore riga ad n componenti ($\boldsymbol{x} \in \mathbb{R}^n$)
$e_i^T = (0, 0, \ldots, 1, 0, \ldots, 0)$ i-esimo versore: vettore con un 1 in posizione i-esima e 0 altrove
$\|x\| = \sqrt{\sum_{j=1}^n x_j^2}$ = norma euclidea del vettore ($\boldsymbol{x} \in \mathbb{R}^n$)
$\mathrm{A} = \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \ldots & a_{mn} \end{bmatrix}$ = matrice $m \times n$
$\arg\min\{f(\boldsymbol{x}) : x \in X\} = \boldsymbol{x}^* \in \mathbb{R}^n$ tale che $f(\boldsymbol{x}^*) = \min\{f(\boldsymbol{x}) : x \in X\}$
$I(x, \varepsilon) = \{y : \|y - x\| < \varepsilon\}$ = intorno aperto di raggio $\varepsilon > 0$ di x
$\frac{\partial f}{\partial x_i}$ derivata parziale prima rispetto a x_i
∇ operatore di derivata parziale
$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \ldots & \frac{\partial f}{\partial x_n} \end{bmatrix}^T$ vettore gradiente
∇^2 operatore di derivata seconda (elementi $\frac{\partial^2 f}{\partial x_i \partial x_j}$)
$\nabla^2 f(\boldsymbol{x}) = H(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ & \vdots & & \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$ matrice Hessiana
$\frac{\partial \boldsymbol{q}}{\partial \boldsymbol{x}} = \boldsymbol{J}$ = matrice Jacobiana $n \times k = [\nabla q_1, \nabla q_2, \ldots, \nabla q_k]$ dove $q_j : \mathbb{R}^n \to \mathbb{R}$, $j = 1, \ldots, k$

2 Introduzione

L'obiettivo di questa dispensa è quello di fornire una introduzione agli algoritmi per l'ottimizzazione numerica per problemi nel continuo.
Molte applicazioni possono essere formulate come problemi di ottimizzazione nel continuo; per esempio

- trovare la traiettoria ottima di un velivolo o del braccio di un robot;
- identificare le proprietà sismiche di un pezzo di crosta terrestre adattando un modello della regione studiata ad un insieme di valori campionati da una rete di stazioni di rilevamento;
- comporre un portafoglio di investimenti che massimizzi il ritorno atteso nel rispetto di un accettabile livello di rischio;
- calcolare la forma ottimale di un componente di una automobile, di un aereomobile o dello scafo di una imbarcazione;
- determinare la conformazione spaziale di una proteina a partire dalla sequenza di aminoacidi che la compongono.

Formulazione matematica Formalmente, l'ottimizzazione consiste nella massimizzazione o minimizzazione di una funzione soggetta a vincoli sulle sue variabili. La notazione utilizzata è la seguente:

- $\boldsymbol{x}$ è il vettore delle *variabili* o delle *incognite*;
- $f(\boldsymbol{x})$ è la *funzione obiettivo* che vogliamo massimizzare o minimizzare;
- $g(\boldsymbol{x})$ sono *vincoli di disuguaglianza*, funzioni scalari di $\boldsymbol{x}$ che definiscono delle *disequazioni* che il vettore delle variabili deve soddisfare;
- $h(\boldsymbol{x})$ sono *vincoli di uguaglianza*, funzioni scalari di $\boldsymbol{x}$ che definiscono delle *equazioni* che il vettore delle variabili deve soddisfare;
- X è la *regione ammissibile*, cioè il luogo dei punti dove il vettore delle variabili soddisfa tutti i vincoli, di disuguaglianza o di uguaglianza.

Con questa notazione il problema di ottimizzazione può venir così formulato:

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ t.c. & \boldsymbol{x} \in X \end{array}$$

dove $X = \{\boldsymbol{x} \in \mathbb{R}^n : g_i(\boldsymbol{x}) \leq 0 \quad i = 1, \ldots, k; \quad h_j(\boldsymbol{x}) = 0 \quad j = 1, \ldots, h\}$
Nel caso in cui $X \subset \mathbb{R}^n$, cioè X è un sottoinsieme proprio di $\mathbb{R}^n$, si parla di *ottimizzazione vincolata*, mentre quando vale $X \equiv \mathbb{R}^n$ si parla di *ottimizzazione non vincolata*.
Quando almeno una delle funzioni $f(\boldsymbol{x}), g(\boldsymbol{x}), h(\boldsymbol{x})$ è non lineare si parla di *Programmazione Non Lineare* (PNL).

2.1 Esempi

Consideriamo il problema di determinare il punto $P = (x_1, x_2)$ più vicino al punto di coordinate $(2, 1)$, con la richiesta che P deve appartenere all'intersezione della superficie interna alla parabola di equazione $x_2 = x_1^2$ con l'area del triangolo di estremi $(0, 0)$, $(2, 0)$ e $(0, 2)$. Il problema può essere modellizzato nel seguente modo

$$\begin{array}{ll} \min & (x_1 - 2)^2 + (x_2 - 1)^2 \\ t.c. & x_1^2 - x_2 \leq 0 \\ & x_1 + x_2 \leq 2 \\ & x_1,\ x_2 \geq 0 \end{array}$$

Esso può essere visto come un problema di PNL leggendo $f(\boldsymbol{x}) = (x_1 - 2)^2 + (x_2 - 1)^2$ ed i vincoli di disuguaglianza come: $g_1(\boldsymbol{x}) = x_1^2 - x_2$, $g_2(\boldsymbol{x}) = x_1 + x_2 - 2$, $g_3(\boldsymbol{x}) = -x_1$ e $g_4(\boldsymbol{x}) = -x_2$. La Figura 2.1 rappresenta in chiaro la regione ammissibile del problema, e,

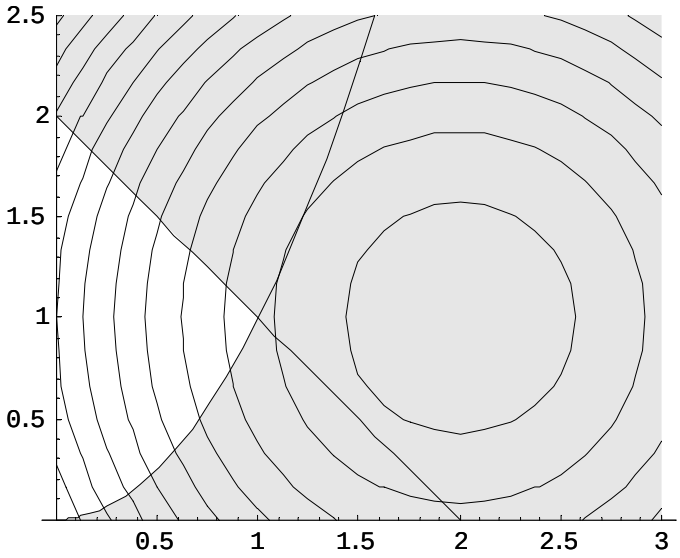


Figura 1: Esempio in due dimensioni

come cerchi concentrici, le curve di livello della funzione obiettivo.

Anche i problemi di Programmazione Lineare Intera (PLI) possono essere visti come problemi di PNL. Cominciamo con la Programmazione Lineare Binaria (PLB):

$$\begin{array}{lll} 1) & \min & \boldsymbol{c}^T \boldsymbol{x} \\ & t.c. & A\boldsymbol{x} = \boldsymbol{b} \\ & & \boldsymbol{x} \in \{0, 1\}^n \end{array}$$

Essa è riconducibile alla PNL mediante, ad esempio, la seguente trasformazione

$$\begin{array}{lll} 2) & \min & \boldsymbol{c}^T \boldsymbol{x} + M\boldsymbol{x}^T(\boldsymbol{1} - \boldsymbol{x}) \\ & t.c. & A\boldsymbol{x} = \boldsymbol{b} \\ & & \boldsymbol{x} \in [0, 1]^n \end{array}$$

I problemi 1) e 2) sono equivalenti. Nella funzione obiettivo del problema 2) abbiamo aggiunto un valore di penalità, $M \gg 0$, grande a sufficienza. Ciascuna componente $Mx_i(1-x_i)$ assume il suo valore massimo in corrispondenza del valore $x_i = 0.5$, mentre tende a zero al tendere della variabile x_i a 0 oppure a 1.
Il problema 2) è un caso particolare di programmazione quadratica, dove le variabili hanno esponente di grado 2 o 1 ed i vincoli sono lineari. La programmazione lineare binaria è un caso particolare di lineare intera ma è anche un caso particolare di programmazione quadratica. In alternativa si può scrivere

$$\begin{array}{lll} 3) & \min & \boldsymbol{c}^T\boldsymbol{x} \\ & t.c. & A\boldsymbol{x} = \boldsymbol{b} \\ & & \boldsymbol{x}^T(\mathbf{1} - \boldsymbol{x}) = \mathbf{0} \\ & & \boldsymbol{x} \in [0,1]^n \end{array}$$

I problemi 1) e 3) sono equivalenti. Poiché le variabili sono non negative l'unico modo per soddisfare il vincolo non lineare $\boldsymbol{x}^T(\mathbf{1}^T - \boldsymbol{x}) = \mathbf{0}$, è richiedere che ogni componente $x_i(1 - x_i)$ sia uguale a zero. Questo si verifica solo quando ciascuna variabile x_i assume il valore 0 oppure 1. Notiamo, infine, che anche la Programmazione Lineare Intera (PLI):

$$\begin{array}{lll} 4) & \min & \boldsymbol{c}^T\boldsymbol{x} \\ & t.c. & A\boldsymbol{x} = \boldsymbol{b} \\ & & \boldsymbol{x} \in Z^n \end{array}$$

è riconducibile alla PNL mediante, ad esempio, la seguente trasformazione

$$\begin{array}{lll} 5) & \min & \boldsymbol{c}^T\boldsymbol{x} \\ & t.c. & A\boldsymbol{x} = \boldsymbol{b} \\ & & \sin(\pi x_j) = 0, \quad j = 1, \ldots, n \end{array}$$

Queste semplici trasformazioni mettono in evidenza come i problemi di PNL costituiscano una generalizzazione dei problemi incontrati nell'ambito della programmazione lineare e lineare intera e che pertanto siano in generale più difficili da risolvere (frequentemente, molto più difficili da risolvere) di quanto non lo siano i pur difficili problemi di PLI.

2.2 Gli algoritmi di ottimizzazione

In generale, non è possibile identificare per via analitica le soluzioni ottime dei problemi di PNL. È possibile, come vedremo nelle sezioni 4 e 4.1, fornire, in qualche caso, le condizioni analitiche che devono essere soddisfatte dalle soluzioni ottime, ed in alcuni, fortunati casi, determinare una soluzione ottima proprio imponendo tali condizioni. Di conseguenza, si faranno di volta in volta delle assunzioni per specializzare le funzioni $f(\boldsymbol{x}), g_i(\boldsymbol{x})$, e $h_i(\boldsymbol{x})$, (continuità, differenziabilità, convessità, ecc.) in modo da dividere i problemi in classi, applicando ai problemi di ciascuna classe i metodi più opportuni. Nella maggior parte dei casi quindi, la

via analitica non permette di fornire una soluzione, anche se le proprietà analitiche possono ancora tornare utili per dire se una data soluzione, in qualche modo generata, è o meno una soluzione ottima. In generale quindi, è necessario ricorrere a tecniche di tipo algoritmico.

Gli algoritmi di ottimizzazione sono iterativi. Iniziano generando un vettore di variabili $\boldsymbol{x}_0$, non necessariamente appartenente alla regione ammissibile X, e generano iterativamente una sequenza di vettori $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$ che sono tutte (sperabilmente) approssimazioni migliori di una soluzione ottima.
Abbiamo già incontrato questa tecnica quando abbiamo risolto i problemi di Programmazione Lineare. L'*algoritmo del Simplesso* è un tipico algoritmo di tipo iterativo che, partendo da una soluzione ammissibile, si sposta in soluzioni vicine alla soluzione corrente, arrestandosi quando nessuna di tali soluzioni migliora il valore della funzione obiettivo corrente e sono soddisfatte particolari condizioni analitiche, che garantiscono di trovarsi nella soluzione ottima. Uno schema analogo viene adottato dagli algoritmi che vedremo nel seguito, con una importante differenza però: nella maggior parte dei casi tali algoritmi si arrestano identificando solo un *minimo locale* del problema, una soluzione che è la migliore nella porzione di regione ammissibile che la circonda, ma che potrebbe essere anche molto diversa (e peggiore) dalla soluzione ottima del problema (cfr. la Sezione 2.4).

Le strategie adottate per passare da un vettore $\boldsymbol{x}_k$ al successivo $\boldsymbol{x}_{k+1}$ fanno la differenza fra un algoritmo ed un altro. La maggior parte degli algoritmi usano i valori della funzione f, delle funzioni g_i ed h_i e, quando conveniente, delle derivate prime e seconde di tali funzioni. Alcuni algoritmi usano l'informazione raccolta durante il processo iterativo di ricerca, mentre altri usano solo informazione locale al vettore $\boldsymbol{x}$ corrente. In ogni caso i buoni algoritmi dovrebbero possedere le seguenti caratteristiche:

- Robustezza. Dovrebbero comportarsi bene su una varietà di problemi della classe per cui sono stati progettati a partire da qualsiasi ragionevole vettore iniziale $\boldsymbol{x}_0$.
- Efficienza. Non dovrebbero richiedere un eccessivo tempo di calcolo o spazio di memoria.
- Accuratezza. Dovrebbero identificare una soluzione con precisione, senza risultare eccessivamente sensibili ad errori nei dati o agli inevitabili errori di arrotondamento che si verificano quando l'algoritmo viene implementato.

Queste caratteristiche sono spesso conflittuali, come vedremo, e occorrerà scendere a compromessi.

2.3 Nozioni di convessità

In ottimizzazione la nozione di convessità è fondamentale. Molti problemi possiedono tale proprietà che in genere li rende più facili da risolvere sia in teoria che in pratica. Il termine convesso può essere applicato sia a insiemi che a funzioni.

> **Definizione 1** *Insieme convesso. Un insieme* $\mathrm{X} \subset \mathbb{R}^n$ *è convesso se comunque presi due punti* $\boldsymbol{x}, \boldsymbol{y} \in \mathrm{X}$*, allora* $\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y} \in \mathrm{X}$*, per ogni* $\lambda \in [0,1]$.

Cioè ogni punto nel segmento che congiunge due qualsiasi punti in X è contenuto in X.

Definizione 2 *Funzione convessa. Una funzione $f : \mathbb{R}^n \to \mathbb{R}$ è convessa se il suo dominio è un insieme convesso $\mathrm{X} \subseteq \mathbb{R}^n$ e comunque presi due punti $\boldsymbol{x}, \boldsymbol{y} \in \mathrm{X}$ vale la relazione:*

$$f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leq \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}) \text{ per ogni } \lambda \in [0,1].$$

In altre parole, il valore della funzione sui punti del segmento che unisce $\boldsymbol{x}$ a $\boldsymbol{y}$ è minore o uguale alla combinazione convessa dei valori che assume la funzione nei punti $\boldsymbol{x}$ e $\boldsymbol{y}$.
Le funzioni convesse e gli insiemi convessi rivestono un ruolo importante nella PNL ed in particolare lo rivestono i problemi dove si minimizza una funzione obiettivo convessa su una regione ammissibile convessa.

Definizione 3 *Un problema di ottimizzazione con funzione obiettivo e regione ammissibile entrambe convesse viene detto problema convesso.*

I problemi convessi sono importanti perché, in genere, più semplici da risolvere. Quindi è importante capire come si può stabilire se una funzione o un insieme sono convessi.
Relativamente agli insiemi convessi valgono le seguenti proprietà:

Proprietà 1 *Siano X e Y due insiemi convessi ed $\alpha > 0$ uno scalare allora*
a) l'insieme $\alpha \mathrm{X}$ è convesso;
b) l'insieme $X + Y$ è convesso;
c) l'insieme $X \bigcap Y$ è convesso.

Relativamente alle funzioni convesse valgono le seguenti proprietà:

Proprietà 2 *Siano $f()$ e $g()$ due funzioni convesse ed $\alpha > 0$ uno scalare allora*
a) la funzione $\alpha f()$ è convessa;
b) la combinazione lineare $f() + g()$ è convessa;
c) la funzione $\max\{f(), g()\}$ è convessa;
d) il luogo dei punti $\boldsymbol{x}$ per i quali vale $f(\boldsymbol{x}) \leq \alpha$ è convesso.

Tali proprietà ci aiutano ad identificare la convessità di una funzione complessa dall'analisi delle sue componenti.
Una immediata e importante conseguenza della Proprietà 1.c e della 2.d è la seguente

Proprietà 3 *Un problema di programmazione non lineare:*

$$\begin{array}{lll} \min & & f(\boldsymbol{x}) \\ t.c. & g_i(\boldsymbol{x}) \leq 0, & i = 1, \ldots, k \\ & h_i(\boldsymbol{x}) = 0, & i = 1, \ldots, h \end{array}$$

è convesso se $f(\boldsymbol{x})$ è convessa, le $g_i(\boldsymbol{x})$ sono convesse e le $h_i(\boldsymbol{x})$ sono lineari.

Osserviamo infine, che la funzione $\max\{f(), g()\}$ è convessa se $f()$ e $g()$ sono convesse, ma la funzione $\max\{f(), g()\}$ non è in genere differenziabile anche se $f()$ e $g()$ lo sono. La richiesta che la funzione obiettivo $f()$ sia differenziabile è fondamentale per la maggior parte dei metodi che vedremo e la trattazione di funzioni non differenziabili risulta molto più complessa del caso di funzioni differenziabili.

2.4 Minimi locali e globali

Data una funzione $f : X \to \mathbb{R}$, con $X \subseteq \mathbb{R}^n$, in generale, ciò che desideriamo è trovare un *ottimo globale* di $f(\boldsymbol{x})$, un punto in cui la funzione assume il suo minimo valore. Formalmente,

Definizione 4 *Un punto $\boldsymbol{x}^* \in X$ è un punto di minimo globale di $f(\boldsymbol{x})$ se:*

$$f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}) \quad \forall \boldsymbol{x} \in X$$

Poiché i nostri algoritmi iterativi non visitano (chiaramente) tutti i punti di X, né (sperabilmente) nemmeno molti, non si è sicuri che la funzione $f(\boldsymbol{x})$ non prenda una ripida discesa in una delle regioni che non sono state visitate dall'algoritmo. La maggior parte degli algoritmi sono quindi in grado di trovare solo *minimi locali*, cioè punti che assumono il valore minimo fra tutti quelli del loro *intorno*. Formalmente,

Definizione 5 *Un punto $\boldsymbol{x}^* \in X$ è un punto di minimo locale di $f(\boldsymbol{x})$ se esiste un intorno aperto $I(\boldsymbol{x}^*, \varepsilon)$ di $\boldsymbol{x}^*$ avente raggio $\varepsilon > 0$ tale che:*

$$f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}) \quad \forall x \in X \cap I(\boldsymbol{x}^*, \varepsilon)$$

Un punto è detto di minimo *in senso stretto* se al posto di $\leq$ abbiamo $<$ nelle precedenti definizioni.
Come detto la maggior parte delle proprietà analitiche e delle tecniche algoritmiche hanno come finalità quella di determinare punti di minimo locale della funzione $f(\boldsymbol{x})$.

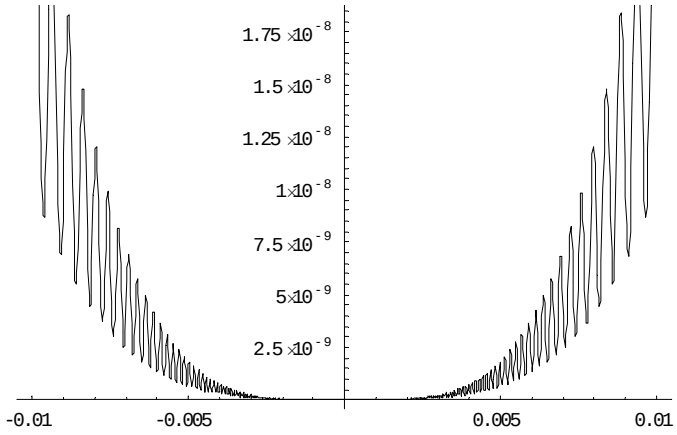


Figura 2: Curva con minimo globale non isolato

La figura illustra l'andamento della funzione

$$f(x) = x^4 \cos(1/x) + 2x^4$$

due volte differenziabile con continuità e che ha un minimo globale in $x^* = 0$ e minimi locali in senso stretto in molti punti prossimi ad esso. Questo esempio, benché costruito ad arte, non è però forzato in quanto, nei problemi di ottimizzazione associati con la determinazione di conformazioni molecolari, la funzione potenziale, che deve essere minimizzata, può avere

un numero di minimi locali che è esponenziale rispetto al numero di variabili. Consideriamo il seguente esempio di ottimizzazione non vincolata definito in $\mathbb{R}^3$.

$$\min f(\boldsymbol{x}) = \sum_{i=1}^{N} \sum_{j=1}^{i-1} \left(\frac{1}{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^{12}} - \frac{2}{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^{6}} \right)$$

con $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N \in \mathbb{R}^3$. La funzione $f(\boldsymbol{x})$ rappresenta l'energia potenziale di un raggruppamento di atomi identici, centrati nei punti $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$, quando esista una forza di attrazione/repulsione fra ogni loro coppia. Si è potuto verificare sperimentalmente che questo modello rappresenta con particolare accuratezza gli stati di energia di alcuni gas ideali, quali il *kripto* (dal greco *kryptos*, nascosto, per la difficoltà di ottenere l'elemento in presenza di aria) e l'*argo* (dal greco *argos*, inattivo, poiché l'elemento non è reattivo) e di alcuni metalli. Si ritiene che la configurazione a minima energia, cioè la soluzione ottima del problema sia quella più stabile in natura.
La soluzione di questo, apparentemente semplice problema, è in realtà estremamente complessa, tenuto conto che si stima esistano $O(e^N)$ configurazioni che corrispondono a minimi locali, ma non globali, di f.

In questi casi la tecnica più frequentemente adottata è quella di campionare sistematicamente, o casualmente, lo spazio delle soluzioni ammissibili, di generare cioè un insieme di soluzioni ammissibili distinte e possibilmente ben ripartite nello spazio delle soluzioni X, e di portare, mediante un algoritmo di ricerca iterativo, ciascun punto campionato al suo minimo locale. Il miglior minimo locale viene scelto come approssimazione della soluzione ottima del problema. La disciplina che si occupa, nell'ambito della PNL, di determinare i punti di minimo globale prende il nome di *Ottimizzazione Globale* (OG). Le tecniche dell'OG si basano comunque sulla disponibilità di algoritmi per l'individuazione di punti di minimo locale ed è perciò su queste tecniche che noi concentriamo la nostra attenzione.

Osserviamo infine che se la funzione $f(\boldsymbol{x})$ e l'insieme che definisce la regione ammissiile X, godono di alcune proprietà molto generali è possibile affermare l'esistenza di almeno un punto di minimo e giustificarne quindi la ricerca.

> **Proprietà 4** *Se una funzione $f(\boldsymbol{x})$ è continua su un insieme $X \subseteq \mathbb{R}^n$ chiuso e limitato, allora $f(\boldsymbol{x})$ ammette un punto di minimo globale in X.*
>
> **Proprietà 5** *Se per una funzione $f(\boldsymbol{x})$ definita su tutto $\mathbb{R}^n$ vale*
>
> $$\lim_{\|\boldsymbol{x}\| \to \infty} f(\boldsymbol{x}) = +\infty$$
>
> *allora $f(\boldsymbol{x})$ ammette minimo globale.*

3 Ottimizzazione non vincolata

Nei problemi di ottimizzazione non vincolata si minimizza la funzione obiettivo senza porre restrizioni al valore che le variabili possono assumere.
Di solito non vi è una prospettiva globale dell'andamento di tale funzione. Tutto quello che sappiamo è il valore di $f(\boldsymbol{x})$ e (forse) di qualche sua derivata in un insieme di punti $\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$. Fortunatamente, questo è spesso tutto quello che serve ai nostri algoritmi per identificare una soluzione affidabile senza impiegare troppo tempo o troppa memoria.

Esempio Supponiamo di cercare una curva che si adatti ad un insieme di dati sperimentali, corrispondenti alle misure $y_1, y_2, \ldots, y_m$ di un segnale campionato agli istanti $t_1, t_2, \ldots, t_m$. Dai dati in nostro possesso e da conoscenze relative all'applicazione, deduciamo che il segnale ha un andamento esponenziale ed oscillatorio, e scegliamo di modellarlo con la funzione

$$\phi(t, \boldsymbol{x}) = x_1 + x_2 e^{-(x_3 - t)^2/x_4} + x_5 \cos(x_6 t)$$

I numeri reali $x_1, x_2, \ldots, x_6$ sono i parametri del modello. Desideriamo sceglierli in modo da far aderire il più possibile i valori $\phi(t, \boldsymbol{x})$, calcolati dal modello, ai dati osservati $y_1, y_2, \ldots, y_m$. Per trasformare il nostro obiettivo in un problema di ottimizzazione, definiamo i residui

$$r_j = y_j - \phi(t_j, \boldsymbol{x}), \quad j = 1, \ldots, m,$$

che misurano la discrepanza fra i dati predetti dal modello e quelli misurati. La nostra stima del vettore $\boldsymbol{x}$ verrà ottenuta risolvendo il problema

$$\min_{x \in \mathbb{R}^6} f(\boldsymbol{x}) = r_1^2(\boldsymbol{x}) + r_2^2(\boldsymbol{x}) + \ldots + r_m^2(\boldsymbol{x}).$$

Questo è un *problema ai minimi quadrati non lineare*, un caso speciale di ottimizzazione non vincolata (cfr. Sezione 8).
Si osservi che il calcolo della sola funzione obiettivo, per un dato valore delle variabili $\boldsymbol{x}$, può essere computazionalmente oneroso, anche se le variabili sono solo 6, dipendendo dal numero m di osservazioni, che può essere anche molto elevato.

Supponiamo ora che il valore $f(\boldsymbol{x}^*)$ della soluzione ottima $\boldsymbol{x}^*$ sia diverso da zero, poiché, come spesso accade, il modello non ha riprodotto esattamente il valore di tutti i punti campionati.

Come possiamo verificare che $\boldsymbol{x}^*$ è effettivamente un minimo per la funzione $f(\boldsymbol{x})$?

Per rispondere dobbiamo definire quali condizioni analitiche sono soddisfatte dai punti di minimo e spiegare come fare a riconoscerle.

4 Condizioni analitiche di ottimalità

La prima questione alla quale occorre dare risposta è la seguente:

Come è possibile riconoscere un minimo locale senza esplorare tutti gli infiniti punti che appartengono al suo intorno?

È possibile ottenere questo risultato quando la funzione non è *spigolosa*, ad esempio quando è due volte differenziabile con continuità o più formalmente se è di classe $\boldsymbol{C}^2$.

> **Definizione 6** *Una funzione $f : \mathbb{R}^n \to \mathbb{R}$ è detta di classe $\boldsymbol{C}^k(\mathrm{X})$ se e solo se essa è differenziabile con continuità almeno k volte in tutti i punti $\boldsymbol{x} \in \mathrm{X} \subseteq \mathbb{R}^n$.*

Possiamo dire che un punto $\boldsymbol{x}^*$ è un minimo locale, magari in senso stretto, esaminando il vettore gradiente e la matrice hessiana, calcolati in $\boldsymbol{x}^*$, $\nabla f(\boldsymbol{x}^*)$ e $H(\boldsymbol{x}^*)$. Lo strumento che ci permette di effettuare questa analisi è lo sviluppo in serie di Taylor di $f(\boldsymbol{x})$, arrestato al più al secondo ordine.

Proprietà 6 *Sia $f : \mathbb{R}^n \to \mathbb{R}$ una funzione differenziabile con continuità. Dati un punto $\boldsymbol{x} \in \mathbb{R}^n$ ed un vettore $\boldsymbol{h} \in \mathbb{R}^n$, vale la seguente relazione*

$$f(\boldsymbol{x} + \boldsymbol{h}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T \boldsymbol{h} + \beta(\boldsymbol{x}, \boldsymbol{h})$$

dove $\beta(\boldsymbol{x}, \boldsymbol{h})$ è un infinitesimo che tende a zero più velocemente della norma di $\boldsymbol{h}$, o, in altri termini, vale

$$f(\boldsymbol{x} + \boldsymbol{h}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x} + t\boldsymbol{h})^T \boldsymbol{h}$$

per qualche $t \in (0, 1)$. Inoltre, se $f(\boldsymbol{x})$ è una funzione di classe $\boldsymbol{C}^2(\mathbb{R}^n)$ si ha

$$f(\boldsymbol{x} + \boldsymbol{h}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T \boldsymbol{h} + \frac{1}{2} \boldsymbol{h}^T H(\boldsymbol{x}) \boldsymbol{h} + \beta_2(\boldsymbol{x}, \boldsymbol{h})$$

dove $\beta_2(\boldsymbol{x}, \boldsymbol{h})$ è un infinitesimo che tende a zero più velocemente della norma al quadrato di $\boldsymbol{h}$, o, in altri termini, valgono

$$f(\boldsymbol{x} + \boldsymbol{h}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T \boldsymbol{h} + \frac{1}{2} \boldsymbol{h}^T H(\boldsymbol{x} + t\boldsymbol{h}) \boldsymbol{h}$$

e

$$\nabla f(\boldsymbol{x} + \boldsymbol{h}) = \nabla f(\boldsymbol{x}) + \int_0^1 H(\boldsymbol{x} + t\boldsymbol{h}) \boldsymbol{h} dt$$

per qualche $t \in (0, 1)$.

Mediante lo sviluppo in serie di Taylor si possono ricavare *condizioni necessarie* di ottimalità: si assume che $\boldsymbol{x}^*$ sia un minimo locale e si studiano proprietà di $\nabla f(\boldsymbol{x}^*)$ e di $H(\boldsymbol{x}^*)$. Tali proprietà risultano più semplici da ricavare se prima introduciamo i concetti di *direzione di discesa* e di *derivata direzionale*.

Definizione 7 *Data una funzione $f : \mathbb{R}^n \to \mathbb{R}$, un vettore $\boldsymbol{d} \in \mathbb{R}^n$ si dice direzione di discesa per f in $\boldsymbol{x}$ se esiste $\overline{\lambda} > 0$ tale che:*

$$f(\boldsymbol{x} + \lambda \boldsymbol{d}) < f(\boldsymbol{x})$$

per ogni $0 < \lambda < \overline{\lambda}$.

In pratica, se siamo in un punto $\boldsymbol{x}$ e ci spostiamo da $\boldsymbol{x}$ lungo $\boldsymbol{d}$, la funzione, almeno per un tratto, decresce. Quindi, le direzioni di discesa sono quelle da seguire per migliorare il valore della funzione obiettivo. Si intuisce quindi che se in un punto $\boldsymbol{x}$ esistono direzioni di discesa, quel punto non può essere un minimo locale. Per sapere se un punto è un minimo locale è necessario scoprire che non esistono direzioni di discesa in quel punto.

Ma come si possono individuare le direzioni di discesa?

Definizione 8 *Sia data una funzione $f : \mathbb{R}^n \to \mathbb{R}$, un vettore $\boldsymbol{d} \in \mathbb{R}^n$ e un punto $\boldsymbol{x}$ dove f è definita. Se esiste il limite:*

$$\lim_{\lambda \to 0^+} \frac{f(\boldsymbol{x} + \lambda \boldsymbol{d}) - f(\boldsymbol{x})}{\lambda}$$

allora tale limite prende il nome di derivata direzionale della funzione f nel punto $\boldsymbol{x}$ lungo la direzione $\boldsymbol{d}$.

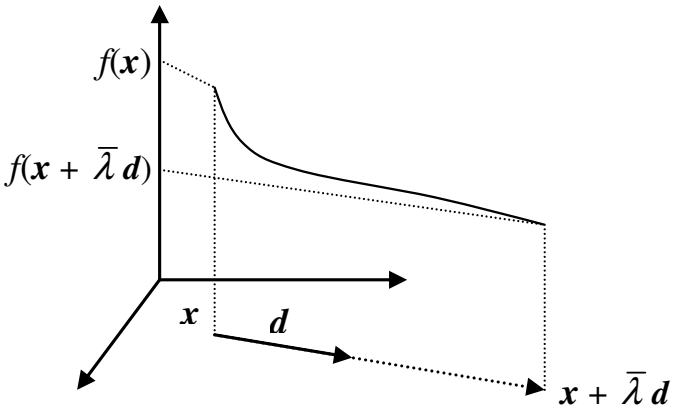


Figura 3: Direzione di discesa

Il significato geometrico della derivata direzionale è il seguente. Dato un punto $\boldsymbol{x} \in X$ e una direzione $\boldsymbol{d} \in \mathbb{R}^n$ si considerano i punti $\boldsymbol{x} + \lambda \boldsymbol{d}$, con $\lambda > 0$, cioè i punti che si incontrano partendo da $\boldsymbol{x}$ e muovendosi lungo la direzione $\boldsymbol{d}$. Il valore della derivata direzionale rappresenta il tasso di variazione del valore della funzione obiettivo, calcolato in $\boldsymbol{x}$, lungo la direzione $\boldsymbol{d}$. In altre parole, come varia il valore di $f(\boldsymbol{x})$ se ci si allontana da $\boldsymbol{x}$ lungo la direzione $\boldsymbol{d}$. Se $\boldsymbol{d} = \boldsymbol{e}_i$, cioè la direzione coincide con quella dell'i-esimo versore degli assi coordinati, allora la derivata direzionale coincide con la derivata parziale della funzione obiettivo nella componente i-esima: $\frac{\partial f}{\partial x_i}$. La derivata direzionale è utile in quanto può essere calcolata anche in punti nei quali una funzione non è differenziabile. Nel caso in cui invece sia possibile derivare la funzione $f(\boldsymbol{x})$, si ricava una interessante relazione fra derivata direzionale e vettore gradiente $\nabla f(\boldsymbol{x})$.

Proprietà 7 *Data una funzione $f : \mathbb{R}^n \to \mathbb{R}$, derivabile con continuità in $\boldsymbol{x} \in \mathbb{R}^n$ la derivata direzionale di f nel punto $\boldsymbol{x}$ lungo la direzione $\boldsymbol{d}$ è data da $\nabla f(\boldsymbol{x})^T \boldsymbol{d}$.*

Dimostrazione Applicando lo sviluppo in serie di Taylor arrestato al primo ordine ad f in $\boldsymbol{x}$ ricaviamo: $f(\boldsymbol{x} + \lambda \boldsymbol{d}) = f(\boldsymbol{x}) + \lambda \nabla f(\boldsymbol{x})^T \boldsymbol{d} + \beta(\boldsymbol{x}, \lambda \boldsymbol{d})$. Dividendo per λ e riarrangiando i termini si ottiene: $\frac{f(\boldsymbol{x}+\lambda \boldsymbol{d}) - f(\boldsymbol{x})}{\lambda} = \nabla f(\boldsymbol{x})^T \boldsymbol{d} + \frac{\beta(\boldsymbol{x}, \lambda \boldsymbol{d})}{\lambda}$. Il limite per $\lambda \to 0$ del termine a sinistra del segno di uguale porta al risultato. □

Il legame fra derivata direzionale e gradiente è proprio quello che ci serve per legare le informazioni che ci dà il vettore gradiente calcolato in un punto $\boldsymbol{x}$ con la presenza di direzioni di discesa in $\boldsymbol{x}$.

Proprietà 8 *Sia data una funzione $f : \mathbb{R}^n \to \mathbb{R}$, derivabile con continuità in $\mathbb{R}^n$. Dati un punto $\boldsymbol{x}$ ed una direzione $\boldsymbol{d} \in \mathbb{R}^n$, se la derivata direzionale di f in $\boldsymbol{x}$ lungo $\boldsymbol{d}$ è negativa, allora la direzione $\boldsymbol{d}$ in $\boldsymbol{x}$ è di discesa.*

Dimostrazione Sappiamo che $\lim_{\lambda \to 0^+} \frac{f(\boldsymbol{x}+\lambda \boldsymbol{d}) - f(\boldsymbol{x})}{\lambda} = \nabla f(\boldsymbol{x})^T \boldsymbol{d}$ e che per ipotesi $\nabla f(\boldsymbol{x})^T \boldsymbol{d} < 0$. Quindi, per λ sufficientemente piccolo, vale $f(\boldsymbol{x} + \lambda \boldsymbol{d}) - f(\boldsymbol{x}) < 0$, che è il significato di direzione di discesa. □

Quindi, qualunque direzione formi un angolo ottuso ($> 90°$) con il gradiente della funzione obiettivo calcolato in $\boldsymbol{x}$ è una direzione di discesa, e tutti i punti sufficientemente vicini a $\boldsymbol{x}$

lungo tale direzione sono caratterizzati dall'avere un valore della funzione f inferiore a quello che ha f in $\boldsymbol{x}$. Inoltre, l'antigradiente $-\nabla f(\boldsymbol{x})$ è esso stesso una direzione di discesa (mentre $\nabla f(\boldsymbol{x})$ è una direzione di salita).

Questo significa che se in un punto $\boldsymbol{x}$ il vettore $\nabla f(\boldsymbol{x})$ non è nullo, allora esistono direzioni di discesa in $\boldsymbol{x}$ e tale punto non può essere un minimo locale.

Teorema 1 *Data una funzione $f : \mathbb{R}^n \to \mathbb{R}$, derivabile con continuità in $\boldsymbol{x}^* \in \mathbb{R}^n$, condizione necessaria affinché il punto $\boldsymbol{x}^*$ sia un minimo locale per f è che il gradiente della funzione calcolato in $\boldsymbol{x}^*$ sia nullo: $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$.*

Dimostrazione Se fosse $\nabla f(\boldsymbol{x}^*) \neq \mathbf{0}$ allora $-\nabla f(\boldsymbol{x}^*)$ sarebbe una direzione di discesa ed esisterebbe un punto vicino a $\boldsymbol{x}^*$, $\boldsymbol{x}^* - \lambda \nabla f(\boldsymbol{x}^*)$, in cui il valore della funzione obiettivo sarebbe $f(\boldsymbol{x}^* - \lambda \nabla f(\boldsymbol{x}^*)) < f(\boldsymbol{x}^*)$, contraddicendo il fatto che $\boldsymbol{x}^*$ sia un minimo locale. □

Il Teorema 1 fornisce delle condizioni molto generali, dette condizioni necessarie di ottimalità del 1° ordine. Un punto che soddisfa tali condizioni si dice *punto stazionario*, e PS indica l'insieme dei punti stazionari per f, ossia: $PS = \{\boldsymbol{x} : \boldsymbol{x} \in \mathbb{R}^n, \nabla f(\boldsymbol{x}) = \mathbf{0}\}$

Naturalmente, i punti stazionari possono essere sia punti di minimo locale, sia punti di ...massimo locale! E, purtroppo, anche punti che non sono né punti di massimo, né punti di minimo. Per dirimere questa nuova questione (punti di minimo, punti di massimo, altri punti) abbiamo bisogno di usare le informazioni che ci può dare la matrice hessiana.

Teorema 2 *Data una funzione $f : \mathbb{R}^n \to \mathbb{R}$ di classe $\boldsymbol{C}^2(\boldsymbol{x}^*)$, condizioni necessarie affinché il punto $\boldsymbol{x}^*$ sia un minimo locale per f sono che il gradiente della funzione calcolato in $\boldsymbol{x}^*$, $\nabla f(\boldsymbol{x}^*)$, sia nullo e che valga la relazione: $\boldsymbol{d}^T H(\boldsymbol{x}^*)\boldsymbol{d} \geq 0$ per ogni $\boldsymbol{d} \in \mathbb{R}^n$.*

Dimostrazione Per lo sviluppo in serie di Taylor arrestato al secondo termine:

$$f(\boldsymbol{x}^* + \lambda \boldsymbol{d}) = f(\boldsymbol{x}^*) + \lambda \nabla f(\boldsymbol{x}^*)^T \boldsymbol{d} + \frac{1}{2}\lambda^2 \boldsymbol{d}^T H(\boldsymbol{x}^*)\boldsymbol{d} + \beta_2(\boldsymbol{x}^*, \lambda \boldsymbol{d})$$

e poiché $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$ si può ricavare

$$\frac{f(\boldsymbol{x}^* + \lambda \boldsymbol{d}) - f(\boldsymbol{x}^*)}{\lambda^2} = \frac{1}{2}\boldsymbol{d}^T H(\boldsymbol{x}^*)\boldsymbol{d} + \frac{\beta_2(\boldsymbol{x}^*, \lambda \boldsymbol{d})}{\lambda^2}$$

Poiché per ipotesi $\boldsymbol{x}^*$ è minimo locale allora, per λ sufficientemente piccolo, il termine a sinistra è non negativo. Risulta quindi $\frac{1}{2}\boldsymbol{d}^T H(\boldsymbol{x}^*)\boldsymbol{d} + \frac{\beta_2(\boldsymbol{x}^*, \lambda \boldsymbol{d})}{\lambda^2} \geq 0$. Passando al limite per $\lambda \to 0$, e osservando che $\boldsymbol{d}$ è una direzione qualsiasi, segue la tesi. □

Le condizioni espresse nel Teorema 2 sono dette condizioni necessarie di ottimalità del 2° ordine. Le condizioni appena espresse non sono però sufficienti. Se il punto $\boldsymbol{x}^* \in PS$ possiamo calcolare la matrice hessiana $H(\boldsymbol{x}^*)$ e se la matrice soddisfa la relazione $\boldsymbol{d}^T H(\boldsymbol{x}^*)\boldsymbol{d} \geq 0$ per ogni $\boldsymbol{d} \in \mathbb{R}^n$ possiamo solo escludere che il punto sia un massimo locale. Tuttavia, se la funzione obiettivo è due volte differenziabile con continuità è possibile enunciare anche (finalmente) delle condizioni sufficienti di ottimalità del 2° ordine. Cioè condizioni che, se soddisfatte da un punto, garantiscono che quel punto sia di minimo.

Teorema 3 *Data una funzione $f : \mathbb{R}^n \to \mathbb{R}$ di classe $\boldsymbol{C}^2(\boldsymbol{x}^*)$, condizioni sufficienti affinché il punto $\boldsymbol{x}^*$ sia un minimo locale in senso stretto per f sono che il gradiente della funzione calcolato in $\boldsymbol{x}^*$, $\nabla f(\boldsymbol{x}^*)$, sia nullo e che valga la relazione: $\boldsymbol{d}^T H(\boldsymbol{x}^*)\boldsymbol{d} > 0$ per ogni $\boldsymbol{d} \in \mathbb{R}^n$.*

Dimostrazione Per lo sviluppo in serie di Taylor arrestato al secondo termine:

$$f(\boldsymbol{x}^* + \lambda \boldsymbol{d}) = f(\boldsymbol{x}^*) + \lambda \nabla f(\boldsymbol{x}^*)^T \boldsymbol{d} + \frac{1}{2}\lambda^2 \boldsymbol{d}^T H(\boldsymbol{x}^*)\boldsymbol{d} + \beta_2(\boldsymbol{x}^*, \lambda \boldsymbol{d})$$

e poiché $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$ si può ricavare

$$\frac{f(\boldsymbol{x}^* + \lambda \boldsymbol{d}) - f(\boldsymbol{x}^*)}{\lambda^2} = \frac{1}{2}\boldsymbol{d}^T H(\boldsymbol{x}^*)\boldsymbol{d} + \frac{\beta_2(\boldsymbol{x}^*, \lambda \boldsymbol{d})}{\lambda^2}$$

Poiché per λ sufficientemente piccolo, il termine a destra è dominato dalla componente $\frac{1}{2}\boldsymbol{d}^T H(\boldsymbol{x}^*)\boldsymbol{d}$, risulta $f(\boldsymbol{x}^* + \lambda \boldsymbol{d}) - f(\boldsymbol{x}^*) > 0$. Osservando che $\boldsymbol{d}$ è una direzione qualsiasi, segue la tesi. □

Si noti che le condizioni sufficienti del secondo ordine appena enunciate garantiscono qualcosa in più delle condizioni necessarie discusse precedentemente, e cioè che il punto in questione è un minimo locale *in senso stretto*. Inoltre, le condizioni appena enunciate sono sufficienti ma non necessarie, cioè un punto può essere un minimo locale in senso stretto e non soddisfare le condizioni sufficienti del secondo ordine. Ad esempio, il punto $x^* = 0$ è un minimo in senso stretto della funzione $f(x) = x^4$, ma in tale punto l'hessiano si annulla e non è quindi definito positivo.

Apparentemente però abbiamo solo spostato il problema. Infatti, non sembra arduo valutare se il vettore gradiente ha o meno tutte le componenti nulle, ma sembra invece computazionalmente complicato stabilire se una matrice quadrata H soddisfi o meno la relazione $\boldsymbol{d}^T H \boldsymbol{d} > 0$ *per ogni* $\boldsymbol{d} \in \mathbb{R}^n$.

Definizione 9 *Una matrice H quadrata di ordine n, si dice (semi)-definita positiva su un insieme $X \subseteq \mathbb{R}^n$ se per ogni $\boldsymbol{d} \in X, \boldsymbol{d} \neq \boldsymbol{0}$, vale*

$$\boldsymbol{d}^T H \boldsymbol{d} > 0 \text{ definita positiva}$$
$$\boldsymbol{d}^T H \boldsymbol{d} \geq 0 \text{ semi-definita positiva}$$

Una matrice H è (semi)-definita negativa se $-H$ è (semi)-definita positiva. Se H non è né (semi)-definita positiva, né (semi)-definita negativa, allora la matrice è indefinita.

Ora, se una matrice è *simmetrica*, come è il caso di ogni matrice hessiana, è possibile stabilire se essa è (semi)-definita positiva osservando il segno dei determinanti delle n sottomatrici quadrate che si ottengono considerando le matrici formate dalle sue prime i righe ed i colonne.

Proprietà 9 *Una matrice H simmetrica è (semi)-definita positiva se e solo se i determinanti di tutti i minori principali della matrice sono $(\geq) > 0$*

Un'altra caratterizzazione della (semi)-definita positività di una matrice H è data dal segno dei suoi autovalori.

Proprietà 10 *Una matrice H simmetrica è (semi)-definita positiva se e solo se ha tutti gli autovalori* $(\geq) > 0$.

Inoltre gli autovalori di H simmetrica possono fornire un limite superiore o inferiore al valore di $\boldsymbol{d}^T H \boldsymbol{d}$ per ogni valore di $\boldsymbol{d}$: $\lambda_{min} \|d\|^2 \leq \boldsymbol{d}^T H \boldsymbol{d} \leq \lambda_{max} \|d\|^2$, dove λ_{min} := autovalore minimo e λ_{max} := autovalore massimo di H.

In pratica quindi, in un punto di stazionarietà $\boldsymbol{x}^*$, si deve poter analizzare efficientemente il segno degli autovalori della matrice hessiana $H(\boldsymbol{x}^*)$, o di matrici che la approssimano, spesso sfruttando una loro rappresentazione come prodotti di matrici triangolari.
Osserviamo infine che, se in un punto $\boldsymbol{x}^*$ si annulla il gradiente, ma l'Hessiana è indefinita allora $\boldsymbol{x}^*$ non è né un punto di minimo né un punto di massimo.

4.1 Condizioni analitiche di ottimalità per problemi convessi

Come abbiamo già detto i problemi convessi rivestono un ruolo particolare in ottimizzazione. La ragione principale è che nel caso di problemi convessi saper risolvere il problema di ottimizzazione locale (determinare cioè un qualsiasi minimo locale) implica saper risolvere il problema di ottimizzazione globale . Infatti, nel caso dei problemi convessi, i minimi locali coincidono con i minimi globali.

Proprietà 11 *In un problema di ottimizzazione convesso (cioè con funzione obiettivo e regione ammissibile entrambe convesse) ogni minimo locale è anche minimo globale.*

Quindi, nel caso di problemi convessi l'identificazione di un minimo locale è sufficiente per la loro soluzione. Di conseguenza è preliminare ad ogni approccio risolutivo stabilire se il problema che si intende risolvere è convesso o meno. Di seguito richiamiamo utili proprietà che permettono di stabilire, dall'analisi del vettore gradiente o della matrice hessiana calcolati in un punto, quando una funzione è convessa nell'intorno di quel punto. Di seguito poi, presenteremo proprietà che risulteranno utili nell'identificare i minimi locali (e quindi globali) di funzioni convesse.

Proprietà 12 *Una funzione $f : \mathrm{X} \to \mathbb{R}$, derivabile con continuità nell'insieme convesso* X*, è convessa su* X *se e solo se, comunque presi due punti $\boldsymbol{x}$ e $\boldsymbol{y} \in \mathrm{X}$, vale la relazione:*

$$f(\boldsymbol{y}) - f(\boldsymbol{x}) \geq \nabla f(\boldsymbol{x})^T (\boldsymbol{y} - \boldsymbol{x})$$

Dimostrazione Se f è convessa allora vale la relazione $\lambda f(\boldsymbol{y}) + (1-\lambda) f(\boldsymbol{x}) \geq f(\lambda \boldsymbol{y} + (1-\lambda)\boldsymbol{x})$ che possiamo riscrivere come

$$f(\boldsymbol{y}) - f(\boldsymbol{x}) \geq \frac{f(\boldsymbol{x} + \lambda(\boldsymbol{y} - \boldsymbol{x})) - f(\boldsymbol{x})}{\lambda}$$

e passando la limite per $\lambda \to 0$ si ha la tesi.
Viceversa, se vale $f(\boldsymbol{y}) - f(\boldsymbol{x}) \geq \nabla f(\boldsymbol{x})^T (\boldsymbol{y} - \boldsymbol{x})$, allora per $\boldsymbol{z} = \lambda \boldsymbol{y} + (1-\lambda)\boldsymbol{x} \in X$, con $\lambda \in [0,1]$ valgono $f(\boldsymbol{y}) - f(\boldsymbol{z}) \geq \nabla f(\boldsymbol{z})^T (\boldsymbol{y} - \boldsymbol{z})$ e $f(\boldsymbol{x}) - f(\boldsymbol{z}) \geq \nabla f(\boldsymbol{z})^T (\boldsymbol{x} - \boldsymbol{z})$. Moltiplicando la prima espressione per λ, la seconda per $(1-\lambda)$ e sommando si ottiene

$$\lambda f(\boldsymbol{y}) + (1-\lambda) f(\boldsymbol{x}) - f(\boldsymbol{z}) \geq \nabla f(\boldsymbol{z})^T (\lambda(\boldsymbol{y} - \boldsymbol{z}) + (1-\lambda)(\boldsymbol{x} - \boldsymbol{z})) = 0,$$

poiché $\lambda(\boldsymbol{y}-\boldsymbol{z})+(1-\lambda)(\boldsymbol{x}-\boldsymbol{z})=\boldsymbol{0}$, e quindi

$$\lambda f(\boldsymbol{y})+(1-\lambda)f(\boldsymbol{x})\geq f(\boldsymbol{z})=f(\lambda\boldsymbol{y}+(1-\lambda)\boldsymbol{x})$$

□

Proprietà 13 *Sia* X *un insieme convesso e sia* $f:\mathrm{X}\to\mathbb{R}$ *di classe* $\boldsymbol{C}^2(\mathrm{X})$*, allora* f *è convessa se e solo se la matrice hessiana* $H(\boldsymbol{x})$ *è semidefinita positiva in* $\boldsymbol{x}$*, per ogni punto* $\boldsymbol{x}\in\mathrm{X}$*.*

In pratica quindi, anche per stabilire la convessità della funzione obiettivo si deve poter analizzare efficientemente il segno degli autovalori della matrice hessiana H.
Inoltre, quando analizzeremo i problemi di ottimizzazione vincolata, sfruttando le proprietà 1, 2 e 3, saremo interessati allo studio delle matrici hessiane delle funzioni $g_i(\boldsymbol{x})$ allo scopo di stabilire se la regione ammissibile sia convessa o meno.

Proprietà 14 *Data una funzione* $f:\mathbb{R}^n\to\mathbb{R}$ *di classe* $\boldsymbol{C}^2(\mathbb{R}^n)$*, sia* $\boldsymbol{x}$ *un punto in cui la matrice hessiana* $H(\boldsymbol{x})$ *è definita positiva. Allora, almeno in un intorno di* $\boldsymbol{x}$*,* f *è strettamente convessa.*

Dimostrazione Dati due punti $\boldsymbol{x}$ e $\boldsymbol{y}$, approssimiamo f con Taylor, nell'intorno di $\boldsymbol{x}$:

$$f(\boldsymbol{y})=f(\boldsymbol{x})+\nabla f(\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{x})+\frac{1}{2}(\boldsymbol{y}-\boldsymbol{x})^TH(\boldsymbol{x})(\boldsymbol{y}-\boldsymbol{x})+\beta_2(\boldsymbol{x},\boldsymbol{y})$$

Poiché $H(\boldsymbol{x})$ è definita positiva vale $(\boldsymbol{y}-\boldsymbol{x})^TH(\boldsymbol{x})(\boldsymbol{y}-\boldsymbol{x})>0$ e, almeno in un intorno di $\boldsymbol{x}$, questa espressione domina su $\beta_2(\boldsymbol{x},\boldsymbol{y})$. Quindi, se al secondo membro sottraiamo questa quantità positiva, otteniamo $f(\boldsymbol{y})-f(\boldsymbol{x})>\nabla f(\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{x})$. □

Come detto le funzioni convesse rivestono un ruolo importante nella PNL poiché sono le uniche per le quali il problema di ottimizzazione globale è, almeno in linea di principio risolubile. In particolare, le condizioni analitiche si semplificano sensibilmente.

Teorema 4 *Data una funzione convessa* $f:\mathbb{R}^n\to\mathbb{R}$ *derivabile con continuità, condizione necessaria e sufficiente affinché il punto* $\boldsymbol{x}^*$ *sia un minimo locale per* f *è che* $\nabla f(\boldsymbol{x}^*)=\boldsymbol{0}$*.*

Dimostrazione Per la convessità di f per ogni coppia di punti $\boldsymbol{x}$ e $\boldsymbol{y}\in\mathbb{R}^n$ vale $f(\boldsymbol{y})-f(\boldsymbol{x})\geq\nabla f(\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{x})$ e poiché nel punto di minimo locale $\boldsymbol{x}^*$ il gradiente si annulla si ricava $f(\boldsymbol{y})\geq f(\boldsymbol{x})$. □

Nella funzione convessa ogni punto di minimo locale è anche di minimo globale (cfr. proprietà 11) quindi se abbiamo due punti distinti a gradiente nullo, allora abbiamo infiniti punti di stazionarietà (a gradiente nullo).

Teorema 5 *Data una funzione convessa* $f:\mathbb{R}^n\to\mathbb{R}$*, l'insieme dei punti di minimo della funzione è convesso.*

Dimostrazione Immediata conseguenza del punto .d della proprietà 2. □

Purtroppo la maggior parte dei problemi di ottimizzazione che si incontrano in pratica non ammettono modellizzazioni convesse. Come conseguenza l'individuazione di un punto di minimo globale può risultare un'operazione molto costosa, in termini computazionali, se non addirittura impossibile. In queste situazioni il meglio che si può fare è individuare un insieme di minimi locali e da essi dedurne stime, più o meno buone, del valore della soluzione ottima.

4.2 Condizioni *empiriche* di ottimalità

Dal punto di vista pratico un algoritmo di tipo iterativo viene fatto terminare se soddisfa un qualche criterio di convergenza di tipo empirico. Si tratta in pratica di trovare delle relazioni di tipo quantitativo che ci permettano di riconoscere il punto corrente, nella successione $\{\boldsymbol{x}_k\}$ di punti visitati dall'algoritmo, come un punto stazionario. In genere la terminazione viene imposta qualora all'iterazione k uno, o una combinazione dei seguenti criteri viene soddisfatta

$$\begin{aligned} ||\boldsymbol{x}_k - \boldsymbol{x}_{k-1}|| &< \varepsilon_1 \\ ||\nabla f(\boldsymbol{x}_k)|| &< \varepsilon_2 \\ ||f(\boldsymbol{x}_k) - f(\boldsymbol{x}_{k-1})|| &< \varepsilon_3 \end{aligned}$$

dove $\varepsilon_1, \varepsilon_2$ e ε_3 sono prescritti valori positivi di tolleranza.

4.3 Programmazione Quadratica

Analizziamo ora una classe di problemi che riveste una importanza particolare nell'ambito delle tecniche di ottimizzazione. Si tratta dei problemi di tipo *quadratico*:

$$\begin{array}{ll} \min & f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T Q \boldsymbol{x} - \boldsymbol{b}^T \boldsymbol{x} \\ t.c. & \boldsymbol{x} \in \mathbb{R}^n \end{array}$$

dove Q è una matrice quadrata simmetrica di ordine n.
Nel caso di funzioni obiettivo quadratiche[1] si ricava immediatamente $\nabla f(\boldsymbol{x}) = Q\boldsymbol{x} - \boldsymbol{b}$ e $H(\boldsymbol{x}) = Q$. In particolare la matrice hessiana è costante e il vettore gradiente è lineare.

Nell'ambito della programmazione quadratica la condizione necessaria di ottimalità del primo ordine diventa: $\nabla f(\boldsymbol{x}) = \boldsymbol{0}$, che si riduce al sistema lineare in n equazioni ed n incognite $Q\boldsymbol{x} = \boldsymbol{b}$. In pratica, quando la matrice Q è invertibile, possiamo, *imponendo* il soddisfacimento della condizione necessaria di ottimalità del primo ordine, ricavare l'unico punto stazionario di $f(\boldsymbol{x})$:

$$\boldsymbol{x}^* = Q^{-1}\boldsymbol{b}$$

Ora, per distinguere se tale punto sia di minimo o di massimo è sufficiente analizzare la matrice hessiana $H = Q$. Come conseguenza della Proprietà 14, se Q è definita positiva allora la funzione quadratica f è convessa ovunque e, in conseguenza del teorema 4, il punto $\boldsymbol{x}^*$ è un minimo globale.

Analizziamo ora i casi che si possono presentare.

4.3.1 Q non è semidefinita positiva

In questo caso non è vero che $\forall\ \boldsymbol{x}$ vale $\boldsymbol{x}^T Q \boldsymbol{x} \geq 0$ e quindi non possiamo limitare la nostra attenzione al solo vettore gradiente. La funzione obiettivo non ha punti di minimo.

4.3.2 Q è definita positiva

In questo caso $\boldsymbol{x}^* = Q^{-1}\boldsymbol{b}$ è l'unico ottimo globale del problema.

[1] Si osservi che tali problemi non hanno nulla a che fare con i *problemi ai minimi quadrati* introdotti nella Sezione 8.

4.3.3 Q è semidefinita positiva

In questo caso se Q è invertibile ricadiamo nel caso precedente, altrimenti in uno dei due sottocasi seguenti

- non esistono soluzioni, cioè non esiste minimo
- ci sono infinite soluzioni.

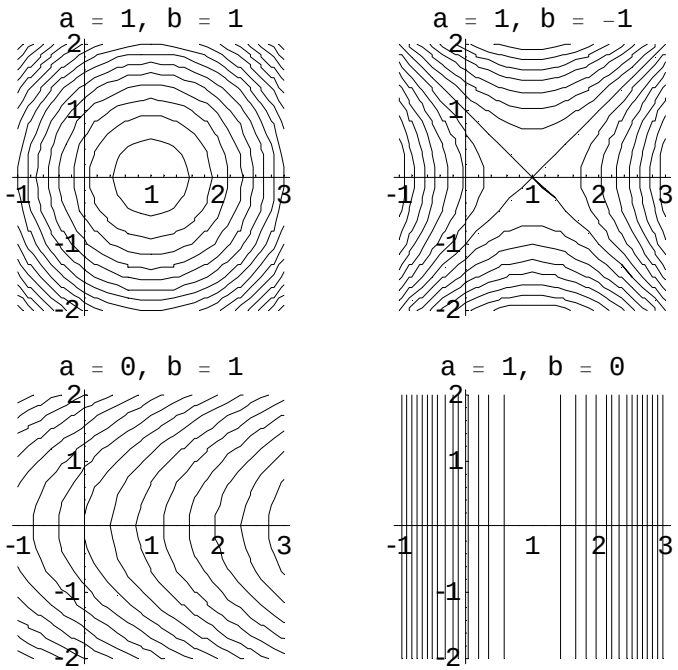


Figura 4: Esempi di funzioni quadratiche

4.3.4 Esempi

Consideriamo la funzione in due variabili $f(x,y) = \frac{1}{2}\left(ax^2 + by^2\right) - x$ che possiamo riscrivere come

$$f(\boldsymbol{x}) = \frac{1}{2}(x,y)\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} - (x,y)\begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Si osservi che in questo esempio, a e b coincidono con gli autovalori di Q, poiché essa è diagonale.

Primo caso: $a > 0, \quad b > 0$.

$\nabla f(\boldsymbol{x}) = \begin{pmatrix} ax - 1 \\ by \end{pmatrix} = \mathbf{0}, \quad \boldsymbol{x}^* = \begin{pmatrix} 1/a \\ 0 \end{pmatrix} = \mathbf{0}$

Q è invertibile e definita positiva. $\boldsymbol{x}^*$ è ottimo globale.

Secondo caso: $a > 0, \quad b < 0$

$\nabla f(\boldsymbol{x}) = \begin{pmatrix} ax - 1 \\ by \end{pmatrix} = \mathbf{0}, \quad \boldsymbol{x}^* = \begin{pmatrix} 1/a \\ 0 \end{pmatrix} = \mathbf{0}$

Q è invertibile ma non semidefinita positiva. $\boldsymbol{x}^*$ è un punto di sella.

Terzo caso: $a = 0, \quad b \neq 0$

$\nabla f(\boldsymbol{x}) = \begin{pmatrix} -1 \\ by \end{pmatrix}$, sempre diverso da $\mathbf{0}$, Q non è invertibile e non esiste soluzione.
Se $b > 0$ ($b < 0$) allora Q è semidefinita positiva (negativa).

Quarto caso: $a \neq 0, \quad b = 0$
$\nabla f(\boldsymbol{x}) = \begin{pmatrix} ax - 1 \\ 0 \end{pmatrix} = \mathbf{0}$. Caso di ∞ soluzioni: i punti in cui il gradiente si annulla sono tutti i punti $(1/a, \gamma)$ con $\gamma \in \mathbb{R}$. Q non è invertibile.
Se $a > 0$ ($a < 0$) allora Q è semidefinita positiva (negativa) e gli infiniti punti stazionari sono tutti punti di minimo (massimo). La Figura 4 illustra i quattro casi.

5 Convergenza e rapidità di convergenza

In tutti i casi, e sono la maggioranza, in cui non si è in grado di calcolare analiticamente l'insieme dei punti stazionari PS, si deve ricorre ad algoritmi di tipo iterativo. Come abbiamo detto all'inizio, tali algoritmi, a partire da un vettore iniziale di variabili $\boldsymbol{x}_0$ generano iterativamente una sequenza di vettori, $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$, quali approssimazioni via a via migliori della soluzione ottima. Prima di descrivere, nella prossima sezione, tali algoritmi in maggior dettaglio, proviamo a rispondere qui alla seguente domanda:

In che modo possiamo classificare la bontà di un algoritmo iterativo nell'ambito della PNL?

Nell'analizzare le prestazioni di un algoritmo di discesa, una prima importante distinzione è stabilire se, e in che modo, esso *converga*.
In particolar modo, quando si parla di *convergenza* nell'ambito della PNL, si intende sempre *convergenza ad un minimo locale*.
È qindi nell'ambito della ricerca di un minimo locale, che faremo distinzione fra *convergenza locale* e *convergenza globale*.

> **Definizione 10** *Algoritmo globalmente convergente. Un algoritmo è globalmente convergente se è convergente ad un punto di PS per qualunque $\boldsymbol{x}_0 \in \mathbb{R}^n$ (termina in un minimo locale da qualunque punto parta).*

> **Definizione 11** *Algoritmo localmente convergente. Un algoritmo è localmente convergente se è convergente solo per $\boldsymbol{x}_0$ tale che $\boldsymbol{x}_0 \in I(\boldsymbol{x}^*, \varepsilon)$, $\boldsymbol{x}^* \in PS$, (termina in un minimo locale solo partendo da un intorno del minimo locale).*

Quindi un algoritmo *globalmente convergente* converge ad un minimo locale da qualunque punto parta, mentre un algoritmo *localmente convergente*, puo' convergere ad un minimo locale, o non convergere affatto, al variare del punto di partenza.

Ma, anche supponendo che le condizioni di convergenza di un algoritmo di discesa siano soddisfatte, e che esso, a partire dal punto $\boldsymbol{x}_0$, converge ad un punto di minimo locale $\boldsymbol{x}^*$, è necessario chiedersi in che modo, cioè con che *velocità*, esso converge. Si tratta cioè di caratterizzare la *rapidità* con cui tale convergenza avviene.
Nell'ambito della PL, PLI e dei problemi di ottimizzazione su grafo, tutti gli algoritmi studiati sono in grado di calcolare la soluzione ottima in un numero finito, eventualmente esponenziale,

di iterazioni. In quell'ambito, è possibile fondare la definizione di *efficienza* di un algoritmo di ottimizzazione sulla base del *numero di iterazioni* necessarie, nel caso peggiore, per giungere alla soluzione ottima.
Nel caso della *programmazione non lineare*, tranne che in casi particolari (ad es. il metodo di Newton per funzioni quadratiche, cfr. la Sezione 6.2.3), gli algoritmi risolutivi producono, addirittura, una successione *infinita* di punti $\boldsymbol{x}_k$. Per misurare il grado di convergenza di un algoritmo è quindi necessario introdurre nuovi criteri.
I metodi più utilizzati per misurare la convergenza adottano come criterio di convergenza il rapporto tra gli scostamenti esistenti, ad un'iterazione ed alla successiva, tra la soluzione corrente $\boldsymbol{x}_k$ ed il punto limite $\boldsymbol{x}^*$ al quale l'algoritmo dovrebbe convergere, cioè:

$$\frac{\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|}{\|\boldsymbol{x}_k - \boldsymbol{x}^*\|}$$

Definizione 12 *Sia $\{\boldsymbol{x}_k\}$ una successione in $\mathbb{R}^n$ che converge a $\boldsymbol{x}^*$. Si dice che la convergenza è Q-lineare se esiste un valore costante $r \in (0,1)$ tale che per ogni $k \geq \overline{k}$ vale*

$$\frac{\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|}{\|\boldsymbol{x}_k - \boldsymbol{x}^*\|} \leq r$$

Questo significa che la distanza dalla soluzione $\boldsymbol{x}^*$ decresce ad ogni iterazione di almeno un fattore costante limitato da 1. Per esempio, la successione $1 + (0.5)^k$ converge Q-linearmente ad 1, con tasso $r = 0.5$. Il prefisso Q significa quoziente, poiché questo tipo di convergenza è definita in termini di quozienti di errori successivi.
Se vale la relazione $r \geq 1$ si dice che l'algoritmo ha convergenza *sublineare* e, aggiungiamo, non è di alcun interesse! Molto più interessanti invece gli algoritmi che presentano una convergenza *superlineare*

Definizione 13 *Sia $\{\boldsymbol{x}_k\}$ una successione in $\mathbb{R}^n$ che converge a $\boldsymbol{x}^*$. Si dice che la convergenza è Q-superlineare se*

$$\lim_{k\to\infty} \frac{\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|}{\|\boldsymbol{x}_k - \boldsymbol{x}^*\|} = 0$$

Si osservi che l'esistenza del limite implica che:

$$\frac{\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|}{\|\boldsymbol{x}_k - \boldsymbol{x}^*\|} \leq r_k \text{ con } r_k \to 0 \text{ quando } k \to \infty.$$

Ma si può far di meglio.

Definizione 14 *Sia $\{\boldsymbol{x}_k\}$ una successione in $\mathbb{R}^n$ che converge a $\boldsymbol{x}^*$. Si dice che la convergenza è Q-quadratica se esiste un valore costante $C > 0$ tale che per ogni $k \geq \overline{k}$ vale*

$$\frac{\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|}{\|\boldsymbol{x}_k - \boldsymbol{x}^*\|^2} \leq C$$

La rapidità di convergenza dipende da r e da C, i cui valori dipendono non solo dall'algoritmo ma anche dalle proprietà del particolare problema. Indipendentemente comunque da tali valori, una successione che converge in modo quadratico convergerà sempre più velocemente di una che converge in modo lineare.

Naturalmente, ogni successione che converge quadraticamente converge anche superlinearmente ed ogni successione che converge superlinearmente converge anche linearmente. È possibile definire anche tassi di convergenza superiori (cubici, o oltre) ma non sono di reale interesse pratico. In generale, diciamo che il Q-ordine di convergenza è $p > 1$ se esiste una costante positiva M tale che per ogni $k \geq \overline{k}$ vale

$$\frac{\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|}{\|\boldsymbol{x}_k - \boldsymbol{x}^*\|^p} \leq M$$

Le seguenti successioni hanno la rapidità di convergenza riportata a fianco.

$\{\boldsymbol{x}_k\} = \frac{1}{k}$ (sublineare); $\{\boldsymbol{x}_k\} = \frac{1}{2^k}$ (lineare);
$\{\boldsymbol{x}_k\} = \frac{1}{k!}$ (superlineare); $\{\boldsymbol{x}_k\} = \frac{1}{2^{2^k}}$ (quadratica);

6 Metodi basati su ricerca lineare

La struttura degli algoritmi iterativi di discesa basati su ricerca lineare per problemi di minimizzazione non vincolata è semplice.

Metodo di discesa;
{
  Scegli $\boldsymbol{x}_0 \in \mathbb{R}^n$; $k := 0$;
  While $\nabla f(\boldsymbol{x}_k) \neq \emptyset$;
  {
    calcola $\boldsymbol{d}_k \in \mathbb{R}^n$; /* direzione di discesa */
    calcola $\alpha_k \in \mathbb{R}$; /* passo lungo $\boldsymbol{d}_k$ */
    $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha \boldsymbol{d}_k$;
    $k := k + 1$;
  }
}

Si genera un punto iniziale $\boldsymbol{x}_0 \in \mathbb{R}^n$, si calcola il valore della funzione $f(\boldsymbol{x}_0)$ e del gradiente $\nabla f(\boldsymbol{x}_0)$. Se vale $\nabla f(\boldsymbol{x}_0) = \mathbf{0}$, allora $\boldsymbol{x}_0 \in PS$. Altrimenti, da $\boldsymbol{x}_0$ ci si sposta in cerca di un punto $\boldsymbol{x}_1$, con un valore di $f(\boldsymbol{x}_1)$ migliore di quello di $f(\boldsymbol{x}_0)$. A questo scopo si sceglie una direzione di discesa $\boldsymbol{d}_0$ e lungo tale direzione ci si muove di un passo opportuno α_0. Trovato $\boldsymbol{x}_1$, se $\nabla f(\boldsymbol{x}_1) = \mathbf{0}$ allora $\boldsymbol{x}_1 \in PS$ e l'algoritmo si arresta, altrimenti si cerca un nuovo punto $\boldsymbol{x}_2$ e così via. L'iterazione generica è quindi semplicemente

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k \tag{1}$$

Gli aspetti critici sono ovviamente la scelta di $\boldsymbol{d}_k$ (direzione di discesa) e di α_k (passo lungo la direzione di discesa).

6.1 Determinazione del passo α_k

Affrontiamo per primo il problema di determinare in modo efficace il valore di α_k nell'ipotesi di conoscere la direzione di discesa $\boldsymbol{d}_k$. La scelta della direzione $\boldsymbol{d}_k$ sarà l'argomento della Sezione 6.2.

Nel seguito, per semplificare la notazione ometteremo il pedice k, relativo all'indice dell'iterazione corrente nella relazione (1).
In primo luogo esprimiamo la variazione della funzione obiettivo f al variare di α:

$$\phi(\alpha) = f(\boldsymbol{x} + \alpha \boldsymbol{d}).$$

Poiché in un determinato punto $\boldsymbol{x}$ e per una determinata direzione $\boldsymbol{d}$ solo il valore α può variare con continuità, ci siamo ricondotti alla soluzione di un problema di PNL in una sola variabile. Dobbiamo cioè risolvere il problema

$$\min \phi(\alpha) = f(\boldsymbol{x} + \alpha \boldsymbol{d}), \quad \alpha > 0.$$

Iniziamo col calcolare la derivata della funzione $\phi(\alpha)$. Per prima cosa indichiamo con $\boldsymbol{y}$ il punto incrementato in funzione di α:

$$\boldsymbol{y} = \boldsymbol{x} + \alpha \boldsymbol{d}.$$

Nell'ipotesi che $f(\boldsymbol{y})$ sia differenziabile con continuità, e che quindi il vettore gradiente $\nabla f(\boldsymbol{y})$ sia una funzione continua, possiamo ricavare la derivata di $\phi(\alpha) = f(\boldsymbol{y}(\alpha))$ come derivata di funzione composta:

$$\phi'(\alpha) = \frac{d\phi}{d\alpha} = \sum_{i=1}^{n} \frac{\partial f(\boldsymbol{y})}{\partial y_i} \frac{dy_i}{d\alpha}.$$

che si può riscrivere (grazie alla definizione di gradiente) come:

$$\phi'(\alpha) = \nabla f(\boldsymbol{y})^T \boldsymbol{d} = \nabla f(\boldsymbol{x} + \alpha \boldsymbol{d})^T \boldsymbol{d}.$$

$\phi'(\alpha)$ rappresenta il tasso di variazione del valore della funzione obiettivo man mano che ci allontaniamo dal punto corrente $\boldsymbol{x}$ lungo la direzione $\boldsymbol{d}$. In particolare, per $\alpha = 0$, $\phi'(0)$ coincide con la derivata direzionale calcolata nel punto $\boldsymbol{x}$, $\nabla f(\boldsymbol{x})^T \boldsymbol{d}$ (cfr. Definizione 8 e Proposizione 7).
Abbiamo ora gli elementi formali che ci permetteranno di individuare il valore migliore per il passo α. Come spesso accade siamo però di fronte ad una scelta: da una lato vorremmo calcolare α in modo da ottenere una sostanziale riduzione del valore di f, ma allo stesso tempo non vorremmo spendere troppo tempo in questo calcolo. L'ideale sarebbe trovare il minimo globale della funzione $\phi(\alpha)$, per valori $\alpha > 0$, ma in generale è troppo costoso identificare tale valore. Anche trovare un minimo locale di $\phi(\alpha)$ con una certa accuratezza richiede generalmente troppe valutazioni della funzione obiettivo f e magari del gradiente $\nabla f(\boldsymbol{x})$. Le strategie più efficaci eseguono una ricerca monodimensionale *non esatta* per identificare una lunghezza del passo che permetta una adeguata riduzione di f col minimo sforzo computazionale.
Gli algoritmi di ricerca monodimensionali generano una sequenza di valori di α, fermandosi quando uno di tali valori soddisfa certe condizioni. Tali algoritmi identificano, in una prima fase, un intervallo entro il quale un adeguato valore del passo esiste, ed eseguono, in una successiva fase, una ricerca di tale valore dentro l'intervallo trovato.

Vediamo ora quali richieste deve soddisfare un buon algoritmo per la determinazione del passo α.
In primo luogo, osserviamo che la semplice diminuzione del valore di $f(\boldsymbol{x})$ ad ogni iterazione dell'algoritmo, non è sufficiente a garantire la convergenza del metodo di discesa ad un punto

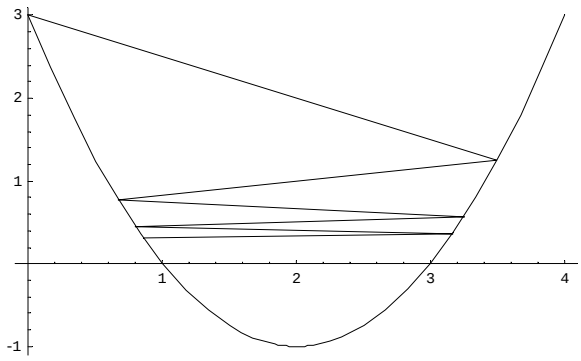


Figura 5: Convergenza a valori errati

stazionario.

Esempio. Consideriamo la funzione di una sola variabile $\phi(\alpha) = \alpha^2 - 4\alpha + 3$, convessa e avente come unico punto di minimo $\alpha^* = 2$, con valore $\phi(2) = -1$, e consideriamo il punto iniziale $\alpha_1 = 0$. Se la successione $\{\alpha_k\}$ di punti visitati dall'algoritmo viene generata dalla formula $\alpha_k = 2 + (-1)^k(1 + 1/k)$, otteniamo che ad ogni iterazione il valore della funzione obiettivo diminuisce: $\phi(\alpha_{k+1}) < \phi(\alpha_k)$, con $k = 1, 2, \ldots$. Come si vede dalla Figura 5 la successione ha due punti limite, $\alpha' = 1$ e $\alpha'' = 3$, nei quali la funzione obiettivo vale 0. Dunque, la successione dei valori $\{\phi(\alpha_k)\}$ converge, ma al valore sbagliato. Il problema, in questo caso, è che la funzione ϕ diminuisce sì ad ogni iterazione, ma sempre di meno. Per evitare questo comportamento possiamo imporre una *condizione di sufficiente riduzione* del valore della funzione f.

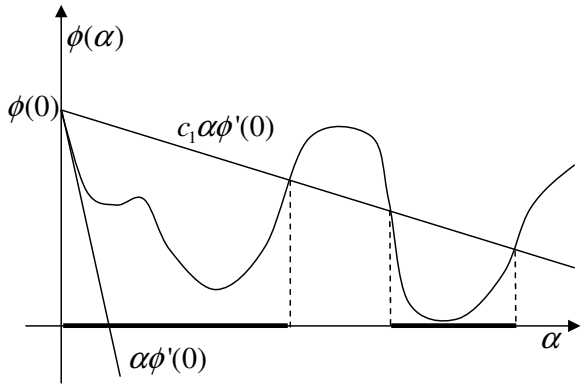


Figura 6: Condizione di Armijo

Condizione 1 *Condizione di sufficiente riduzione della f. Ad ogni iterazione dell'algoritmo, il punto incrementato deve soddisfare la seguente diseguaglianza:*

$$f(\boldsymbol{x} + \alpha \boldsymbol{d}) \leq f(\boldsymbol{x}) + c_1 \alpha \nabla f(\boldsymbol{x})^T \boldsymbol{d}, \tag{2}$$

con $c_1 \in (0, 1)$. In altre parole, la riduzione di f deve essere proporzionale sia alla lunghezza del passo α, sia alla derivata direzionale $\nabla f(\boldsymbol{x})^T \boldsymbol{d}$. La disuguaglianza (2) è detta *condizione di Armijo*. In termini di $\phi(\alpha)$ l'espressione diviene: $\phi(\alpha) \leq \phi(0) + \alpha c_1 \phi'(0)$. Come evidenzia

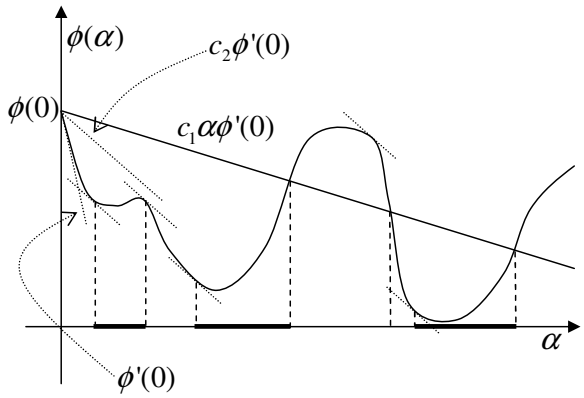


Figura 7: Condizioni di Wolfe

la Figura 6, consideriamo validi i valori di $\alpha > 0$ per i quali il grafico di $\phi(\alpha)$ sta al di sotto della retta di pendenza $c_1\phi'(0)$ (dunque negativa), passante per il punto $(0, \phi(0))$.
La condizione di sufficiente riduzione da sola non basta ad assicurare che l'algoritmo faccia progressi ragionevoli in quanto essa è soddisfatta da tutti i valori di α sufficientemente piccoli. Per evitare passi troppo piccoli si aggiunge una seconda condizione.

Condizione 2 *Condizione di curvatura. Ad ogni iterazione dell'algoritmo, il punto incrementato deve soddisfare la seguente disuguaglianza:*

$$\nabla f(\boldsymbol{x} + \alpha \boldsymbol{d})^T \boldsymbol{d} \geq c_2 \nabla f(\boldsymbol{x})^T \boldsymbol{d} \tag{3}$$

con $c_2 \in (c_1, 1)$. In termini di $\phi(\alpha)$ l'espressione diviene: $\phi'(\alpha) \geq c_2\phi(0)$. La pendenza della funzione $\phi(\alpha)$ nei valori di α accettabili deve essere maggiore, di un fattore costante, alla pendenza che la funzione assume nel punto 0. Questo ha senso, poiché se la pendenza $\phi'(\alpha)$ è molto negativa abbiamo l'indicazione che possiamo ridurre f significativamente muovendoci ulteriormente nella direzione scelta. In sostanza, la condizione di curvatura vincola il passo α ad essere abbastanza lungo da percepire una significativa diminuzione (in valore assoluto) della derivata direzionale, fatto questo che indica un avvicinamento al minimo della funzione obiettivo. Si noti che anche i valori di α per i quali $\phi'(\alpha) > 0$ soddisfano la condizione di curvatura. Un valore tipicamente usato per c_1 è 10^{-4}, mentre il valore di c_2 dipende dal metodo usato per la scelta della direzione (0.9 se stiamo usando Newton o quasi-Newton e 0.1 se usiamo i metodi a gradiente coniugato). Le due condizioni appena introdotte sono note come *condizioni di Wolfe*. Come si vede dalla Figura 7, una valore del passo α può soddisfare le condizioni di Wolfe senza essere particolarmente vicino ad un minimo locale di $\phi(\alpha)$. Tali condizioni possono essere rinforzate per garantire che la lunghezza del passo stia almeno in un intorno di un minimo locale. Le *condizioni di Wolfe in senso forte* richiedono che α soddisfi, oltre alla condizione di Armijo, anche la disequazione:

$$|\nabla f(\boldsymbol{x} + \alpha \boldsymbol{d})^T \boldsymbol{d}| \leq c_2 |\nabla f(\boldsymbol{x})^T \boldsymbol{d}| \tag{4}$$

con $0 < c_1 < c_2 < 1$. In pratica si impedisce alla derivata $\phi'(\alpha)$ di essere troppo positiva. Le considerazioni appena introdotte vengono utilizzate all'interno di metodi iterativi, metodi ai quali dobbiamo ricorrere ogniqualvolta non si è in grado di calcolare analiticamente i valori di α per i quali $\phi'(\alpha) = 0$. Per nostra fortuna le condizioni sopra riportate valgono per un'ampia classe di funzioni.

Teorema 6 *Data una direzione di discesa $\boldsymbol{d}$ per una funzione $f(\boldsymbol{x})$ di classe C^1, se la funzione $\phi(\alpha)$ non è inferiormente limitata per $\alpha > 0$, allora esistono $0 < \alpha_1 < \alpha_2$ tali che per ogni $\alpha \in [\alpha_1, \alpha_2]$ valgono le condizioni di Wolfe anche in senso forte.*

Vediamo ora alcuni metodi per determinare il valore del passo α in modo da tener conto delle indicazioni fornite dalle condizioni di Wolfe.

6.1.1 Backtracking e metodo di Armijo

Questo metodo iterativo genera i valori di α in modo abbastanza accurato e converge con accettabile rapidità, pur facendo a meno della verifica puntuale della *condizione di curvatura* e utilizzando esplicitamente solo la condizione di sufficiente riduzione della funzione f. L'algoritmo è di tipo *backtracking*. Indicheremo con $\alpha_1, \alpha_2, \ldots, \alpha_i, \ldots$ i valori di α generati alle varie iterazioni, mentre α^* indica il valore restituito dall'algoritmo, che verrà quindi utilizzato come passo nell'iterazione (1).

Metodo di Armijo;
{
  Scegli $\alpha_0 \in \mathbb{R}$; $\alpha := \alpha_0$;
  While $f(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k) > f(\boldsymbol{x}_k) + \alpha c_1 \nabla f(\boldsymbol{x}_k)^T \boldsymbol{d}_k$;
    $\alpha := \sigma\alpha$; `/* backtracking */`
  $\alpha^* := \alpha$;
}

L'approccio backtracking consiste nello scegliere, inizialmente, un valore α_0. Se già α_0 soddisfa la condizione di sufficiente riduzione, il procedimento termina e restituisce α_0. Altrimenti, si moltiplica α_0 per un fattore di contrazione $0 < \sigma < 1/2$ e si prova il valore così generato. Il procedimento prosegue in questo modo fino a trovare un valore $\alpha^* = \sigma^i \alpha_0$ tale da soddisfare la condizione $\phi(\alpha^*) \leq \phi(0) + \alpha^* c_1 \phi'(0)$. L'idea è che il valore α^* restituito dal metodo, oltre a soddisfare tale condizione, non sia troppo piccolo, in quanto il valore trovato all'iterazione precedente, ossia $\sigma^{i-1}\alpha_0$, non era stato ritenuto soddisfacente, ossia era ancora troppo grande. La versione base, nella quale il metodo backtracking prevede di utilizzare ad ogni iterazione sempre lo stesso valore del fattore di contrazione σ, prende il nome di *metodo di Armijo*.

Teorema 7 *Dato un punto $\boldsymbol{x}$ e una direzione di discesa $\boldsymbol{d}$, il metodo di Armijo determina in un numero finito di iterazioni un punto α^* tale che:*

$$f(\boldsymbol{x} + \alpha^* \boldsymbol{d}) < f(\boldsymbol{x})$$

Dimostrazione Per assurdo, se l'algoritmo non terminasse mai sarebbe sempre vera la relazione

$$\frac{f(\boldsymbol{x} + \sigma^j \alpha_0 \boldsymbol{d}) - f(\boldsymbol{x})}{\sigma^j \alpha_0} > c_1 \nabla f(\boldsymbol{x})^T \boldsymbol{d}$$

e al limite per $i \to \infty$ si avrebbe $\sigma^j \alpha_0 \to 0$, cioè $\nabla f(\boldsymbol{x})^T \boldsymbol{d} > c_1 \nabla f(\boldsymbol{x})^T \boldsymbol{d}$ che è impossibile per $0 < c_1 < 1$ poiché $\boldsymbol{d}$ è direzione di discesa ($\nabla f(\boldsymbol{x})^T \boldsymbol{d} < 0$). □

6.1.2 Metodo di ricerca esatto

Si tratta di calcolare analiticamente i valori α^* per i quali

$$\phi'(\alpha^*) = \nabla f(\boldsymbol{y})^T \boldsymbol{d} = \nabla f(\boldsymbol{x} + \alpha^* \boldsymbol{d})^T \boldsymbol{d} = 0.$$

Come si vede dalla Figura 8 è possibile dare una interpretazione geometrica alle condizioni

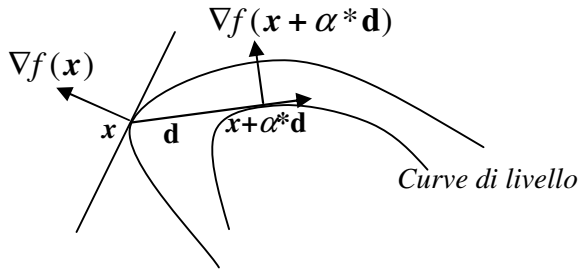


Figura 8: Scelta del passo ottima

analitiche che richiedono che un punto di minimo per la funzione $\phi(\alpha)$ sia un punto stazionario, quindi un punto in cui $\phi'(\alpha) = 0$. Consideriamo una direzione di discesa $\boldsymbol{d}$, che forma quindi con $\nabla f(\boldsymbol{x})$ un angolo ottuso, e spostiamoci in tale direzione. Il primo valore $\alpha^* > 0$ tale che $\phi'(\alpha^*) = 0$ individua un punto $\boldsymbol{x} + \alpha^* \boldsymbol{d}$, in cui l'angolo tra $\boldsymbol{d}$ e $\nabla f(\boldsymbol{x} + \alpha^* \boldsymbol{d})$ è un angolo retto. Proseguendo ulteriormente nella direzione $\boldsymbol{d}$ ci muoveremmo lungo una retta che giace nel piano tangente nel punto $\boldsymbol{x} + \alpha^* \boldsymbol{d}$ alla curva di livello passante per $\boldsymbol{x} + \alpha^* \boldsymbol{d}$ senza ottenere quindi alcun ulteriore miglioramento. Purtroppo solo raramente è possibile effettuare una ricerca esatta, essendo questa di solito onerosa dal punto di vista computazionale. Un caso però, merita un trattamento a parte: il caso di funzioni quadratiche (cfr. 4.3).

Nel caso di funzioni quadratiche con matrice Q simmetrica e definita positiva,

$$f(\boldsymbol{x}) = \frac{1}{2} \boldsymbol{x}^T Q \boldsymbol{x} - \boldsymbol{b}^T \boldsymbol{x},$$

fissata una direzione di discesa $\boldsymbol{d}$, si ricava

$$\phi(\alpha) = f(\boldsymbol{x} + \alpha \boldsymbol{d}) = \frac{1}{2} (\boldsymbol{x} + \alpha \boldsymbol{d})^T Q (\boldsymbol{x} + \alpha \boldsymbol{d}) - \boldsymbol{b}^T (\boldsymbol{x} + \alpha \boldsymbol{d}).$$

Derivando $\phi(\alpha)$ rispetto ad α e uguagliando a zero si ottiene l'espressione

$$(Q(\boldsymbol{x} + \alpha \boldsymbol{d}) - \boldsymbol{b})^T \boldsymbol{d} = 0$$

dalla quale si ricava per α il valore ottimo

$$\alpha^* = -\frac{\boldsymbol{x}^T Q \boldsymbol{d} - \boldsymbol{b}^T \boldsymbol{d}}{\boldsymbol{d}^T Q \boldsymbol{d}} = -\frac{\nabla f(\boldsymbol{x})^T \boldsymbol{d}}{\boldsymbol{d}^T Q \boldsymbol{d}}. \tag{5}$$

Tale risultato ci tornerà utile quando per studiare le proprietà di convergenza di alcuni algoritmi li applicheremo a problemi di tipo quadratico.

6.1.3 Metodo di interpolazione

Questo metodo risulta efficace se la funzione $\phi(\alpha)$ non varia in maniera troppo accentuata da punto a punto. Esso può essere visto come un miglioramento del metodo di Armijo (cfr. Sezione 6.1.1) ed è parte integrante delle tecniche più sofisticate per la scelta di α.
Supponiamo che il valore iniziale α_0 sia dato. Se esso soddisfa la condizione di sufficiente riduzione, $\phi(\alpha_0) \leq \phi(0) + \alpha_0 c_1 \phi'(0)$ terminiamo la ricerca. Altrimenti, costruiamo una approssimazione quadratica di $\phi(\alpha)$ determinando i valori a, b, e c di un polinomio di secondo grado $\phi_q(\alpha) = a\alpha^2 + b\alpha + c$ in modo che soddisfi le condizioni $\phi_q(0) = \phi(0), \phi_q'(0) = \phi'(0)$ e $\phi_q(\alpha_0) = \phi(\alpha_0)$.

> Dalla prima relazione ricaviamo $c = \phi(0)$. Inoltre, da $\phi_q'(\alpha) = 2a\alpha + b$, e dalla seconda relazione ricaviamo $b = \phi'(0)$. Infine da $\phi_q(\alpha_0) = a\alpha_0^2 + \phi'(0)\alpha_0 + \phi(0)$ e dalla terza relazione ricaviamo $a = (\phi(\alpha_0) - \phi'(0)\alpha_0 - \phi(0))/\alpha_0^2$.

Tale parabola è convessa in quanto $a > 0$ (lo si dimostri per esercizio). Ora si cerca il minimo, α_1 della parabola, imponendo $\phi_q'(\alpha) = 2a\alpha + b = 0$. Otteniamo così il valore:

$$\alpha_1 = \frac{-\phi'(0)\alpha_0^2}{2(\phi(\alpha_0) - \phi'(0)\alpha_0 - \phi(0))}$$

A questo punto se α_1 soddisfa la condizione di sufficiente riduzione abbiamo terminato, altrimenti iteriamo il procedimento tenendo conto anche del nuovo valore appena calcolato, passando cioè ad un'approssimazione cubica. La determinazione del polinomio dipenderà dal passaggio della parabola

$$\phi_c(\alpha) = a\alpha^3 + b\alpha^2 + \alpha\phi'(0) + \phi(0)$$

per tre punti: $\phi_c(0) = \phi(0)$, $\phi_c'(0) = \phi'(0)$, $\phi_c(\alpha_0) = \phi(\alpha_0)$ e $\phi_c(\alpha_1) = \phi(\alpha_1)$. Si ricava un sistema lineare di due equazioni nelle due incognite a e b la cui soluzione è

$$\begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{\alpha_0^2\alpha_1^2(\alpha_1 - \alpha_0)} \begin{bmatrix} \alpha_0^2 & -\alpha_1^2 \\ -\alpha_0^3 & \alpha_1^3 \end{bmatrix} \begin{bmatrix} \phi(\alpha_1) - \phi(0) - \phi'(0)\alpha_1 \\ \phi(\alpha_0) - \phi(0) - \phi'(0)\alpha_0 \end{bmatrix}.$$

Derivando rispetto ad α ed uguagliando a zero si ricava che un minimo locale di $\phi_c(\alpha)$, a_2, cade nell'intervallo $[0, a_1]$ ed è dato da

$$\alpha_2 = \frac{-b + \sqrt{b^2 - 3a\phi'(0)}}{3a}.$$

Si verificano le condizioni di sufficiente riduzione per a_2 e se necessario si riapplica il metodo, escludendo il primo punto calcolato, tra quelli presenti, nell'approssimazione cubica. Come si può osservare questa procedura genera una sequenza decrescente di valori di α, tale che ogni α_i non è troppo più piccolo del suo predecessore α_{i-1}. Questa versione è pensata per effettuare il minor numero possibile di calcoli del vettore gradiente $\nabla f(\boldsymbol{x})$. L'interpolazione cubica è estremamente efficace e garantisce, di solito, un tasso di convergenza quadratico.

6.1.4 Inizializzazione del passo α_0

La scelta del valore al quale inizializzare α_0 dipende dal tipo di tecnica adottata per scegliere la direzione di discesa $\boldsymbol{d}$. Nei metodi di Newton e quasi-Newton, si dovrebbe sempre scegliere $\alpha_0 = 1$. Questo permette di usare sempre il passo unitario ogniqualvolta le condizioni di Wolfe

sono soddisfatte, garantendo il mantenimento dell'alto tasso di convergenza di questi metodi. Con le tecniche di gradiente o di gradiente coniugato, è invece importante usare l'informazione corrente per effettuare delle buone inizializzazioni.
Una strategia molto diffusa si basa sull'assunzione che non vari eccessivamente, fra una iterazione e la successiva, il valore della derivata direzionale. Quindi si sceglie α_0 in modo che $\alpha_0 \nabla f(\boldsymbol{x}_k)^T \boldsymbol{d}_k = \alpha_{k-1} \nabla f(\boldsymbol{x}_{k-1})^T \boldsymbol{d}_{k-1}$:

$$\alpha_0 = \alpha_{k-1} \frac{\nabla f(\boldsymbol{x}_{k-1})^T \boldsymbol{d}_{k-1}}{\nabla f(\boldsymbol{x}_k)^T \boldsymbol{d}_k}.$$

Un'altra tecnica, approssima la funzione $\phi(\alpha)$ con una forma quadratica, imponendo l'uso dei tre valori: $f(\boldsymbol{x}_{k-1})$, $f(\boldsymbol{x}_k)$ e $\nabla f(\boldsymbol{x}_{k-1})^T \boldsymbol{d}_{k-1}$ e ricavando il minimo

$$\alpha_0 = \frac{2(f(\boldsymbol{x}_k) - f(\boldsymbol{x}_{k-1}))}{\phi'(0)}.$$

In questo caso, se si adotta il criterio di scegliere $\alpha_0 = \min\{1, 1.01\alpha_0\}$ si garantisce sempre la possibilità di scegliere il passo unitario, mantenendo il tasso di convergenza superlineare dei metodi di Newton e quasi-Newton.

6.1.5 Tecniche che non calcolano le derivate

Nei casi in cui non sia possibile, o troppo costoso, il calcolo del vettore gradiente, si possono adottare tecniche che non fanno uso delle informazioni del primo ordine. In questi casi è necessario individuare un intervallo di ricerca $[0, \alpha_0]$ ove cercare α^*, nel quale la funzione sia convessa o pseudo-convessa (convessa almeno nell'intervallo che si sta considerando). In questo caso si possono applicare tecniche che richiedono solo il calcolo della funzione obiettivo $f(\boldsymbol{x})$. Fra di esse ricordiamo:

- Sezione Aurea
- Fibonacci

e rimandiamo ad altri testi per un approfondimento.

6.2 Scelta della direzione $\boldsymbol{d}_k$

In questa sezione affrontiamo il problema di determinare la direzione $\boldsymbol{d}$ nei metodi basati su ricerca lineare. Iniziamo proponendo una condizione la cui soddisfazione, assieme al rispetto delle condizioni introdotte per α, garantisce la convergenza degli algoritmi di discesa.

Condizione 3 *Una direzione di discesa* $\boldsymbol{d}$ *soddisfa la condizione d'angolo se esiste* $1 > \varepsilon > 0$ *tale che:*

$$\nabla f(\boldsymbol{x})^T \boldsymbol{d} \leq -\varepsilon \, \|\nabla f(\boldsymbol{x})\| \cdot \|\boldsymbol{d}\| \tag{6}$$

La condizione impone di non scegliere mai direzioni troppo vicine al piano tangente alle curve di livello passanti per il punto corrente. L'interpretazione geometrica risulta più chiara se scriviamo la seguente espressione

$$\cos\theta_k = \frac{-\nabla f(\boldsymbol{x})_k^T \boldsymbol{d}_k}{||\nabla f(\boldsymbol{x})_k|| \cdot ||\boldsymbol{d}_k||} \tag{7}$$

dove θ_k misura l'angolo fra la direzione $\boldsymbol{d}_k$ e la direzione dell'antigradiente. In tal modo possiamo riscrivere la condizione d'angolo come

$$\varepsilon \leq \cos\theta_k. \tag{8}$$

Si osservi che per $\boldsymbol{d} = -\nabla f(\boldsymbol{x})$ abbiamo $\cos\theta_k = 1$, quindi l'antigradiente certamente soddisfa la relazione (8). Nonostante la sua semplicità, tale ragionevole condizione è estremamente potente in quanto il suo soddisfacimento, assieme al soddisfacimento delle condizioni di Wolfe per il valore del passo α, è sufficiente a garantire la convergenza *globale* di qualsiasi algoritmo di discesa.

Teorema 8 *Sia data una funzione f limitata inferiormente in $\mathbb{R}^n$, di classe C^1 in un insieme aperto S contenente l'insieme di livello $L_f(\boldsymbol{x}_0) = \{\boldsymbol{x} : f(\boldsymbol{x}) \leq f(\boldsymbol{x}_0)\}$, dove $\boldsymbol{x}_0$ è il punto di partenza di un algoritmo iterativo di tipo (1). Si supponga che ad ogni iterazione k la direzione $\boldsymbol{d}_k$ soddisfi la condizione d'angolo ed il passo α_k soddisfi le condizioni di Wolfe, per opportuni valori dei parametri c_1, c_2 ed ε. Si supponga inoltre che il gradiente $\nabla f(\boldsymbol{x})$ sia Lipschitz continuo in S, cioè che esista una costante $L > 0$ tale che il gradiente $\nabla f(\boldsymbol{x})$ soddisfi la seguente relazione*

$$||\nabla f(\boldsymbol{x}) - \nabla f(\tilde{\boldsymbol{x}})|| \leq L||\boldsymbol{x} - \tilde{\boldsymbol{x}}||, \text{ per tutte le coppie di punti } \boldsymbol{x}, \tilde{\boldsymbol{x}} \in S.$$

Allora la successione dei punti $\{\boldsymbol{x}_k\}$ è tale che il gradiente di $f(\boldsymbol{x})$ si annulla in un numero finito di passi o converge a zero asintoticamente:

$$\begin{gathered}\text{esiste } k \text{ con } \nabla f(\boldsymbol{x}_k) = \boldsymbol{0} \\ \text{oppure la successione } \{\nabla f(\boldsymbol{x}_k)\} \to \boldsymbol{0}\end{gathered}$$

Dimostrazione Dalla condizione di curvatura (3)

$$\nabla f(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k)^T \boldsymbol{d}_k \geq c_2 \nabla f(\boldsymbol{x}_k)^T \boldsymbol{d}_k$$

ricaviamo

$$(\nabla f(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k) - \nabla f(\boldsymbol{x}_k))^T \boldsymbol{d}_k \geq (c_2 - 1)\nabla f(\boldsymbol{x}_k)^T \boldsymbol{d}_k$$

mentre la condizione di Lipschitz e la disuguaglianza di Cauchy-Schwarz[2] implicano che

$$(\nabla f(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k) - \nabla f(\boldsymbol{x}_k))^T \boldsymbol{d}_k \leq \alpha_k L||\boldsymbol{d}_k||^2.$$

Combinando queste due relazioni ricaviamo

$$\alpha_k \geq \frac{c_2 - 1}{L} \frac{\nabla f(\boldsymbol{x}_k)^T \boldsymbol{d}_k}{||\boldsymbol{d}_k||^2}.$$

Se sostituiamo il valore di α_k così ottenuto nella condizione di Armijo (2)

$$f(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k) \leq f(\boldsymbol{x}_k) + c_1 \alpha_k \nabla f(\boldsymbol{x}_k)^T \boldsymbol{d}_k,$$

[2] La disuguaglianza di Cauchy-Schwarz afferma che se $\boldsymbol{x}$ e $\boldsymbol{y}$ sono vettori di $\mathbb{R}^n$ allora vale la relazione $(\boldsymbol{x}^T\boldsymbol{y})^2 \leq (\boldsymbol{x}^T\boldsymbol{x})(\boldsymbol{y}^T\boldsymbol{y})$

ricaviamo

$$f(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k) \leq f(\boldsymbol{x}_k) - c_1 \frac{1 - c_2}{L} \frac{(\nabla f(\boldsymbol{x}_k)^T \boldsymbol{d}_k)^2}{||\boldsymbol{d}_k||^2}.$$

Usando la definizione (7) possiamo riscrivere questa relazione come

$$f(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k) \leq f(\boldsymbol{x}_k) - c \cos^2 \theta_k ||\nabla f(\boldsymbol{x}_k)||^2,$$

dove $c = c_1(1 - c_2)/L$. Sommando questa espressione su tutti gli indici minori o uguali a k ricaviamo

$$f(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k) \leq f(\boldsymbol{x}_0) - c \sum_{j=0}^{k} \cos^2 \theta_j ||\nabla f(\boldsymbol{x}_j)||^2.$$

Poiché la funzione $f(\boldsymbol{x})$ è inferiormente limitata, ne segue che il valore $f(\boldsymbol{x}_0) - f(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k)$ è limitato da una qualche costante positiva per tutti i k. Questo significa che vale la seguente relazione, detta *condizione di Zoutendijk*

$$\sum_{j=0}^{\infty} \cos^2 \theta_j ||\nabla f(\boldsymbol{x}_j)||^2 < \infty. \tag{9}$$

La condizione di Zoutendijk implica che

$$\cos^2 \theta_j ||\nabla f(\boldsymbol{x}_j)||^2 \to 0$$

Ora, poiché per ipotesi vale la condizione d'angolo

$$0 < \varepsilon \leq \cos \theta_k \quad \text{per tutti i valori di } k$$

segue immediatamente che

$$\lim_{k \to \infty} ||\nabla f(\boldsymbol{x}_j)|| = 0.$$

□

Si osservi che la condizione sulle linee di livello equivale a chiedere che la funzione non sia inferiormente illimitata ed è quindi ragionevole poiché altrimenti il problema non sarebbe ben posto. La condizione che il gradiente sia Lipschitz continuo è spesso richiesta per poter dimostrare la convergenza locale degli algoritmi ed è spesso soddisfatta in pratica.
Vediamo ora, nel dettaglio, alcuni modi differenti di scegliere la direzione $\boldsymbol{d}$.

6.2.1 Metodo del gradiente

Nel metodo del gradiente la direzione, $\boldsymbol{d}$, coincide con l'antigradiente: $\boldsymbol{d} = -\nabla f(\boldsymbol{x})$. Tale direzione massimizza il decremento della funzione obiettivo. Nei metodi iterativi si ha quindi

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k \nabla f(\boldsymbol{x}_k). \tag{10}$$

Per ricavare una buona stima del valore del passo ottimo α_k possiamo approssimare $f(\boldsymbol{x})$ con una funzione quadratica, ottenuta fermandoci al secondo ordine nello sviluppo in serie di Taylor di f.

$$f(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k) \approx \phi(\alpha) = f(\boldsymbol{x}_k) + \alpha \boldsymbol{d}_k^T \nabla f(\boldsymbol{x}_k) + \frac{1}{2}\alpha^2 \boldsymbol{d}_k^T H(\boldsymbol{x}_k) \boldsymbol{d}_k$$

Calcoliamo ora la derivata di $\phi(\alpha)$ rispetto ad α come fatto nella Sezione 6.1.2, ottenendo

$$\phi'(\alpha) = \boldsymbol{d}_k^T \nabla f(\boldsymbol{x}_k) + \alpha \boldsymbol{d}_k^T H(\boldsymbol{x}_k) \boldsymbol{d}_k$$

Se vale la condizione $\boldsymbol{d}_k^T H(\boldsymbol{x}_k) \boldsymbol{d}_k \neq 0$, allora possiamo imporre $\phi'(\alpha) = 0$ e ricavare

$$\alpha = -\frac{\boldsymbol{d}_k^T \nabla f(\boldsymbol{x}_k)}{\boldsymbol{d}_k^T H(\boldsymbol{x}_k) \boldsymbol{d}_k}$$

Metodo del gradiente;
{
    Scegli $\boldsymbol{x}_0 \in \mathbb{R}^n$; $k := 0$;
    While $\nabla f(\boldsymbol{x}_k) \neq \emptyset$;
    {
        $\boldsymbol{d}_k := -\nabla f(\boldsymbol{x}_k)$ /* direzione di discesa */
        calcola $\alpha_k \in \mathbb{R}$; /* passo lungo $\boldsymbol{d}_k$ */
        $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha \boldsymbol{d}_k$;
        $k := k + 1$;
    }
}

Nell'ipotesi di ricerca lineare esatta ed approssimazione mediante una funzione quadratica, il passo iterativo (10) diviene:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \frac{\nabla f(\boldsymbol{x}_k)^T \nabla f(\boldsymbol{x}_k)}{\nabla f(\boldsymbol{x}_k)^T H(\boldsymbol{x}_k) \nabla f(\boldsymbol{x}_k)} \nabla f(\boldsymbol{x}_k) \tag{11}$$

La direzione dell'antigradiente sembra la scelta più ovvia in un metodo di discesa: seguire ad ogni iterazione la direzione ed il passo di massimo decremento del valore della funzione obiettivo. Ma fino a che punto questa scelta è vincente? Per scoprirlo occorre analizzare il tasso di convergenza del metodo, e per farlo occorre scegliere una classe di problemi per la quale tale analisi sia praticabile.

6.2.2 Analisi del metodo del gradiente

Analizziamo le prestazioni del metodo del gradiente quando viene applicato a *problemi quadratici* con matrice Q definita positiva

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T Q \boldsymbol{x} - \boldsymbol{b}^T \boldsymbol{x}. \tag{12}$$

In questo caso il passo iterativo (11) diviene

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \frac{(Q\boldsymbol{x}_k - \boldsymbol{b})^T (Q\boldsymbol{x}_k - \boldsymbol{b})}{(Q\boldsymbol{x}_k - \boldsymbol{b})^T Q (Q\boldsymbol{x}_k - \boldsymbol{b})} (Q\boldsymbol{x}_k - \boldsymbol{b})$$

Per questo tipo di problemi siamo in grado di ricavare analiticamente il valore della soluzione ottima, $\boldsymbol{x}^* = Q^{-1}\boldsymbol{b}$, e di conseguenza possiamo utilizzare questa informazione (di solito non disponibile, altrimenti non avremmo bisogno degli algoritmi iterativi...) per calcolare la forma esatta della riduzione dell'errore di approssimazione tra iterazioni successive.
Per far ciò useremo la norma pesata

$$||\boldsymbol{x}||_Q^2 = \frac{1}{2}\boldsymbol{x}^T Q\boldsymbol{x}$$

e quindi la riduzione dell'errore calcolata sarà:

$$\frac{\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|_Q}{\|\boldsymbol{x}_k - \boldsymbol{x}^*\|_Q}$$

Come possiamo interpretare tale norma? Per rispondere dobbiamo effettuare alcune trasformazioni. Innanzitutto riscriviamo $f(\boldsymbol{x}^*)$.

$$f(\boldsymbol{x}^*) = \frac{1}{2}\boldsymbol{x}^{*T}Q\boldsymbol{x}^* - \boldsymbol{b}^T\boldsymbol{x}^* = \frac{1}{2}\boldsymbol{x}^{*T}Q\boldsymbol{x}^* - (Q\boldsymbol{x}^*)^T\boldsymbol{x}^* = -\frac{1}{2}\boldsymbol{x}^{*T}Q\boldsymbol{x}^*.$$

Ora misuriamo il quadrato della distanza fra il punto corrente $\boldsymbol{x}_k$ ed il punto di ottimo con la nuova metrica:

$$||\boldsymbol{x}_k - \boldsymbol{x}^*||_Q^2 = \frac{1}{2}(\boldsymbol{x}_k - \boldsymbol{x}^*)^T Q(\boldsymbol{x}_k - \boldsymbol{x}^*) = \frac{1}{2}\boldsymbol{x}_k^T Q\boldsymbol{x}_k + \frac{1}{2}\boldsymbol{x}^{*T}Q\boldsymbol{x}^* - \boldsymbol{x}_k^T Q\boldsymbol{x}^* = f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*). \quad (13)$$

dove, di nuovo, abbiamo fatto uso della relazione $Q\boldsymbol{x}^* = \boldsymbol{b}$ e del fatto che $\boldsymbol{x}_k^T\boldsymbol{b} = \boldsymbol{b}^T\boldsymbol{x}_k$.
Questa norma misura quindi la distanza fra il valore corrente della funzione obiettivo ed il valore ottimo.
Con un po' di passaggi è possibile ricavare il seguente risultato intermedio.

Lemma 1 *Se l'algoritmo del gradiente viene applicato ad un problema di programmazione quadratica convesso la forma esatta della riduzione dell'errore di approssimazione tra iterazioni successive misurata con la norma pesata $||\cdot||_Q$ è*

$$\frac{\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|_Q}{\|\boldsymbol{x}_k - \boldsymbol{x}^*\|_Q} = \left\{1 - \frac{((Q\boldsymbol{x}_k - \boldsymbol{b})^T(Q\boldsymbol{x}_k - \boldsymbol{b}))^2}{((Q\boldsymbol{x}_k - \boldsymbol{b})^T Q(Q\boldsymbol{x}_k - \boldsymbol{b}))((Q\boldsymbol{x}_k - \boldsymbol{b})^T Q^{-1}(Q\boldsymbol{x}_k - \boldsymbol{b}))}\right\}^{\frac{1}{2}}$$

Dimostrazione Vedi Appendice □

Abbiamo quindi una misura del decremento di f ad ogni iterazione. Il termine fra parentesi è però decisamente difficile da interpretare. Per nostra fortuna tale quantità si dimostra dominabile tramite una funzione degli autovalori di Q come ci dimostra il seguente teorema dovuto a Kantorovic.

Teorema 9 *Data una matrice Q definita positiva vale la seguente relazione*

$$\frac{(\boldsymbol{x}^T\boldsymbol{x})^2}{(\boldsymbol{x}^T Q\boldsymbol{x})(\boldsymbol{x}^T Q^{-1}\boldsymbol{x})} \geq \frac{4\lambda_m\lambda_M}{(\lambda_m + \lambda_M)^2}$$

valida per ogni $\boldsymbol{x} \in \mathbb{R}^n$ e dove λ_M e $\lambda_m > 0$ sono l'autovalore massimo e minimo, rispettivamente, di Q.

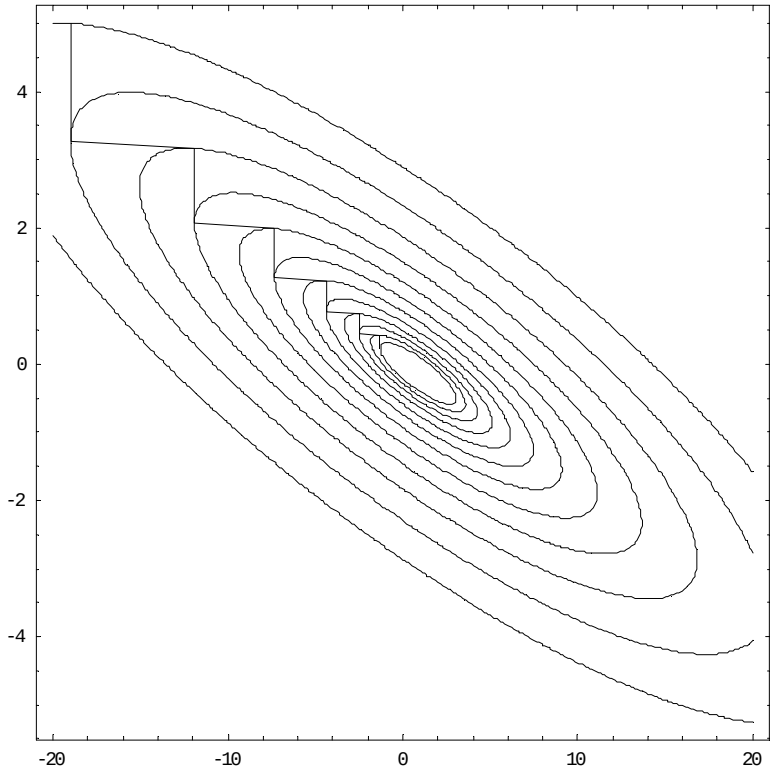


Figura 9: Andamento zigzagante del metodo del gradiente

Dimostrazione Vedi Appendice □

Combinando i due precedenti risultati possiamo ricavare il seguente fondamentale risultato.

Teorema 10 *Il metodo del gradiente con scelta del passo ottimale, quando viene applicato al problema quadratico* $\min \frac{1}{2}\boldsymbol{x}^T Q\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x}$ *dove* Q *è una matrice simmetrica definita positiva, produce una successione* $\{\boldsymbol{x}_k\}$ *tale che:*

$$\frac{\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|_Q}{\|\boldsymbol{x}_k - \boldsymbol{x}^*\|_Q} \leqslant \left(\frac{\lambda_M - \lambda_m}{\lambda_M + \lambda_m}\right) \tag{14}$$

dove λ_M *e* λ_m *sono l'autovalore massimo e minimo, rispettivamente di* Q*.*

Dimostrazione Denotando $\nabla f(\boldsymbol{x}_k) = (Q\boldsymbol{x}_k - \boldsymbol{b})$ con $\boldsymbol{g}_k$ possiamo scrivere il risultato del Lemma 1 come

$$\frac{\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|_Q}{\|\boldsymbol{x}_k - \boldsymbol{x}^*\|_Q} = \left\{1 - \frac{(\boldsymbol{g}_k^T\boldsymbol{g}_k)^2}{(\boldsymbol{g}_k^T Q\boldsymbol{g}_k)(\boldsymbol{g}_k^T Q^{-1}\boldsymbol{g}_k)}\right\}^{\frac{1}{2}}$$

e dal Teorema di Kantorovic ricaviamo

$$\frac{\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|_Q}{\|\boldsymbol{x}_k - \boldsymbol{x}^*\|_Q} \leq \left\{1 - \frac{4\lambda_m\lambda_M}{(\lambda_m + \lambda_M)^2}\right\}^{\frac{1}{2}} = \left(\frac{\lambda_M - \lambda_m}{\lambda_M + \lambda_m}\right)$$

□

La disuguaglianza (14) mostra che l'algoritmo del gradiente converge con tasso di convergenza lineare.

Esempio La Figura 9 illustra l'andamento delle prime 12 iterazioni dell'algoritmo del gradiente applicato alla funzione quadratica (12) con $Q = \begin{pmatrix} 3 & 12 \\ 12 & 70 \end{pmatrix}$, definita positiva, con $\boldsymbol{b} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, a partire dal punto iniziale $\boldsymbol{x}_0 = \begin{pmatrix} -19 \\ 5 \end{pmatrix}$. Per via analitica si ricava $\boldsymbol{x}^* = \begin{pmatrix} 29/33 \\ -3/22 \end{pmatrix}$. In questo caso $\lambda_M/\lambda_m = 78.7297$.

In generale, al crescere del rapporto λ_M/λ_m (detto *numero condizione* di Q) le prestazioni dell'algorimo degradano e l'andamento zigzagante presentato in Figura 9 si accentua. Per converso, come caso speciale, l'algoritmo del gradiente converge in un solo passo quando $\lambda_m = \lambda_M$, cioè quando tutti gli autovalori di Q sono uguali, e cioè quando Q è un multiplo della matrice identità. In questo caso la direzione punta all'ottimo e le curve di livello sono cerchi concentrici. Anche se la disuguaglianza (14) dà una stima per eccesso, essa risulta una accurata indicazione del comportamento dell'algorimo per valori di $n > 2$.

Quanto detto vale quando l'algoritmo del gradiente viene applicato a funzioni quadratiche, ma cosa succede se lo applichiamo a funzioni qualsiasi? In pratica, il tasso di convergenza del metodo del gradiente è essenzialmente lo stesso per generiche funzioni obiettivo non lineari. Infatti, assumendo che ad ogni iterazione la lunghezza del passo α sia effettuata in modo ottimo, si può dimostrare il seguente teorema.

> **Teorema 11** *Si supponga $f : \mathbb{R}^n \to \mathbb{R}$ di classe C^2 e che le soluzioni generate dall'algorimo del gradiente, con scelta del passo α ottima, convergano al punto $\boldsymbol{x}^*$, dove la matrice hessiana $H(\boldsymbol{x}^*)$ è definita positiva. Sia*
>
> $$r \in \left(\frac{\lambda_M - \lambda_m}{\lambda_M + \lambda_m}, 1 \right)$$
>
> *dove λ_M e λ_m sono l'autovalore massimo e minimo, rispettivamente di $H(\boldsymbol{x}^*)$. Allora esiste un $\overline{k}$ tale che per tutti i $k > \overline{k}$ vale*
>
> $$\frac{f(\boldsymbol{x}_{k+1}) - f(\boldsymbol{x}^*)}{f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*)} \leq r^2.$$

In generale, non possiamo aspettarci che il tasso di convergenza migliori se si adotta una scelta non ottima del passo α. Il Teorema 11 ci mostra che il metodo del gradiente può essere inaccettabilmente lento a convergere, anche quando la matrice hessiana è ragionevolmente ben condizionata (cioè con un valore del numero condizione non troppo grande).

Osserviamo ora, che generalmente il metodo del gradiente è globalmente convergente: è sufficiente che il passo venga scelto in modo da soddisfare le condizioni di Wolfe. Tuttavia la convergenza è estremamente lenta.
Quali sono le ragioni per delle prestazioni così deludenti?
Analizziamo innanzitutto il costo computazionale del metodo. Nel calcolare la direzione di

discesa esso richiede il calcolo del gradiente di $f(\boldsymbol{x})$, che ha un costo computazionale dell'ordine di $O(n^2)$. Nella scelta del passo α, se utilizziamo il metodo di Armijo ci limitiamo ad aggiungere un numero più o meno costante di valutazioni di $f(\boldsymbol{x})$ e di $\nabla f(\boldsymbol{x})$, mentre se utilizziamo l'approssimazione quadratica il costo aumenta sensibilmente: è necessario calcolare l'hessiana $H(\boldsymbol{x})$ con un costo computazionale dell'ordine di $O(n^3)$. In pratica, nessuno utilizza il metodo del gradiente con scelta del passo ottima, proprio per l'elevato costo computazionale. D'altronde è solo quando scegliamo il passo in modo ottimo che garantiamo la convergenza lineare! Quindi abbiamo convergenza lineare al costo dell'ordine di $O(n^3)$ per iterazione. La spiegazione per le prestazioni deludenti, non sta quindi nel basso costo computazionale del metodo, ma nell'uso non efficace delle informazioni elaborate: la matrice Hessiana, quando usata, non contribuisce alla scelta della direzione, ma solo del passo, inoltre il metodo non tiene alcuna traccia nelle scelte all'iterazione corrente della storia passata.
Diviene naturale chiedersi se non si possa far di meglio, nell'ipotesi di poter spendere comunque al più $O(n^3)$ passi di calcolo ad ogni iterazione. Il prossimo metodo risponde positivamente a questa domanda.

6.2.3 Metodo di Newton

Analogamente al metodo del gradiente, anche il metodo di Newton si basa sul concetto di minimizzare un'approssimazione quadratica della funzione f. In questo caso però, si desidera ricavare da tale approssimazione contemporaneamente il valore di α_k e quello di $\boldsymbol{d}_k$.
Sia $f(\boldsymbol{x})$ una funzione con Hessiana continua. Per valori sufficientemente piccoli della norma del vettore incremento $\boldsymbol{h}_k = \alpha_k \boldsymbol{d}_k$ è possibile scrivere:

$$f(\boldsymbol{x}_k + \boldsymbol{h}_k) \approx q(\boldsymbol{h}_k) = f(\boldsymbol{x}_k) + \boldsymbol{h}_k^T \nabla f(\boldsymbol{x}_k) + \frac{1}{2}\boldsymbol{h}_k^T H(\boldsymbol{x}_k)\boldsymbol{h}_k$$

Annullando il gradiente di $q(\boldsymbol{h}_k)$ si ricava

$$\nabla q(\boldsymbol{h}_k) = \nabla f(\boldsymbol{x}_k) + H(\boldsymbol{x}_k)\boldsymbol{h}_k = \boldsymbol{0},$$

da cui, se la matrice hessiana è non singolare, possiamo ottenere

$$\boldsymbol{h}_k = -H(\boldsymbol{x}_k)^{-1}\nabla f(\boldsymbol{x}_k).$$

Quindi quando la matrice hessiana nel punto $\boldsymbol{x}_k$ è invertibile, ad esempio quando essa è definita positiva in quel punto, il vettore $\boldsymbol{h}_k$ è ottimo unico.

Il metodo di Newton *puro* ha quindi la forma

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - H(\boldsymbol{x}_k)^{-1}\nabla f(\boldsymbol{x}_k) \tag{15}$$

Come si comporta il metodo di Newton? Iniziamo con l'osservare che quando esso viene applicato a problemi quadratici

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T Q\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x}$$

e la matrice Q è definita positiva, il metodo trova l'ottimo in un passo. Infatti si ricava

$$\boldsymbol{x}_1 = \boldsymbol{x}_0 - H(\boldsymbol{x}_0)^{-1}\nabla f(\boldsymbol{x}_0) = \boldsymbol{x}_0 - Q^{-1}(Q\boldsymbol{x}_0 - \boldsymbol{b}) = \boldsymbol{x}_0 - \boldsymbol{x}_0 + Q^{-1}\boldsymbol{b} = Q^{-1}\boldsymbol{b}.$$

Se invece la matrice hessiana non è definita positiva, il metodo *non converge*.

Metodo di Newton Puro;
{
    Scegli $\boldsymbol{x}_0 \in \mathbb{R}^n$; $k := 0$;
    While $\nabla f(\boldsymbol{x}_k) \neq \emptyset$;
    {
        $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - H(\boldsymbol{x}_k)^{-1}\nabla f(\boldsymbol{x}_k)$;
        $k := k + 1$;
    }
}

Quanto osservato per i problemi quadratici ha importanti conseguenze:

> *le proprietà di convergenza del metodo sono fortemente legate alla definita positività della matrice hessiana. Quando la funzione $f(\boldsymbol{x})$ è di tipo generale la qualità della direzione dipende essenzialmente dal fatto che la matrice hessiana calcolata nel punto corrente $\boldsymbol{x}_k$ sia definita positiva o meno.*

Infatti, quando la matrice non è definita positiva in un punto, la direzione in quel punto può non essere una direzione di discesa.
Nei casi in cui invece la matrice hessiana è definita positiva, ed in particolare è tale in un intorno di un punto stazionario, vale il seguente fondamentale risultato:

> **Teorema 12** *Data una funzione f di classe C^2, per la quale la matrice hessiana è definita positiva nell'intorno $I(\boldsymbol{x}^*, \varepsilon)$ di un punto stazionario $\boldsymbol{x}^*$ e soddisfa la seguente condizione:*
>
> $$\text{esiste } L > 0 : ||H(\boldsymbol{y}) - H(\boldsymbol{x})|| \leq L||\boldsymbol{y} - \boldsymbol{x}|| \text{ per ogni } \boldsymbol{x}, \boldsymbol{y} \in I(\boldsymbol{x}^*, \varepsilon)$$
>
> *allora la successione $\{\boldsymbol{x}_k\}$ che si ottiene applicando il metodo di Newton puro è tale che, se il punto iniziale $\boldsymbol{x}_0 \in I(\boldsymbol{x}^*, \varepsilon)$, la successione converge a $\boldsymbol{x}^*$ con tasso di convergenza quadratico.*

Il miglioramento rispetto al metodo del gradiente è notevole in fatto di rapidità di convergenza. Ma tale miglioramento viene ad un costo, non solo computazionale: per convergere il metodo richiede di partire vicino ad un punto stazionario. Abbiamo quindi solo una proprietà di *convergenza locale* e non di convergenza globale. Inoltre, il teorema appena enunciato richiede qualcosa in più del fatto che la matrice hessiana sia definita positiva ad ogni iterazione. Infatti, la definita positività della matrice hessiana non basta a far sì che il metodo converga, poiché le direzioni di Newton potrebbero non soddisfare alla condizione d'angolo, potrebbero cioè essere quasi ortogonali al gradiente.
Se invece valgono opportune condizioni sugli autovalori dell'inversa della matrice hessiana, condizioni che ne implicano la definita positività, allora la direzione di Newton soddisfa la condizione d'angolo.

> **Teorema 13** *Se $0 < m < \lambda_m H(\boldsymbol{x}_k)^{-1} \leq \lambda_M H(\boldsymbol{x}_k)^{-1} < M$ per ogni k (proprietà che implica che sia H che H^{-1} siano definite positive), la direzione di Newton soddisfa la condizione d'angolo con $\varepsilon = \frac{m}{M}$:*
>
> $$\nabla f(\boldsymbol{x}_k)^T \boldsymbol{d}_k < -\frac{m}{M}||\nabla f(\boldsymbol{x}_k)|| \cdot ||\boldsymbol{d}_k||$$

Per riottenere una proprietà di *convergenza globale* è quindi necessario modificare il metodo.

Una prima modifica consiste nell'adattare dinamicamente la lunghezza del passo, che nel metodo puro è tenuto costantemente al valore $\alpha = 1$. Si tratta cioè di applicare ad ogni iterazione l'algoritmo di Armijo a partire da $\alpha_0 = 1$.

Una seconda modifica riguarda la scelta della direzione. Se la matrice hessiana non è definita positiva, si può perturbarla, perturbando di conseguenza la direzione $\boldsymbol{d}_k$. Delle diverse tecniche proposte in letteratura, nessuna è esente da controindicazioni. Un metodo semplice ed efficace consiste nel passare ad una matrice $(H(\boldsymbol{x}_k) + \gamma I)$ che, per tutti i valori di γ superiori ad una certa soglia (dipendente da $(H(\boldsymbol{x}_k))$, è definita positiva. Si risolve il sistema

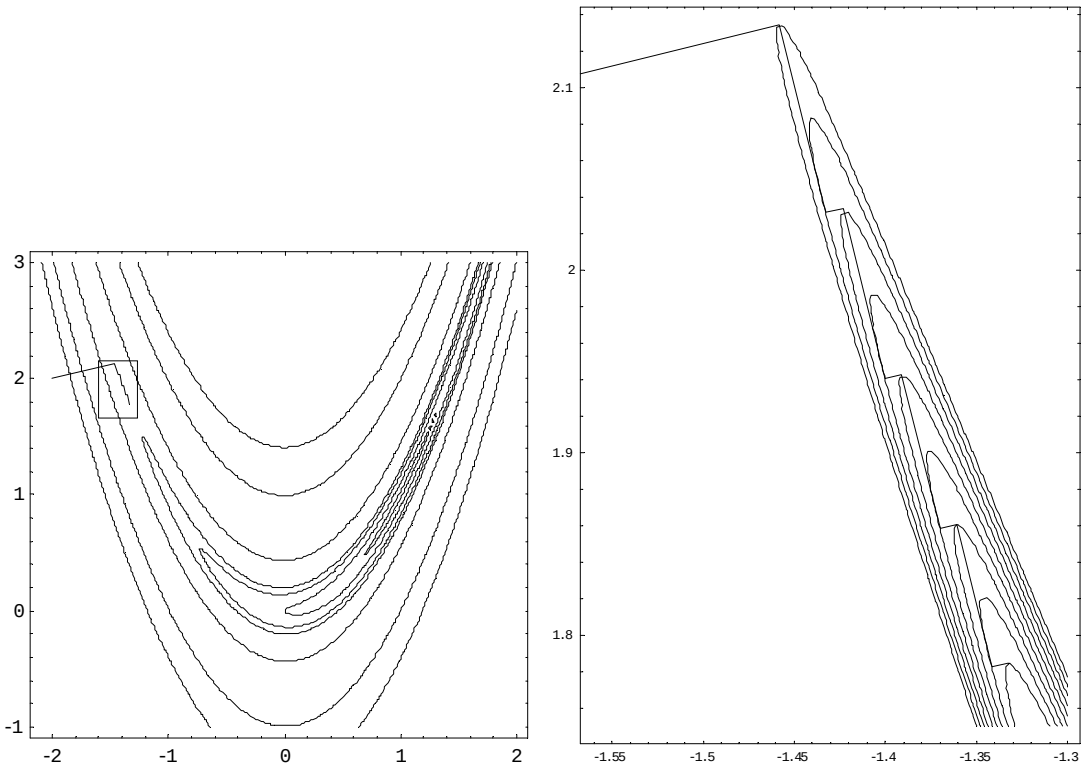
$$(H(\boldsymbol{x}_k) + \gamma I)\boldsymbol{h}_k = -\nabla f(\boldsymbol{x}_k)$$

osservando che al crescere di γ, $\boldsymbol{h}_k \to -\nabla f(\boldsymbol{x}_k)$.
In alternativa, se non si desidera modificare la matrice hessiana, si può alternare l'utilizzo della direzione di Newton, con la direzione dell'antigradiente, scegliendo di volta in volta quale direzione adottare come esito di test sulla condizione d'angolo.

Dunque, se ad ogni iterazione l'Hessiana è definita positiva, e si garantisce il soddisfacimento della condizione d'angolo, il metodo di discesa che utilizza la direzione di Newton converge globalmente ad un punto stazionario. Peraltro, è possibile dimostrare che, da una certa iterazione k in poi, il passo $\alpha_k = 1$ soddisfa le condizioni del metodo di Armijo, e dunque si può di fatto usare il metodo di Newton puro, e la convergenza quadratica è ancora assicurata.

Riassumendo, se la matrice hessiana non è sempre definita positiva il metodo di discesa può risultare inapplicabile per due motivi: o perchè in alcuni punti la matrice hessiana risulta singolare, oppure perchè la direzione di Newton non è una direzione di discesa. Inoltre, può anche accadere che, pur risultando una direzione di discesa, la direzione di Newton e il gradiente siano talmente prossimi ad essere ortogonali che non conviene seguire la direzione di Newton. A questi inconvenienti è possibile ovviare in modo molto semplice, utilizzando la direzione dell'antigradiente quando non sia possibile o conveniente seguire la direzione di Newton.

Metodo di Newton Modificato;
{
  Scegli $\boldsymbol{x}_0 \in \mathbb{R}^n$; $k := 0$;
  While $\nabla f(\boldsymbol{x}_k) \neq \emptyset$;
  {
    if $H(\boldsymbol{x}_k)$ è singolare **then** $\boldsymbol{d}_k$:= $-\nabla f(\boldsymbol{x}_k)$;
    else
    {
      $\boldsymbol{s}$:= $-H(\boldsymbol{x}_k)^{-1}\nabla f(\boldsymbol{x}_k)$;
      if $|\nabla f(\boldsymbol{x}_k)^T \boldsymbol{s}| < \varepsilon ||\nabla f(\boldsymbol{x}_k)|| \cdot ||\boldsymbol{s}||$ **then** $\boldsymbol{d}_k$:= $-\nabla f(\boldsymbol{x}_k)$;
      else
        if $\boldsymbol{s}$ è direzione di discesa **then** $\boldsymbol{d}_k$:=$\boldsymbol{s}$;
                                else $\boldsymbol{d}_k$:=$-\boldsymbol{s}$;
    }
    calcola $\alpha_k \in \mathbb{R}$; /*passo lungo $\boldsymbol{d}_k$*/
    $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$;
    $k := k + 1$;
  }
}



(a) Curve di livello (b) Dieci iterazioni di gradiente

Figura 10: Il metodo del gradiente applicato alla funzione di Rosenbrock.

Il metodo di Newton modificato in modo da considerare tutte queste possibilità è mostrato in figura, dove $\varepsilon > 0$ è una quantità sufficientemente piccola. In sostanza, come si può vedere il metodo sceglie come direzione di discesa quella di Newton, la sua opposta o l'antigradiente, e successivamente effettua una ricerca lineare con metodi standard, in modo da garantire la convergenza globale. L'algoritmo così ottenuto soddisfa la condizione d'angolo e dunque le proprietà di convergenza globale dell'algoritmo sono conservate.

Osservazione. Per trovare $\boldsymbol{h}$ non si inverte $H(\boldsymbol{x})$, ma si risolve il sistema $H(\boldsymbol{x})\boldsymbol{h} = -\nabla f(\boldsymbol{x})$ mediante fattorizzazione di $H(\boldsymbol{x})$ come matrice in forma di prodotto LDL^T, con L diagonale e triangolare inferiore ed L^T triangolare superiore. Dalla fattorizzazione si può inoltre ricavare se la matrice hessiana è definita positiva.

Un ulteriore aspetto problematico del metodo di Newton, modificato o meno, è il richiedere la valutazione delle derivate seconde di $f(\boldsymbol{x})$ per il calcolo di $H(\boldsymbol{x}_k)$. In alternativa, si può ricorrere alle *differenze finite*. Si tratta di approssimare $H(\boldsymbol{x}_k)$ mediante $\frac{1}{2}(\overline{H}_k + \overline{H}_k^T)$ dove la colonna i-esima di $\overline{H}_k$ è $\frac{1}{h}[\nabla f(\boldsymbol{x}_k + \boldsymbol{h}_i) - \nabla f(\boldsymbol{x}_k)]$, con $\boldsymbol{h}_i = (\dots 0, h_i, 0 \dots)^T$, ed h_i è un opportuno incremento dell'i-esima variabile. Purtroppo la matrice risultante può non essere definita positiva (mentre questo serve), si devono calcolare n gradienti e risolvere un sistema lineare ad ogni passo, con un costo quindi dell'ordine di $O(n^3)$ per iterazione.

In sintesi il costo computazionale del metodo di Newton è $O(n^3)$ ad ogni iterazione ed è quindi proponibile solo per sistemi di dimensioni non superiori a qualche decina di migliaia di variabili.

6.2.4 Confronto fra i metodi del gradiente e di Newton

Esemplifichiamo le differenti prestazioni del metodo del gradiente e di quello di Newton su una funzione non convessa nota in letteratura come funzione di Rosenbrock

$$f(x, y) = 100(y - x^2)^2 + (1 - x)^2.$$

Tale funzione ammette un minimo assoluto in $(1, 1)$ di valore 0. Se applichiamo alla funzione di Rosenbrock il metodo del gradiente otteniamo l'andamento riportato in Figura 10, dove nella figura di destra viene riportato ingrandito, il riquadro, evidenziato a sinistra, relativo alle prime dieci iterazioni del metodo, a partire del punto $(-2, 2)$. Tale andamento prosegue per centinaia di iterazioni prima di raggiungere il punto di minimo, con un percorso che segue l'andamento a *banana* evidenziato dalle curve di livello.
Se alla stessa funzione, e a partire dallo stesso punto, applichiamo l'algoritmo di Newton *puro*, con $\alpha = 1$, otteniamo l'andamento riportato in Figura 11, dove il punto di ottimo viene raggiunto alla quinta iterazione. Si evidenzia come alla seconda iterazione lo spostamento non porti ad un punto migliore rispetto a quello di partenza, con una direzione che è praticamente perpendicolare al gradiente ed un passo lunghissimo. Ciò nonostante l'algoritmo converge molto più rapidamente.

6.2.5 Metodi quasi Newton

L'esempio precedente ci invoglia a chederci se non si possano raggiungere le prestazioni del metodo di Newton ad un costo computazionale inferiore all'$O(n^3)$ richiesto ad ogni iterazione.

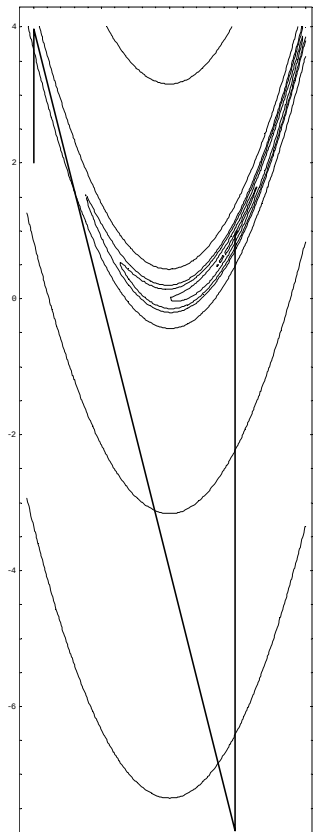


Figura 11: Il metodo di Newton applicato alla funzione di Rosenbrock.

Tale costo vale sia per il metodo puro che per le varie varianti di Newton modificato. In particolare, l'algoritmo del gradiente ha un costo computazionle dell'ordine di $O(n^2)$ ad iterazione, se si accetta di individuare la lunghezza del passo con metodi di ricerca lineare non esatti, come, ad esempio, quello di Armijo. In mancanza di alternative valide, il metodo del gradiente risulterebbe l'unico praticabile in presenza di funzioni di grandissime dimensioni. Per nostra fortuna, *esistono alternative valide*.

I metodi quasi Newton sono stati introdotti per ovviare all'eccessivo carico computazionale dei metodi di Newton puro e modificato, senza perdere le caratteristiche di convergenza dell'algoritmo. Lo scopo che questi metodi si prefiggono è di ridurre il costo computazionale ad ogni iterazione mediante l'approssimazione dell'inversa della matrice hessiana, $H(\boldsymbol{x}_k)^{-1}$ con una matrice definita positiva G_k, tale che la matrice al passo successivo, G_{k+1}, possa essere ricavata da G_k in $O(n^2)$ passi.
L'idea è quella

(a) di ricavare la direzione come risultato dell'applicazione di un operatore lineare al vettore gradiente calcolato nel punto corrente:

$$\boldsymbol{d}_k := -G_k \nabla f(\boldsymbol{x}_k)$$

(b) di calcolare il passo α_k mediante tecniche di ricerca monodimensionale,

(c) di calcolare il nuovo punto nel modo consueto

$$\boldsymbol{x}_{k+1} := \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$$

Per semplicità si inizializza G_0 con la matrice identità. Mentre per l'aggiornamento si opera nel seguente modo.
Come per il metodo di Newton puro, approssimiamo la funzione $f(\boldsymbol{x}_k+\boldsymbol{h}_k)$ con una espressione quadratica $q(\boldsymbol{h}_k)$, dove $\boldsymbol{h}_k = \alpha_k \boldsymbol{d}_k$:

$$f(\boldsymbol{x}_k + \boldsymbol{h}_k) \approx q(\boldsymbol{h}_k) = f(\boldsymbol{x}_k) + \boldsymbol{h}_k^T \nabla f(\boldsymbol{x}_k) + \frac{1}{2}\boldsymbol{h}_k^T H(\boldsymbol{x}_k)\boldsymbol{h}_k$$

da cui possiamo scrivere

$$\nabla q(\boldsymbol{h}_k) = \nabla f(\boldsymbol{x}_k) + H(\boldsymbol{x}_k)\boldsymbol{h}_k \approx \nabla f(\boldsymbol{x}_k + \boldsymbol{h}_k) = \nabla f(\boldsymbol{x}_{k+1}).$$

Se scriviamo $\boldsymbol{h}_k = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k$ e definiamo $\boldsymbol{p}_k := \nabla f(\boldsymbol{x}_{k+1}) - \nabla f(\boldsymbol{x}_k)$ possiamo utilizzare l'approssimazione quadratica per ricavare $\boldsymbol{p}_k \approx H(\boldsymbol{x}_k)\boldsymbol{h}_k$, ottenendo la seguente

Relazione guida

$$H(\boldsymbol{x}_k)^{-1}\boldsymbol{p}_k \approx \boldsymbol{h}_k$$

Quindi, dopo aver inizializzato la matrice G_0 con la matrice identità, imponiamo ad ogni iterazione k, che la matrice G_{k+1} soddisfi la relazione guida come uguaglianza:

$$G_{k+1}\boldsymbol{p}_k = \boldsymbol{h}_k \tag{16}$$

Il modo in cui viene soddisfatta l'equazione (16) differenzia le diverse varianti dei metodi quasi Newton.

Metodo quasi-Newton;
{
    Scegli $\boldsymbol{x}_0 \in \mathbb{R}^n$; $k := 0$;
    $G_0 := I$;
    While $\nabla f(\boldsymbol{x}_k) \neq \emptyset$;
    {
        $\boldsymbol{d}_k := -G_k \nabla f(\boldsymbol{x}_k)$ /* direzione di discesa */
        calcola $\alpha_k \in \mathbb{R}$; /* passo lungo $\boldsymbol{d}_k$ */
        $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha \boldsymbol{d}_k$;
        calcola G_{k+1} a partire da G_k;
        $k := k + 1$;
    }
}

6.2.6 Formula di aggiornamento di rango uno

Con la formula di aggiornamento di *rango uno* si impone

$$G_{k+1}\boldsymbol{p}_k = G_k\boldsymbol{p}_k + c\boldsymbol{u}\boldsymbol{u}^T\boldsymbol{p}_k = \boldsymbol{h}_k$$

cioè $G_{k+1} = G_k + c\boldsymbol{u}\boldsymbol{u}^T$.
Se scegliamo il coefficiente c in modo che valga la relazione $c\boldsymbol{u}^T\boldsymbol{p}_k = 1$, cioè

$$c := 1/(\boldsymbol{u}^T\boldsymbol{p}_k),$$

e imponiamo $\boldsymbol{u} = \boldsymbol{h}_k - G_k\boldsymbol{p}_k$, la formula di aggiornamento di G_{k+1} a partire da G_k diviene

$$G_{k+1} = G_k + \frac{(\boldsymbol{h}_k - G_k\boldsymbol{p}_k)(\boldsymbol{h}_k - G_k\boldsymbol{p}_k)^T}{(\boldsymbol{h}_k - G_k\boldsymbol{p}_k)^T\boldsymbol{p}_k}.$$

Come è facile verificare la matrice G_{k+1} può essere ricavata dalla matrice G_k con un costo computazionale dell'ordine di $O(n^2)$. Abbiamo quindi raggiunto il primo obiettivo: il costo computazionale rimane equivalente a quello del metodo del gradiente, almeno dal punto di vista asintotico. Vi è un secondo risultato positivo che si ottiene valutando le prestazioni di questo metodo quando esso viene applicato, al solito, alle funzioni quadratiche.

> **Proprietà 15** *G_k converge a $H(\boldsymbol{x}_k)^{-1}$ per funzioni quadratiche anche se α_k non viene calcolato in modo ottimo.*

Permangono però dei problemi:

- G_k non rimane definita positiva ad ogni iterazione e quindi non ci garantisce che la direzione scelta sia sempre una direzione di discesa.
- il valore $\boldsymbol{u}^T\boldsymbol{p}_k$ può diventare molto piccolo ($\boldsymbol{u}$ può diventare prossimo al vettore nullo) introducendo problemi di instabilità numerica.

Per tali ragioni è stata introdotta una differente formula di aggiornamento.

6.2.7 Formula di aggiornamento di rango due (DFP)

Con la formula di aggiornamento di *rango due* si impone

$$G_{k+1} = G_k + c\boldsymbol{u}\boldsymbol{u}^T + d\boldsymbol{v}\boldsymbol{v}^T.$$

In questo caso, se si sceglie di imporre le relazioni

$$\boldsymbol{u} = \boldsymbol{h}_k, \quad c\boldsymbol{u}^T\boldsymbol{p}_k = 1, \quad \boldsymbol{v} = G_k\boldsymbol{p}_k \quad e \quad d\boldsymbol{v}^T\boldsymbol{p}_k = -1,$$

si ottiene la formula di aggiornamento seguente:

$$G_{k+1} = G_k + \frac{\boldsymbol{h}_k\boldsymbol{h}_k^T}{\boldsymbol{h}_k^T\boldsymbol{p}_k} - \frac{G_k\boldsymbol{p}_k\boldsymbol{p}_k^TG_k}{\boldsymbol{p}_k^TG_k\boldsymbol{p}_k}.$$

Questa formula di aggiornamento è alla base del cosiddetto *metodo DFP*, dai nomi degli autori Davidon, Fletcher e Powell, che gode di diverse proprietà.

> **Proprietà 16** *Se $f(\boldsymbol{x})$ è una funzione qualsiasi il metodo DFP*
>
> - *se G_0 è definita positiva (es. $G_1 = I$) genera matrici G_k definite positive, quindi ad ogni passo le direzioni $\boldsymbol{d}_k$ sono di discesa;*
> - *ha un costo computazionale dell'ordine $O(n^2)$ per iterazione;*

- *ha rapidità di convergenza superlineare.*

Il costo computazionale rimane dell'ordine di quello del metodo del gradiente, inoltre le direzioni sono sempre direzioni di discesa e, importantissimo, ha una rapidità di convergenza notevolmente migliore di quella del metodo del gradiente, anche se non tale da uguagliare il metodo di Newton (quando quest'ultimo converge).

Proprietà 17 *Se $f(\boldsymbol{x})$ è una funzione quadratica il metodo DFP con ricerca esatta del passo α_k*

- *termina in n iterazioni con $G_{n+1} = H(\boldsymbol{x}_n)^{-1}$;*
- *genera direzioni che soddisfano la seguente condizione*[3]*: $\boldsymbol{d}_i^T H \boldsymbol{d}_j = 0$, per ogni $i \neq j$;*
- *quando $G_0 = I$, genera gradienti coniugati (vedi nota a piè di pagina);*
- *$G_{k+1}\boldsymbol{p}_i = \boldsymbol{h}_i$ con $i = 1, \ldots, k$ (la relazione (16) è soddisfatta in maniera retroattiva).*

La precedente proprietà significa, in pratica, che quando questo metodo viene applicato alle funzioni quadratiche, esso converge in un numero di iterazioni al più pari alla dimensione dello spazio di ricerca, cioè in n passi.

Proprietà 18 *Se $f(\boldsymbol{x})$ è una funzione convessa il metodo DFP con ricerca esatta del passo α_k converge.*

Questo significa che la convergenza è garantita anche quando la matrice G_0 non è inizializzata alla matrice identità.

Ma si può fare di meglio.

6.2.8 Formula di aggiornamento di rango due inversa (BFGS)

In questa versione dei metodi *quasi Newton* si cerca una matrice B_k che approssimi direttamente la matrice Hessiana $H(\boldsymbol{x}_k)$ e non la sua inversa. Partiamo cioè dall'espressione $H(\boldsymbol{x}_k)\boldsymbol{h}_k \approx \boldsymbol{p}_k$ e imponiamo alla matrice B_k di soddisfarla come uguaglianza

$$B_{k+1}\boldsymbol{h}_k = \boldsymbol{p}_k \tag{17}$$

Cercando direttamente una formula di aggiornamento di *rango due* si impone

$$B_{k+1} = B_k + c\boldsymbol{u}\boldsymbol{u}^T + d\boldsymbol{v}\boldsymbol{v}^T$$

e con passaggi analoghi a quelli visti nella sezione (6.2.7) si perviene all'espressione

$$B_{k+1} = B_k + \frac{\boldsymbol{p}_k\boldsymbol{p}_k^T}{\boldsymbol{p}_k^T\boldsymbol{h}_k} - \frac{B_k\boldsymbol{h}_k\boldsymbol{h}_k^T B_k}{\boldsymbol{h}_k^T B_k \boldsymbol{h}_k}$$

dove vengono sostanzialmente invertiti i ruoli di $\boldsymbol{h}$ e $\boldsymbol{p}$.
Naturalmente il metodo di Newton richiede di approssimare l'inversa di $H(\boldsymbol{x}_k)$, e dalla precedente espressione si può ricavare sia la G_k, che la relativa formula di aggiornamento

$$G_{k+1} = G_k + \left(1 + \frac{\boldsymbol{p}_k G_k \boldsymbol{p}_k^T}{\boldsymbol{h}_k^T\boldsymbol{p}_k}\right)\frac{\boldsymbol{h}_k\boldsymbol{h}_k^T}{\boldsymbol{h}_k^T\boldsymbol{p}_k} - \left(\frac{\boldsymbol{h}_k\boldsymbol{p}_k^T G_k + G_k\boldsymbol{p}_k\boldsymbol{h}_k^T}{\boldsymbol{h}_k^T\boldsymbol{p}_k}\right).$$

[3] Direzioni che soddisfano tali condizioni sono dette *direzioni coniugate* e vengono descritte nella sezione 6.3

Questa formula è nota come BFGS, dai nomi degli autori Broyden, Fletcher, Goldfarb e Shanno.
Il metodo che ne deriva è più robusto del metodo DFP rispetto agli errori di arrotondamento ed in particolare vale la seguente

> **Proprietà 19** *Data una funzione $f(\boldsymbol{x})$ il metodo BFGS converge globalmente purché la ricerca monodimensionale di α_k soddisfi la condizione di sufficiente riduzione della $f(\boldsymbol{x})$, relazione (2), e la condizione di Wolfe, relazione (3) (cfr. sezione 6.1).*

6.2.9 Famiglia di Broyden

Le formule di aggiornamento appena introdotte si possono vedere come particolarizzazioni di una famiglia di formule dette di *Broyden*

$$G_{k+1} = (1-\phi)G_{k+1}^{DPF} + \phi G_{k+1}^{BFGS},$$

dove per le migliori proprietà e la stabilità del metodo si adotta $0 \leq \phi \leq 1$.

6.3 Metodi alle direzioni coniugate

Questi metodi sono stati sviluppati per risolvere efficacemente i modelli quadratici del tipo

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T Q\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x}$$

Tali metodi convergono esattamente in un numero finito di iterazioni quando vengono applicati a funzioni quadratiche definite positive. Metodi che hanno tale proprietà di terminazione quadratica sono molto ricercati in quanto ci si aspetta che abbiano buone prestazioni anche per funzioni non quadratiche purché applicati nell'intorno di un minimo locale. Questo perché, dalle proprietà dello sviluppo in serie di Taylor, funzioni molto generali ben approssimano la forma quadratica nell'intorno di un minimo.

Due vettori $\boldsymbol{u}$ e $\boldsymbol{v}$ sono detti ortogonali se il loro prodotto scalare è nullo, cioè $\boldsymbol{u}^T\boldsymbol{v} = 0$. Similmente, due vettori sono detti *coniugati* rispetto ad una matrice simmetrica definita positiva Q se vale la relazione

$$\boldsymbol{u}^T Q\boldsymbol{v} = 0.$$

Il metodo si basa essenzialmente sulla seguente proprietà.

> **Teorema 14** *Siano $\boldsymbol{u}_i$, con $i = 1, \ldots, n$, un insieme di vettori di $\mathbb{R}^n$ mutuamente coniugati rispetto ad una matrice simmetrica definita positiva Q. Tali vettori formano una base per $\mathbb{R}^n$, cioè, per ogni vettore $\boldsymbol{x} \in \mathbb{R}^n$ vale*
>
> $$\boldsymbol{x} = \sum_{i=1}^{n} \frac{\boldsymbol{u}_i(Q\boldsymbol{x})\boldsymbol{u}_i}{\boldsymbol{u}_i Q\boldsymbol{u}_i} = \sum_{i=1}^{n} \alpha_i \boldsymbol{u}_i, \text{ dove } \alpha_i = \frac{\boldsymbol{u}_i Q\boldsymbol{x}}{\boldsymbol{u}_i Q\boldsymbol{u}_i}$$

Useremo la precedente proprietà all'interno di un metodo iterativo

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$$

che, a partire dal vettore iniziale $\boldsymbol{x}_0$, ci permetterà di determinare il minimo $\boldsymbol{x}^* = Q^{-1}\boldsymbol{b}$ della funzione quadratica $f(\boldsymbol{x})$ in un numero di passi minore o uguale ad n, senza la necessità di

invertire la matrice Q.
Il metodo usa come direzioni di discesa un insieme di n vettori $\boldsymbol{d}_i$, con $i = 0, \ldots, n-1$, mutuamente coniugati rispetto a Q (che per ora assumeremo essere disponibili e successivamente vedremo come ricavare): $\boldsymbol{d}_i^T Q \boldsymbol{d}_j = 0$, per $i \neq j$.

Se osserviamo i passaggi del metodo iterativo possiamo ricavare che

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k = \boldsymbol{x}_{k-1} + \alpha_{k-1}\boldsymbol{d}_{k-1} + \alpha_k \boldsymbol{d}_k = \ldots = \boldsymbol{x}_0 + \sum_{i=0}^{k} \alpha_i \boldsymbol{d}_i$$

Poiché i vettori $\boldsymbol{d}_j$ formano una base per $\mathbb{R}^n$, noi sappiamo che ogni vettore $\boldsymbol{x}$ si può esprimere come $\boldsymbol{x}_0 + \sum_{i=0}^{n-1} \alpha_i \boldsymbol{d}_i$. Ora, avendo a disposizione le direzioni coniugate $\boldsymbol{d}_i$, desideriamo scoprire quali valori devono assumere i moltiplicatori α_i affinché $\boldsymbol{x}_0 + \sum_{i=0}^{n-1} \alpha_i \boldsymbol{d}_i = \boldsymbol{x}^* = Q^{-1}b$.

Poiché il punto di partenza $\boldsymbol{x}_0$ è dato e le $\boldsymbol{d}_i$ sono assegnate, è possibile trasformare la funzione $f(\boldsymbol{x})$ in una funzione del vettore dei moltiplicatori incogniti α_i:

$$f(\boldsymbol{x}) = f(\boldsymbol{x}_0 + \sum_{i=0}^{n-1} \alpha_i \boldsymbol{d}_i) = F(\boldsymbol{\alpha}).$$

Abbiamo quindi

$$\min F(\boldsymbol{\alpha}) = \min \frac{1}{2} \left(\boldsymbol{x}_0 + \sum\nolimits_{i=0}^{n-1} \alpha_i \boldsymbol{d}_i\right)^T Q \left(\boldsymbol{x}_0 + \sum\nolimits_{i=0}^{n-1} \alpha_i \boldsymbol{d}_i\right) - \boldsymbol{b}^T \left(\boldsymbol{x}_0 + \sum\nolimits_{i=0}^{n-1} \alpha_i \boldsymbol{d}_i\right) =$$

$$= \sum\nolimits_{i=0}^{n-1} \min \left(\frac{1}{2} \left(\boldsymbol{x}_0 + \alpha_i \boldsymbol{d}_i\right)^T Q \left(\boldsymbol{x}_0 + \alpha_i \boldsymbol{d}_i\right) - \boldsymbol{b}^T \left(\boldsymbol{x}_0 + \alpha_i \boldsymbol{d}_i\right)\right) = \sum\nolimits_{i=0}^{n-1} \min F_i\left(\alpha_i\right)$$

dove si è sfruttato il fatto che $\boldsymbol{d}_i^T Q \boldsymbol{d}_j = 0$, per $i \neq j$. In questo modo, il problema originale è stato trasformato nella somma di n problemi monodimensionali indipendenti, nelle incognite $\alpha_0, \ldots, \alpha_{n-1}$:

$$\min \left(\frac{1}{2}\alpha_i^2 \boldsymbol{d}_i^T Q \boldsymbol{d}_i + \alpha_i (Q\boldsymbol{x}_0)^T \boldsymbol{d}_i - \alpha_i \boldsymbol{b}^T \boldsymbol{d}_i\right),$$

dove in questo ultimo passaggio si è sfruttato il fatto che, poiché Q è simmetrica, $\boldsymbol{x}_i^T Q \boldsymbol{d}_i = \boldsymbol{d}_i^T Q \boldsymbol{x}_i$, $\boldsymbol{x}^T Q = (Q\boldsymbol{x})^T$, e si sono trascurati i termini indipendenti da α_i.
Risolvendo gli n problemi indipendenti, troveremo gli α_i^* ottimi e da questi ricaveremo

$$\boldsymbol{x}^* := \boldsymbol{x}_0 + \sum\nolimits_{i=0}^{n-1} \alpha_{\mathrm{i}}{}^* \boldsymbol{d}_i$$

Se deriviamo $F_i(\alpha_i)$ in funzione di α_i e eguagliamo a zero, otteniamo:

$$\alpha_i \boldsymbol{d}_i^T Q \boldsymbol{d}_i + (Q\boldsymbol{x}_0 - \boldsymbol{b})^T \boldsymbol{d}_i = 0$$

e quindi

$$\alpha_i{}^* = -\frac{(Q\boldsymbol{x}_0 - \boldsymbol{b})^T \boldsymbol{d}_i}{\boldsymbol{d}_i^T Q \boldsymbol{d}_i} = -\frac{\nabla f(\boldsymbol{x}_0)^T \boldsymbol{d}_i}{\boldsymbol{d}_i^T Q \boldsymbol{d}_i}$$

Riprendiamo ora il metodo iterativo

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k = \boldsymbol{x}_0 + \sum_{i=0}^{k} \alpha_i \boldsymbol{d}_i$$

e verifichiamo che, se utilizziamo i moltiplicatori ottimi α_i^* appena determinati, esso converge al passo n-esimo alla soluzione ottima.
Sostituendo a α_i i moltiplicatori ottimi α_i^* ricaviamo

$$\boldsymbol{x}_n = \boldsymbol{x}_0 - \sum_{i=0}^{n-1} \frac{(Q\boldsymbol{x}_0 - \boldsymbol{b})^T \boldsymbol{d}_i}{\boldsymbol{d}_i^T Q \boldsymbol{d}_i} \boldsymbol{d}_i =$$

$$\boldsymbol{x}_n = \boldsymbol{x}_0 - \sum_{i=0}^{n-1} \frac{\boldsymbol{d}_i^T (Q\boldsymbol{x}_0) \boldsymbol{d}_i}{\boldsymbol{d}_i^T Q \boldsymbol{d}_i} + \sum_{i=0}^{n-1} \frac{\boldsymbol{d}_i^T (Q(Q^{-1}\boldsymbol{b})) \boldsymbol{d}_i}{\boldsymbol{d}_i^T Q \boldsymbol{d}_i}$$

$$\boldsymbol{x}_n = \boldsymbol{x}_0 - \boldsymbol{x}_0 + Q^{-1}\boldsymbol{b} = Q^{-1}\boldsymbol{b} = \boldsymbol{x}^*$$

dove, nell'ultimo passaggio, abbiamo sfruttato il Teorema 14.

L'applicabilità del metodo dipende dal fatto che siano note n direzioni mutuamente coniugate. Il modo più diretto è calcolare gli autovettori di Q che godono la proprietà di essere sia mutuamente ortogonali che mutuamente coniugati rispetto a Q. Tuttavia la loro determinazione è altrettanto costosa quanto la soluzione del problema quadratico. Un modo alternativo è quello di modificare il processo di ortogonalizzazione di Gram-Schmidt in modo da generare direzioni coniugate invece che direzioni ortogonali. Questo approccio tuttavia richiede di memorizzare l'intero insieme di direzioni.

6.3.1 Il metodo di gradiente coniugato

Il metodo di *gradiente coniugato* è un metodo alle direzioni coniugate con la speciale proprietà di generare la direzione corrente $\boldsymbol{d}_k$ sulla base della sola conoscenza (e memorizzazione) del vettore $\boldsymbol{d}_{k-1}$.
Nel richiamare brevemente il metodo ci svincoliamo anche dal suo utilizzo nell'ambito dei problemi di programmazione quadratica, e consideriamo la minimizzazione di funzioni $f(\boldsymbol{x})$ qualsiasi.
L'approccio proposto da Fletcher-Reeves (1964) propone le seguenti formule per ottenere direzioni coniugate:

$$\boldsymbol{d}_1 = -\nabla f(\boldsymbol{x}_0)$$

$$\boldsymbol{d}_{k+1} = -\nabla f(\boldsymbol{x}_k) + \beta_k \boldsymbol{d}_k \text{ e } \beta_k = \frac{\|\nabla f(\boldsymbol{x}_k)\|^2}{\|\nabla f(\boldsymbol{x}_{k-1})\|^2} \text{ per } k = 1, \ldots, n-1$$

Inoltre ad ogni passo è necessario eseguire una ricerca monodimensionale per determinare un minimo, o una sua approssimazione, per il passo α. Tuttavia, affinché le direzioni generate siano sempre di discesa è necessario che il passo α sia scelto ad ogni iterazione in modo da soddisfare le condizioni di Wolfe in senso forte (vedi formula (4)).
In alternativa, per β_k, si possono usare le formule di Polak-Ribiére (1969):

$$\beta_k = \frac{(\nabla f(\boldsymbol{x}_k) - \nabla f(\boldsymbol{x}_{k-1}))^T \nabla f(\boldsymbol{x}_k)}{\|\nabla f(\boldsymbol{x}_{k-1})\|^2}$$

che hanno la proprietà di ritornare alla direzione del gradiente ogniqualvolta la scelta della direzione coniugata non sia la scelta migliore, cioè quando $\nabla f(\boldsymbol{x}_k) - \nabla f(\boldsymbol{x}_{k-1}) \approx 0$.
Si può dimostrare che per funzioni quadratiche i due metodi sono equivalenti e determinano la soluzione ottima se si adottano gli α_i^* ottimi.
Poichè non richiede operazioni matriciali questo approccio è adatto per problemi di grandi dimensioni.

7 Metodi di Trust-Region

Sia i metodi basati su ricerca lineare che i metodi di tipo *Trust-Region* determinano lo spostamento ad ogni iterazione con l'aiuto di un modello quadratico della funzione obiettivo. Essi usano però questo modello in modo diverso. Come abbiamo visto fino ad ora, i metodi basati su ricerca lineare, usano il modello quadratico per generare una direzione $\boldsymbol{d}_k$ e poi si concentrano sulla scelta del passo α_k migliore. I metodi Trust-Region, invece, definiscono una regione (la Trust-Region, o Regione di Confidenza, appunto), di solito una (iper)sfera, intorno alla soluzione corrente, $\boldsymbol{x}_k$, nell'ipotesi che in tale regione il modello quadratico sia una rappresentazione adeguata della funzione obiettivo. Essi determinano simultaneamente la direzione e il passo che minimizzano il modello quadratico. Se la scelta del passo che deriva da questo approccio non soddifa opportuni criteri, si rimane nel punto corrente, si riduce l'ampiezza della regione, in pratica il raggio dell'(iper)sfera, e si ripete il processo. Altrimenti, il nuovo punto corrente diventa la soluzione ottima del problema quadratico, e, se è il caso, si incrementa l'ampiezza della regione.
In generale, direzione e passo cambiano ogni volta che viene modificata l'ampiezza della Trust-Region. È quindi questa ampiezza, cioè il raggio dell'(iper)sfera, il punto chiave di questo metodo. Se la regione è troppo piccola, l'algoritmo perde l'opportunità di fare un passo lungo verso il minimo locale. Viceversa, se essa è troppo larga il modello quadratico potrebbe essere molto lontano dal rappresentare la funzione obiettivo nella Trust-Region, così da rendere necessaria una sua riduzione, perdendo una (o più) iterazione(i) senza, di fatto, muoversi. Dal punto di vista pratico si sceglie l'ampiezza dalla Trust-Region in base alle prestazioni che ha avuto l'algoritmo durante le iterazioni precedenti.

Nel seguito denoteremo con m_k la funzione modello adottata. Inoltre assumenremo che m_k sia un modello quadratico basato sullo sviluppo in serire di Taylor di $f(\boldsymbol{x})$ attorno al punto $\boldsymbol{x}_k$,

$$f(\boldsymbol{x}_k + \boldsymbol{p}) \approx f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)^T \boldsymbol{p} + \frac{1}{2}\boldsymbol{p}^T H(\boldsymbol{x}_k)\boldsymbol{p}.$$

Se si denota con B_k una approssimazione della matrice hessiana mediante una qualche matrice simmetrica, il modello m_k viene definito come

$$m_k(\boldsymbol{p}) = f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)^T \boldsymbol{p} + \frac{1}{2}\boldsymbol{p}^T B_k \boldsymbol{p},$$

La differenza fra $f(\boldsymbol{x}_k + \boldsymbol{p})$ e $m_k(\boldsymbol{p})$ è dell'ordine di $O(||\boldsymbol{p}||^2)$ che si riduce a $O(||\boldsymbol{p}||^3)$ quando B_k coincide con la matrice hessiana. In quest'ultimo caso il metodo prende il nome di Trust-Region Newton.

Ad ogni passo cerchiamo una soluzione al sottoproblema

$$\min_{\boldsymbol{p}\in\mathbb{R}^n} m_k(\boldsymbol{p}) = f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)^T \boldsymbol{p} + \frac{1}{2}\boldsymbol{p}^T B_k \boldsymbol{p}, \text{ t.c. } ||\boldsymbol{p}|| \leq \Delta_k. \tag{18}$$

dove si denota con Δ_k l'ampiezza della Trust Region. Ci siamo spostati da un problema di minimo non vincolato ad un problema di ottimizzazione vincolata, dove sia la funzione obiettivo che i vincoli sono quadratici.
Quando la matrice B_k è definita positiva e vale la relazione $||B_k^{-1}\nabla f(\boldsymbol{x}_k)|| \leq \Delta_k$, la soluzione del problema (18) è semplicemente il minimo non vincolato $\boldsymbol{p}^B = -B_k^{-1}\nabla f(\boldsymbol{x}_k)$. In questo

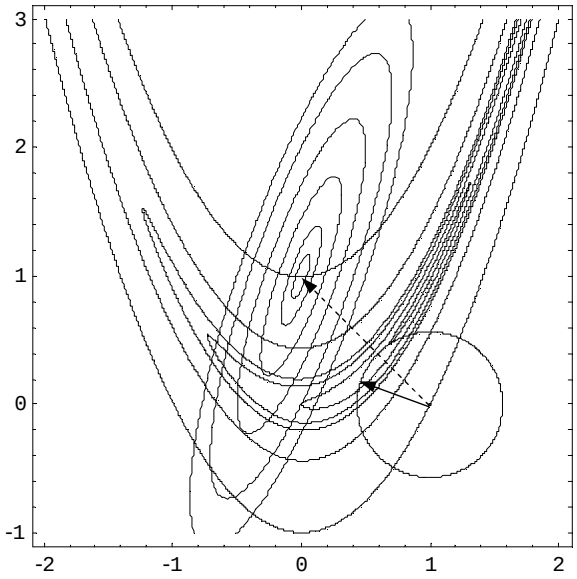


Figura 12: Esempio di modello quadratico per la funzione di Rosenbrock.

caso chiamiamo $\boldsymbol{p}^B$ *full step*. In tutti gli altri casi occorre operare sul raggio Δ_k.
Lo schema di un generico metodo Trust-Region è quindi definito da

- stabilire Δ_k
- risolvere il problema di minimo vincolato

In Figura 12 vediamo le curve di livello della funzione di Rosenbrock ed il relativo modello quadratico calcolato nel punto $(1, 0)$, $m_k(\boldsymbol{p}) = 601p_1^2 + 100p_2^2 - 400p_1p_2 + 400p_1 - 200p_2 + 100$. L'ottimo del modello quadratico non vincolato sta nel punto $(0, 1)$ indicato dalla freccia tratteggiata, mentre l'ottimo del modello con $\Delta = 0.4$ è indicato dalla freccia piena.

Come si è detto la scelta del raggio Δ_k della Trust Region è la chiave per un buon funzionamento del metodo. Tale scelta si basa su considerazioni attorno al rapporto fra la *riduzione attuale* del valore della funzione obiettivo determinata dal passo e la *riduzione predetta* dal modello

$$\rho_k = \frac{f(\boldsymbol{x}_k) - f(\boldsymbol{x}_{k+1})}{m_k(\boldsymbol{0}) - m_k(\boldsymbol{p}_k)}. \tag{19}$$

A seconda del segno di ρ_k possiamo avere vari casi. Se $\rho_k \leq 0$ il passo è peggiorante (o non migliorante) e viene rifiutato. Se è all'incirca pari a 1, m_k è una buona approssimazione della funzione obiettivo e quindi si può pensare di espandere la regione. Se ρ_k è positivo ma molto più piccolo di 1 la regione non viene modificata, mentre se il valore è prossimo a zero o negativo la regione viene ridotta, riducendo Δ_k.

Metodo Trust Region;
{
`Scegli` $\boldsymbol{x}_0 \in \mathbb{R}^n$; $k := 0$; $\overline{\Delta} > 0$, $\Delta_0 \in (0, \overline{\Delta})$; $\eta \in [0, \frac{1}{4}]$;
While $\nabla f(\boldsymbol{x}_k) \neq \emptyset$;
{
$\boldsymbol{p}_k := \texttt{argmin}\{m_k(\boldsymbol{p}),\ \texttt{t.c.}\quad ||\boldsymbol{p}|| \leq \Delta_k\}$;
$\rho_k := \frac{f(\boldsymbol{x}_k) - f(\boldsymbol{x}_k + \boldsymbol{p}_k)}{m_k(\boldsymbol{0}) - m_k(\boldsymbol{p}_k)}$;
if $\rho_k < \frac{1}{4}$ **then**
$\Delta_{k+1} := \frac{1}{4}\Delta_k$
else
if $\rho_k > \frac{3}{4}$ `e` $||\boldsymbol{p}_k|| = \Delta_k$ **then**
$\Delta_{k+1} := \min\{2\Delta_k, \overline{\Delta}\}$
else
$\Delta_{k+1} := \Delta_k$
if $\rho_k > \eta$ **then**
$\boldsymbol{x}_{k+1} := \boldsymbol{x}_k + \boldsymbol{p}_k$
else
$\boldsymbol{x}_{k+1} := \boldsymbol{x}_k$
$k := k + 1$;
}
}

In questo algoritmo $\overline{\Delta}$ è un limite per eccesso sulla lunghezza dei passi e la regione viene allargata solo quando, nella soluzione ottima del problema (18), $||\boldsymbol{p}_k||$ raggiunge il bordo della regione.

Affinché il metodo Trust-Region possa funzionare in pratica è necessario risolvere efficientemente il sottoproblema (18), ed è su questo punto che concentriamo ora la nostra attenzione. Prima di procedere alla descrizione degli algoritmi vogliamo però caratterizzare in modo analitico le soluzioni del problema (18) mediante il seguente teorema, dove per semplicità adottiamo la notazione $\boldsymbol{g} = \nabla f(\boldsymbol{x})$ e sottintendiamo il pedice k, relativo all'iterazione corrente.

Teorema 15 *Il vettore $\boldsymbol{p}^*$ è una soluzione globale ottima del problema di Trust-Region*

$$\min_{\boldsymbol{p} \in \mathbb{R}^n} m(\boldsymbol{p}) = f(\boldsymbol{x}) + \boldsymbol{g}^T\boldsymbol{p} + \frac{1}{2}\boldsymbol{p}^T B\boldsymbol{p},\ \textit{t.c.}\ ||\boldsymbol{p}|| \leq \Delta. \tag{20}$$

se e solo se $\boldsymbol{p}^*$ è ammissibile ed esiste uno scalare $\lambda \geq 0$ tale che sono soddisfatte le seguenti condizioni

$$(B + \lambda I)\boldsymbol{p}^* = -\boldsymbol{g}, \tag{21}$$

$$\lambda(\Delta - ||\boldsymbol{p}^*||) = 0, \tag{22}$$

$$(B + \lambda I) \quad \textit{è semidefinita positiva.} \tag{23}$$

Si osservi che la condizione (22) è una condizione di scarto complementare, ed implica che almeno una delle due quantità non negative, λ e $(\Delta - ||\boldsymbol{p}^*||)$ deve essere zero. E quindi quando la soluzione giace strettamente dentro la regione di confidenza deve essere $\lambda = 0$ e

quindi $\boldsymbol{p}^* = -B^{-1}\boldsymbol{g}$, dalla (21), con B definita positiva dalla (23). Altrimenti si ha $||\boldsymbol{p}^*|| = \Delta$ e λ può assumere un valore positivo. In questo caso, dalla (21) si ha

$$\lambda \boldsymbol{p}^* = -B\boldsymbol{p}^* - \boldsymbol{g} = -\nabla m(\boldsymbol{p}^*)$$

cioè la soluzione $\boldsymbol{p}^*$ è collineare con l'antigradiente del modello quadratico.

7.1 Il punto di Cauchy

Come abbiamo visto, in particolare nel Teorema 8, la convergenza globale degli algoritmi basati su ricerca lineare viene garantita dal rispetto di condizioni non molto stringenti, sia sulla direzione $\boldsymbol{d}$, sia sul passo α. Qualcosa di simile accade con gli algoritmi di tipo Trust-Region. Infatti, per la convergenza globale di tali metodi è sufficiente trovare, ad ogni iterazione k, una soluzione approssimata, $\boldsymbol{p}_k$, del problema (18) che appartenga alla Trust-Region e garantisca una *sufficiente riduzione* del modello quadratico.
Tale riduzione viene quantificata in relazione al cosiddetto *punto di Cauchy*, che nel seguito verrà denotato come $\boldsymbol{p}_k^C$. Tale punto non e' altro che il punto di minimo del modello quadratico lungo la direzione dell'antigradiente.
Sia

$$\boldsymbol{p}_k^S = -\frac{\Delta_k}{||\nabla f(\boldsymbol{x}_k)||}\nabla f(\boldsymbol{x}_k)$$

la direzione dell'antigradiente *normalizzata* in ragione dell'attuale ampiezza della Trust-Region. Cerchiamo ora il minimo τ_k del problema monodimensionale

$$\min_{\tau \geq 0} \mu_k(\tau) = \min_{\tau \geq 0} m_k(\tau \boldsymbol{p}_k^S) \text{ t.c. } ||\tau \boldsymbol{p}_k^S|| \leq \Delta_k$$

e definiamo $\boldsymbol{p}_k^C = \tau_k \boldsymbol{p}_k^S$.
Cominciamo col riscrivere esplicitamente il problema monodimensionale, dopo aver adottato la notazione $\boldsymbol{g}_k$ al posto di $\nabla f(\boldsymbol{x}_k)$,

$$\mu_k(\tau) = f(\boldsymbol{x}_k) - \tau \frac{\Delta_k}{||\boldsymbol{g}_k||}\boldsymbol{g}_k^T\boldsymbol{g}_k + \frac{1}{2}\tau^2 \frac{\Delta_k^2}{||\boldsymbol{g}_k||^2}\boldsymbol{g}_k^T B_k \boldsymbol{g}_k.$$

Quando $\boldsymbol{g}_k^T B_k \boldsymbol{g}_k \leq 0$ e $\boldsymbol{g}_k \neq \boldsymbol{0}$ la funzione $\mu_k(\tau)$ decresce monotonicamente al crescere di τ e quindi τ_k è il più grande valore che soddisfa i limiti della Trust-Region, cioè $\tau_k = 1$.
Quando $\boldsymbol{g}_k^T B_k \boldsymbol{g}_k > 0$, $\mu_k(\tau)$ definisce un problema quadratico convesso, così il valore τ_k è il minimo fra il valore limite 1 e l'ottimo di tale problema. Tale ottimo si ricava derivando $\mu_k(\tau)$ rispetto a τ e imponendo l'uguaglianza a zero

$$\mu_k'(\tau) = -\frac{\Delta_k}{||\boldsymbol{g}_k||}\boldsymbol{g}_k^T\boldsymbol{g}_k + \tau \frac{\Delta_k^2}{||\boldsymbol{g}_k||^2}\boldsymbol{g}_k^T B_k \boldsymbol{g}_k = 0.$$

Quindi

$$\tau^* = \frac{||\boldsymbol{g}_k||^3}{\Delta_k \boldsymbol{g}_k^T B_k \boldsymbol{g}_k}.$$

Riassumendo abbiamo

$$\boldsymbol{p}_k^C = -\tau_k \frac{\Delta_k}{||\boldsymbol{g}_k||}\boldsymbol{g}_k$$

dove

$$\tau_k = \begin{cases} 1 & \text{se } \boldsymbol{g}_k^T B_k \boldsymbol{g}_k \leq 0 \\ \min\{\tau^*, 1\} & \text{altrimenti} \end{cases}$$

Ricavare il punto $\boldsymbol{p}_k^C$ ha un costo computazionale dell'ordine $O(n^2)$ e, come detto, la convergenza dei metodi di Trust-Region è garantita quando ogni passo $\boldsymbol{p}_k$ determina una riduzione del modello quadratico che è almeno qualche fissato multiplo positivo della riduzione fornita dal punto di Cauchy.
Dal punto di vista pratico, l'uso diretto dei punti di Cauchy nella soluzione del problema quadratico (18) comporta gli stessi inconvenienti dell'uso del metodo del gradiente all'interno dei metodi basati su ricerca lineare: convergenza lineare e prestazioni scadenti. È quindi necessario fare in modo che la matrice B_k rivesta un qualche ruolo nella determinazione della direzione oltre che del passo. Delle differenti varianti presenti in letteratura ci limitiamo ad introdurre un metodo che si può adottare quando la matrice B_k è definita positiva (ad esempio nei casi affrontati nella Sezione 8).

7.2 Il metodo *Dogleg*

Per semplicità anche nel seguito sottointenderemo il pedice k, relativo all'iterazione corrente, adotteremo ancora una volta la convenzione $\boldsymbol{g} = \nabla f(\boldsymbol{x})$, e faremo quindi riferimento al problema

$$\min_{\boldsymbol{p} \in \mathbb{R}^n} m(\boldsymbol{p}) = f(\boldsymbol{x}) + \boldsymbol{g}^T \boldsymbol{p} + \frac{1}{2} \boldsymbol{p}^T B \boldsymbol{p}, \text{ t.c. } ||\boldsymbol{p}|| \leq \Delta. \tag{24}$$

Il metodo *Dogleg*, *a zampa di cane*, viene impiegato quando la matrice B è definita positiva. In tal caso la soluzione del problema quadratico non vincolato è $\boldsymbol{p}^B = -B^{-1}\boldsymbol{g}$, e tale soluzione viene scelta, cioè $\boldsymbol{p}^* = \boldsymbol{p}^B$, ogniqualvolta $||\boldsymbol{p}^B|| \leq \Delta$, e quindi più frequentemente quando i valori di Δ sono grandi.
Quando la dimensione della regione Δ è piccola rispetto alla norma di $\boldsymbol{p}^B$, il vincolo $||\boldsymbol{p}|| \leq \Delta$ garantisce che il termine quadratico possa essere trascurato nella soluzione di (20) e che si possa utilizzare la seguente approssimazione

$$\boldsymbol{p}^* \approx -\Delta \frac{\boldsymbol{g}}{||\boldsymbol{g}||}.$$

Per valori di Δ intermedi si deve procedere con scelte più sofisticate.
Il metodo *Dogleg* cerca una soluzione approssimata di $\boldsymbol{p}^*$ come una concatenazione di due segmenti. Il primo segmento va dal punto corrente al punto che minimizza m lungo la direzione dell'antigradiente (vedi formula (11)), cioè

$$\boldsymbol{p}^U = -\frac{\boldsymbol{g}^T \boldsymbol{g}}{\boldsymbol{g}^T B \boldsymbol{g}} \boldsymbol{g}$$

mentre il secondo segmento unisce $\boldsymbol{p}^U$ con $\boldsymbol{p}^B$. La traiettoria che ne deriva può venir descritta formalmente come $\tilde{\boldsymbol{p}}(\tau)$ per $\tau \in [0, 2]$, dove

$$\tilde{\boldsymbol{p}}(\tau) = \begin{cases} \tau \boldsymbol{p}^U, & \tau \in [0, 1] \\ \boldsymbol{p}^U + (\tau - 1)(\boldsymbol{p}^B - \boldsymbol{p}^U), & \tau \in [1, 2] \end{cases}$$

Il metodo *Dogleg* sceglie il valore di $\boldsymbol{p}$ che minimizza il modello quadratico m lungo il cammino $\tilde{\boldsymbol{p}}(\tau)$, vincolato alla dimensione della regione. La soluzione di tale problema risulta semplice grazie alla seguente

Proprietà 20 *Se la matrice B è definita positiva allora*

- $\|\tilde{\boldsymbol{p}}(\tau)\|$ *è monotona crescente*
- $m(\tilde{\boldsymbol{p}}(\tau))$ *è monotona decrescente*

Tale proprietà comporta che il cammino $\tilde{\boldsymbol{p}}(\tau)$ intersechi il confine della regione in un solo punto nei casi in cui $\|\boldsymbol{p}^B\| \geq \Delta$, e non lo intersechi affatto altrimenti. Nel caso in cui $\|\boldsymbol{p}^B\| \geq \Delta$ si può calcolare il valore esatto di τ risolvendo l'equazione quadratica scalare

$$||\boldsymbol{p}^U + (\tau - 1)(\boldsymbol{p}^B - \boldsymbol{p}^U)|| = \Delta$$

Dal punto di vista della convergenza, la proprietà 20 garantisce che il metodo *Dogleg* comporti ad ogni passo una riduzione che è almeno pari a quella garantita dal punto di Cauchy.

8 Problemi ai Minimi Quadrati

All'inizio della Sezione 3 abbiamo introdotto come esempio, il problema di determinare i valori ottimi di un insieme di parametri di una funzione in modo da minimizzare la somma dei quadrati degli scarti fra i valori assunti dalla funzione ed i valori di un insieme di dati campionati. Riprendiamo l'esempio semplificandolo un po'.

Esempio Supponiamo di cercare una curva che si adatti ad un insieme di dati sperimentali, $y_1, y_2, \ldots, y_4$, campionati agli istanti $t_1, t_2, \ldots, t_4$. Il nostro modello ha la forma della funzione

$$\phi(t_i, \boldsymbol{x}) = e^{t_i x_1} + \cos(t_i x_2)$$

Il modello non lineare ha due parametri, x_1, x_2. Desideriamo sceglierli in modo da far aderire il più possibile i valori $\phi(t_i, \boldsymbol{x})$ ai dati y_i. Nel seguito indichiamo con

$$r_i(\boldsymbol{x}) = \phi(t_i, \boldsymbol{x}) - y_i, \quad i = 1, \ldots, m,$$

i cosiddetti residui, la somma dei cui quadrati intendiamo minimizzare.

Il problema è noto come *problema ai minimi quadrati* non lineare (least-squares problem) e la funzione obiettivo da minimizzare assume la forma

$$\min_{x \in \mathbb{R}^n} f(\boldsymbol{x}) = \frac{1}{2} \sum_{i=1}^{m} r_i^2(\boldsymbol{x})$$

dove ogni residuo r_i è una funzione da $\mathbb{R}^n$ in $\mathbb{R}$. Nel seguito si assumerà che $m \geq n$ anche se di solito la relazione è $m \gg n$. Si osservi che il calcolo della sola funzione obiettivo, per un dato valore delle variabili $\boldsymbol{x}$, può essere computazionalmente oneroso, anche se le variabili sono poche, dipendendo dal numero m di osservazioni, che di solito è molto elevato.
Per vedere come la forma speciale della funzione obiettivo$f(\boldsymbol{x})$ renda il problema di ottimizzazione ai minimi quadrati più semplice della maggior parte dei problemi di ottimizzazione non lineare, cominciamo col raggruppare i residui r_i in un *vettore dei residui* $\boldsymbol{r}(\boldsymbol{x})$ che vedremo come una funzione da $\mathbb{R}^n$ in $\mathbb{R}^m$:

$$\boldsymbol{r}(\boldsymbol{x}) = (r_1(\boldsymbol{x}), r_2(\boldsymbol{x}), \ldots, r_m(\boldsymbol{x}))^T.$$

Con questa notazione possiamo scrivere

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{r}(\boldsymbol{x})^T\boldsymbol{r}(\boldsymbol{x}) = \frac{1}{2}||\boldsymbol{r}(\boldsymbol{x})||^2.$$

Inoltre, le derivate di $f(\boldsymbol{x})$ si posso esprimere in forma compatta utilizzando lo Jacobiano

$$\boldsymbol{J}(\boldsymbol{x}) = \left[\frac{\partial r_i(\boldsymbol{x})}{\partial x_j}\right]_{\substack{i=1,2,\ldots,m \\ j=1,2,\ldots,n}} = \begin{bmatrix} \nabla r_1(\boldsymbol{x})^T \\ \nabla r_2(\boldsymbol{x})^T \\ \vdots \\ \nabla r_m(\boldsymbol{x})^T \end{bmatrix}$$

dove $\nabla r_i(\boldsymbol{x})$, con $i = 1, ..., m$, denota il vettore gradiente dei residui $r_i(\boldsymbol{x})$.
Con questa notazione il vettore gradiente e la matrice hessiana di $f(\boldsymbol{x})$ si possono esprimere come segue:

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} \sum_{i=1}^m \frac{\partial r_i(\boldsymbol{x})}{\partial x_1} r_i(\boldsymbol{x}) \\ \vdots \\ \sum_{i=1}^m \frac{\partial r_i(\boldsymbol{x})}{\partial x_n} r_i(\boldsymbol{x}) \end{pmatrix} = \boldsymbol{J}(\boldsymbol{x})^T\boldsymbol{r}(\boldsymbol{x}), \tag{25}$$

$$H(\boldsymbol{x}) = \boldsymbol{J}(\boldsymbol{x})^T\boldsymbol{J}(\boldsymbol{x}) + \sum_{i=1}^m \nabla^2 r_i(\boldsymbol{x})\ r_i(\boldsymbol{x}) \tag{26}$$

dove $\nabla^2 r_i(\boldsymbol{x})$ è la matrice hessiana dei residui.

Continua esempio. La matrice Jacobiana dell'esempio precedente risulta

$$\boldsymbol{J}(\boldsymbol{x}) = \begin{bmatrix} t_1 e^{t_1 x_1} & -t_1 \sin(t_1 x_2) \\ t_2 e^{t_2 x_1} & -t_2 \sin(t_2 x_2) \\ t_3 e^{t_3 x_1} & -t_3 \sin(t_3 x_2) \\ t_4 e^{t_4 x_1} & -t_4 \sin(t_4 x_2) \end{bmatrix},$$

il vettore gradiente di $f(\boldsymbol{x})$ diviene

$$\nabla f(\boldsymbol{x}) = \boldsymbol{J}(\boldsymbol{x})^T r(\boldsymbol{x}) = \begin{pmatrix} \sum_{i=1}^4 t_i e^{t_i x_1} r_i(\boldsymbol{x}) \\ -\sum_{i=1}^4 t_i \sin(t_i x_2) r_i(\boldsymbol{x}) \end{pmatrix},$$

il primo termine della matrice hessiana $H(\boldsymbol{x})$ diviene

$$\boldsymbol{J}(\boldsymbol{x})^T\boldsymbol{J}(\boldsymbol{x}) = \begin{pmatrix} \sum_{i=1}^4 t_i^2 e^{2t_i x_1} & -\sum_{i=1}^4 t_i^2 e^{t_i x_1} \sin(t_i x_2) \\ -\sum_{i=1}^4 t_i^2 e^{t_i x_1} \sin(t_i x_2) & \sum_{i=1}^4 t_i^2 \sin^2(t_i x_2) \end{pmatrix},$$

la matrice hessiana dei residui diviene

$$\nabla^2 r_i(\boldsymbol{x}) = \begin{pmatrix} t_i^2 e^{t_i x_1} & 0 \\ 0 & -t_i^2 \cos(t_i x_2) \end{pmatrix},$$

ed infine la matrice hessiana $H(\boldsymbol{x})$ risulta

$$H(\boldsymbol{x}) = \begin{pmatrix} \sum_{i=1}^4 t_i^2 e^{t_i x_1}(e^{t_i x_1} + r_i(\boldsymbol{x})) & -\sum_{i=1}^4 t_i^2 e^{t_i x_1} \sin(t_i x_2) \\ -\sum_{i=1}^4 t_i^2 e^{t_i x_1} \sin(t_i x_2) & \sum_{i=1}^4 t_i^2 (\sin^2(t_i x_2) - \cos(t_i x_2) r_i(\boldsymbol{x})) \end{pmatrix}.$$

In molte applicazioni le derivate pariziali dei residui e quindi lo Jacobiano sono calcolabili in modo efficiente, per cui risulta utile applicare la formula (25). Inoltre è parimenti possibile calcolare il primo termine $\boldsymbol{J}(\boldsymbol{x})^T\boldsymbol{J}(\boldsymbol{x})$ della matrice hessiana in modo non dispendioso, non dovendo calcolare le derivate parziali seconde di $r_i(\boldsymbol{x})$. Questo aspetto è il punto distintivo dei problemi ai minimi quadrati ed è quello che viene sfruttato dagli algoritmi specializzati a risolvere tali problemi. L'idea infatti è che spesso il primo termine della matrice hessiana nell'espressione (26) è più rilevante del secondo, o poiché sono piccoli i residui $r_i(\boldsymbol{x})$, o poiché sono piccole le norme delle loro matrici hessiane $\nabla^2 r_i(\boldsymbol{x})$.
Nelle prossime due sezioni vedremo due metodi di risoluzione per problemi ai minimi quadrati non lineari, il primo appartenente alla famiglia dei metodi basati su ricerca lineare, mentre il secondo è l'antesignano dei metodi di Trust-Region.

8.1 Il metodo di Gauss-Newton

Il metodo di Gauss-Newton è un metodo basato su ricerca lineare

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k$$

che si può interpretare come un metodo di Newton modificato. Per semplicità, nel seguito lasceremo cadere il pedice k relativo all'iterazione corrente. Per ricavare la direzione $\boldsymbol{p}$, invece di risolvere il sistema $H(\boldsymbol{x})\boldsymbol{p} = -\nabla f(\boldsymbol{x})$ si risolve un analogo sistema dove la matrice hessiana viene approssimata trascurando il secondo termine nell'espressione (26):

$$\boldsymbol{J}(\boldsymbol{x})^T\boldsymbol{J}(\boldsymbol{x})\boldsymbol{p}^{GN} = -\boldsymbol{J}(\boldsymbol{x})^T\boldsymbol{r}(\boldsymbol{x}). \tag{27}$$

Questo modello presenta una serie di vantaggi.

- Innanzitutto, in termini di tempo di calcolo, il risparmio dato dal non dover calcolare le derivate seconde dei residui può essere determinante in molte applicazioni, oltre al fatto che il termine $\boldsymbol{J}^T\boldsymbol{J}$ risulta semplice da calcolare una volta calcolato il termine $\boldsymbol{J}^T\boldsymbol{r}$.

- In secondo luogo, l'approssimazione introdotta spesso non è significativa, e questo comporta che il metodo di Gauss-Newton ha un tasso di convergenza simile a quello del metodo di Newton, specialmente in prossimità della soluzione ottima.

- Un terzo vantaggio è che, ogniqualvolta la matrice $\boldsymbol{J}$ ha rango pieno e il gradiente $\nabla f(\boldsymbol{x})$ è non nullo, la direzione $\boldsymbol{p}^{GN}$ è una direzione di discesa per $f(\boldsymbol{x})$ e quindi adatta ad essere utilizzata nei metodi basati su ricerca lineare. Per riconoscere tale proprietà si ricava

  $$(\boldsymbol{p}^{GN})^T\nabla f(\boldsymbol{x}) = (\boldsymbol{p}^{GN})^T\boldsymbol{J}^T\boldsymbol{r} = -(\boldsymbol{p}^{GN})^T\boldsymbol{J}^T\boldsymbol{J}\boldsymbol{p}^{GN} = -||\boldsymbol{J}\boldsymbol{p}^{GN}||^2 \leq 0.$$

  Tale disuguaglianza è stretta, a meno di non essere in un punto stazionario dove si annulla il gradiente $\boldsymbol{J}^T\boldsymbol{r} = \boldsymbol{0}$ e quindi $\boldsymbol{J}\boldsymbol{p}^{GN} = \boldsymbol{0}$.

- Un ulteriore vantaggio deriva dal fatto che $\boldsymbol{p}^{GN}$ è la soluzione del modello ai minimi quadrati lineare

  $$\min_{\boldsymbol{p}\in\mathbb{R}^n} f(\boldsymbol{p}) = \frac{1}{2}||\boldsymbol{J}\boldsymbol{p} + \boldsymbol{r}||^2.$$

  per il quale sono disponibili algoritmi specializzati molto efficienti basati principalmente su particolari fattorizzazioni della matrice $\boldsymbol{J}^T\boldsymbol{J}$.

Le implementazioni più diffuse del metodo Gauss-Newton eseguono una ricerca lineare per la determinazione del passo α lungo la direzione $\boldsymbol{p}^{GN}$, e richiedono che il valore di α soddisfi le condizioni di Armijo (vedi equazione (2)) e di Wolfe (vedi equazione (3)).

Il metodo di Gauss-Newton converge globalmente qualora si possa garantire che lo Jacobiano $\boldsymbol{J}(\boldsymbol{x})$ soddisfi la seguente relazione

$$||\boldsymbol{J}(\boldsymbol{x})\boldsymbol{z}|| \geq \gamma||\boldsymbol{z}|| \tag{28}$$

in tutti i punti $\boldsymbol{x}$ nell'intorno dell'insieme di livello $\{\boldsymbol{x}|f(\boldsymbol{x}) \leq f(\boldsymbol{x}_0)\}$, dove $\gamma > 0$ e $\boldsymbol{x}_0$ è la soluzione di partenza dell'algoritmo.
Nei casi in cui non vale la relazione (28) si dice che la matrice $\boldsymbol{J}$ è *rank-deficient* e come conseguenza la matrice di coefficienti $\boldsymbol{J}^T\boldsymbol{J}$ risulta singolare. Il sistema (27) ammette ancora soluzione, in effetti ne ammette infinite, e questo provoca difficoltà di convergenza. In questi casi risulta più efficace il metodo descritto nella prossima sezione.

8.2 Il metodo di Levenberg-Marquardt

Mentre il metodo di Gauss-Newton è un metodo basato su ricerca lineare, il metodo di Levenberg-Marquardt è un metodo di tipo Trust-Region. Anche in questo caso la matrice hessiana viene approssimata limitandosi a considerare il primo termine dell'espressione (26) ed il metodo che si ricava risulta efficace anche in presenza di matrici $\boldsymbol{J}$ *rank-deficient*.
Il problema che viene risolto ad ogni iterazione è

$$\min_{\boldsymbol{p}\in\mathbb{R}^n} \frac{1}{2}||\boldsymbol{J}\boldsymbol{p} + \boldsymbol{r}||^2, \text{ t.c. } ||\boldsymbol{p}|| \leq \Delta. \tag{29}$$

dove ancora una volta si è omesso il pedice k relativo all'iterazione corrente. Il relativo modello quadratico risulta

$$\min_{\boldsymbol{p}\in\mathbb{R}^n} m(\boldsymbol{p}) = \frac{1}{2}||\boldsymbol{r}||^2 + \boldsymbol{p}^T\boldsymbol{J}^T\boldsymbol{r} + \frac{1}{2}\boldsymbol{p}^T\boldsymbol{J}^T\boldsymbol{J}\boldsymbol{p}. \tag{30}$$

Il Teorema 15 ci permette di caratterizzare la soluzione del problema (29) nel seguente modo.
- Quando la soluzione $\boldsymbol{p}^{GN}$ dell'equazione (27) è interna alla regione di confidenza ($||\boldsymbol{p}^{GN}|| < \Delta$) questa è anche la soluzione del problema (29).
- Altrimenti esiste un valore $\lambda > 0$ tale che la soluzione $\boldsymbol{p}^{LM}$ del problema (29) sta nel bordo della regione di confidenza ($||\boldsymbol{p}^{LM}|| = \Delta$) e vale la relazione

$$(\boldsymbol{J}^T\boldsymbol{J} + \lambda I)\boldsymbol{p}^{LM} = -\boldsymbol{J}^T\boldsymbol{r}.$$

I metodi più efficaci cercano di identificare iterativamente il valore λ che permette di soddisfare quest'ultima uguaglianza. Essi si basano su particolari fattorizzazioni della matrice $\boldsymbol{J}^T\boldsymbol{J} + \lambda I$.
Il metodo di Levenberg-Marquardt converge sotto le lasche ipotesi di convergenza dei metodi di Trust-Region.

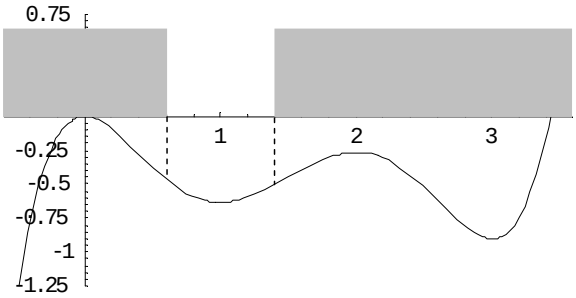


Figura 13: Esempio di funzione obiettivo convessa solo nell'insieme ammissibile.

9 Ottimizzazione vincolata

Consideriamo ora il generico problema di ottimizzazione vincolata

$$\begin{array}{lll} \min & f(\boldsymbol{x}) & \\ g_j(\boldsymbol{x}) & \leq 0 & j = 1, \ldots, k; \\ h_j(\boldsymbol{x}) & = 0 & j = 1, \ldots, h \end{array}$$

con $\boldsymbol{x} \in \mathbb{R}^n$.
Inizieremo con lo studiare le condizioni analitiche di ottimalità. Nel caso dei problemi non vincolati, tutti i minimi locali soddisfano le condizioni necessarie di ottimalità e, almeno in linea di principio, i minimi locali possono venir cercati all'interno dell'insieme dei punti stazionari. Nel caso dei problemi vincolati, invece, non sempre si possono ricavare tutti i minimi locali anche quando risulta possibile imporre il soddisfacimento delle condizioni analitiche.
Nel seguito analizzeremo in primo luogo la regione ammissibile e le relative proprietà, in seguito distingueremo i problemi vincolati analizzando separatemente il caso con soli vincoli di uguaglianza, quello con soli vincoli di disuguaglianza ed infine il caso generale.
Inoltre tratteremo a parte il caso di vincoli lineari ed i problemi quadratici.

Osserviamo innanzitutto che la presenza di vincoli può sia rendere più semplice un problema non vincolato difficile, sia rendere difficile un problema che in assenza di vincoli sarebbe semplice.
In Figura 13 vediamo l'andamento della funzione monodimensionale

$$f(x) = \frac{1}{5}x^5 - \frac{3}{2}x^4 + \frac{11}{3}x^3 - 3x^2.$$

Tale funzione è non convessa in $\mathbb{R}$, mentre è convessa in opportuni intervalli. Ad esempio, nell'intervallo evidenziato $X = [0.6, 1.4]$. In particolare, qualsiasi tecnica di ottimizzazione monodimensionale a partire da un punto $x_0 \in X$ determinerebbe rapidamente l'ottimo globale del problema.
Al contrario invece, in Figura 14 vediamo le curve di livello e la regione ammissibile, in chiaro, del problema

$$\begin{array}{ll} \min & f(x,y) = (x-1)^2 + (y+1)^2 \\ & g_1(x,y) = 1 + \frac{1}{4}\sin(8x) - y \leq 0; \\ & g_2(x,y) = -y \leq 0. \end{array}$$

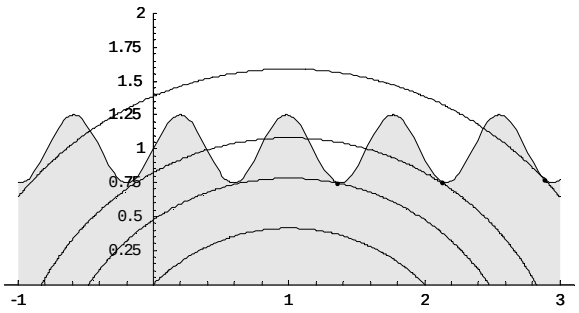


Figura 14: Esempio di funzione obiettivo convessa e infiniti minimi locali.

La funzione obiettivo è convessa ed ammette un unico punto stazionario, $(1, -1)$, che sarebbe un minimo globale per il problema non vincolato. Il problema vincolato ammette invece infiniti minimi locali, tre dei quali evidenziati in figura.

9.1 Condizioni analitiche: vincoli di uguaglianza

Analizziamo in primo luogo un problema vincolato da soli vincoli di uguaglianza

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ h_j(\boldsymbol{x}) & = 0 \quad j = 1, \ldots, h < n \end{array}$$

Nel 1760 Lagrange trasformò questo problema vincolato in un problema non vincolato mediante l'introduzione dei cosiddetti moltiplicatori di Lagrange, λ_j, con $j = 1, \ldots, h$ nella formulazione della cosiddetta *funzione Lagrangiana*

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{j=1}^{h} \lambda_j h_j(\boldsymbol{x}) = f(\boldsymbol{x}) + \lambda^T \boldsymbol{h}(\boldsymbol{x}).$$

Le condizioni necessarie per l'esistenza di un minimo del problema vincolato con vincoli di uguaglianza possono essere date in termini della funzione Lagrangiana e dei moltiplicatori di Lagrange.

Teorema 16 *Sono date una funzione $f(\boldsymbol{x})$ ed h funzioni $h_j(\boldsymbol{x})$, con $j = 1, \ldots, h$ di classe C^1. Condizioni necessarie, nell'ipotesi che i vettori gradienti delle funzioni h_j calcolati nel punto $\boldsymbol{x}^*$ siano tra loro linearmente indipendenti, affinché $\boldsymbol{x}^*$ sia un minimo locale del problema con vincoli di uguaglianza è che esista $\boldsymbol{\lambda}^*$ tale che $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ sia un punto stazionario della funzione $L(\boldsymbol{x}, \boldsymbol{\lambda})$, cioè:*

$$\frac{\partial L}{\partial x_i} = \frac{\partial f(\boldsymbol{x}^*)}{\partial x_i} + \sum_{j=1}^{h} \lambda_j^* \frac{\partial h_j(\boldsymbol{x}^*)}{\partial x_i} = 0, \quad i = 1, 2, \ldots, n \tag{31}$$

$$\frac{\partial L}{\partial \lambda_j} = h_j(\boldsymbol{x}^*) = 0, \quad j = 1, 2, \ldots, h \tag{32}$$

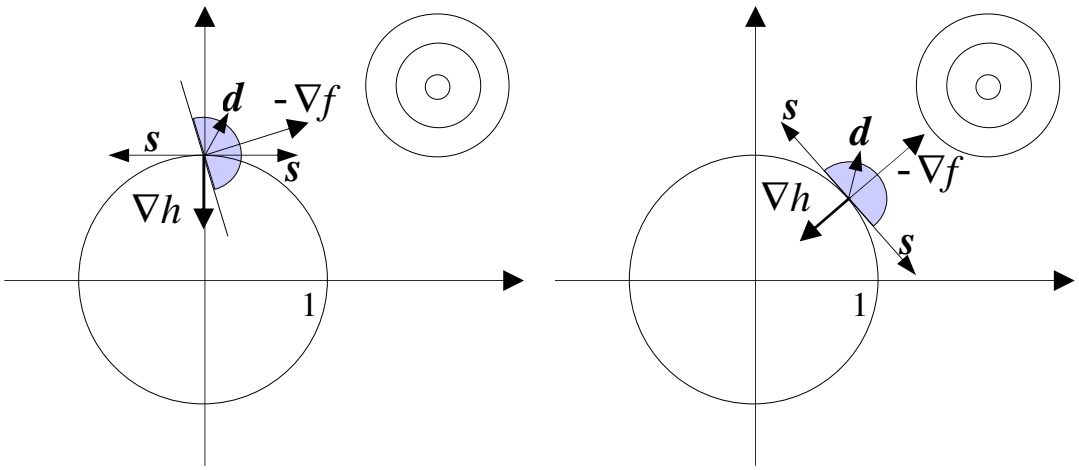
Le condizioni rappresentano un sistema di $n + h$ equazioni in $n + h$ incognite. Il secondo gruppo di h condizioni coincide con la richiesta che i vincoli di uguaglianza siano soddisfatti nel punto di ottimo. Il primo gruppo di n condizioni coincide con $\nabla f(\boldsymbol{x}^*) + J(\boldsymbol{x}^*)^T \boldsymbol{\lambda}^* = \boldsymbol{0}$, ovvero

$$-\nabla f(\boldsymbol{x}^*) = \sum_{j=1}^{h} \lambda_j^* \nabla h_j(\boldsymbol{x}^*)$$

che dal punto di vista geometrico esprime la richiesta che l'antigradiente della funzione obiettivo si possa ottenere come combinazione lineare dei gradienti dei vincoli di uguaglianza.
Esempio. Consideriamo il problema

$$\begin{aligned} \min \quad & f(x,y) = (x-2)^2 + (y-2)^2 \\ & h_1(x,y) = 1 - x^2 - y^2 = 0. \end{aligned}$$

Il punto di ottimo è $(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$ e in tale punto l'antigradiente $-\nabla f(x,y) = -(2(x-2), 2(y-2))^T$ vale $(4-\sqrt{2}, 4-\sqrt{2})^T$. Nello stesso punto il gradiente di h_1, $\nabla h(x,y) = (-2x, -2y)^T$, vale $-(\sqrt{2}, \sqrt{2})$. Come illustra la Figura 15 di destra i due vettori sono collineari e $\lambda^* = \frac{4-\sqrt{2}}{\sqrt{2}}$.



(a) Punto non ottimo

(b) Punto ottimo

Figura 15: Condizione di ottimalità per problemi con vincoli di uguaglianza.

Ma che cosa caratterizza tale punto geometricamente? Osserviamo nella Figura 15 di sinistra il punto non ottimale $P_1 = (0,1)$. In tale punto sono evidenziate le direzioni $\boldsymbol{s}$ ortogonali al vettore $\nabla h(x,y)$, cioè tali che $\nabla h(x,y)^T \boldsymbol{s} = 0$. L'insieme di queste direzioni definisce un iperpiano (in questo caso una retta), diciamo F, che rappresenta l'approssimazione al primo ordine della funzione $h(x,y)$ in P_1. In P_1 è inoltre evidenziato, mediante un semicerchio ombreggiato, il sottospazio, diciamo D, definito da tutte le direzioni di discesa $\boldsymbol{d}$, che formano cioè con l'antigradiente $-\nabla f(x,y)$ un angolo θ tale che $\cos\theta > 0$, quindi tali che

$$\nabla f(x,y)^T \boldsymbol{d} < 0.$$

Il punto P_1 non è ottimo in quanto in tale punto esistono direzioni che appartengono sia ad F che a D, seguendo le quali, almeno per un tratto infinitesimo, si migliora la funzione obiettivo e si continua a soddisfare il vincolo di uguaglianza.
Nella Figura 15 di destra il punto è ottimo proprio perché in tale punto non esistono direzioni che appartengono a $F \cap D$: le direzioni miglioranti non appartengono alla linearizzazione del vincolo di uguaglianza. Tale condizione si ha quando i vettori $-\nabla f(x,y)$ e $\nabla h(x,y)$ sono

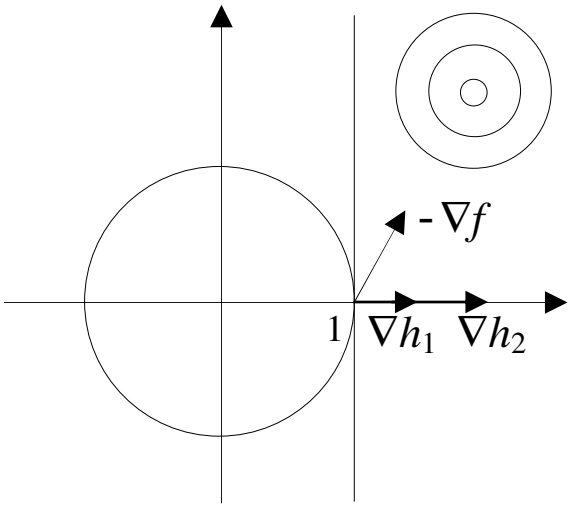


Figura 16: Esempio di gradienti di vincoli di uguaglianza linearmente dipendenti.

collineari.

Il Teorema 16 fa però menzione dell'ulteriore richiesta che:

> [...] i vettori gradienti delle funzioni h_j calcolati nel punto $\boldsymbol{x}^*$ siano tra loro linearmente indipendenti [...]

Vediamo con un esempio il significato geometrico di tale richiesta.

Esempio. Consideriamo il problema

$$\begin{array}{ll} \min & f(x,y) = (x-2)^2 + (y-2)^2 \\ & h_1(x,y) = x^2 + y^2 - 1 = 0; \\ & h_2(x,y) = x - 1 = 0. \end{array}$$

Il punto di ottimo è l'unico punto ammissibile $(1,0)$. In tale punto i vettori $\nabla h_1(x,y)$ e $\nabla h_2(x,y)$ sono $(2,0)^T$ e $(1,0)^T$, rispettivamente, sono collineari e quindi linearmente dipendenti. Nello stesso punto l'antigradiente della funzione obiettivo, $-\nabla f(x,y)$, vale $(2,4)^T$ e per nessun valore di λ_1 e λ_2 può venir soddisfatto il sistema di equazioni

$$\begin{pmatrix} 2 \\ 4 \end{pmatrix} = \lambda_1 \begin{pmatrix} 2 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

D'altro canto se la funzione obiettivo fosse stata $f(x,y) = (x-2)^2 + y^2$ il sistema avrebbe avuto soluzione anche in presenza di vincoli i cui gradienti non sono linearmente indipendenti. Questo perché l'antigradiente della funzione obiettivo sarebbe stato comunque generabile mediante una combinazione lineare del sottoinsieme dei gradienti dei soli vincoli linearmente indipendenti (in questo caso uno solo).

Siamo in presenza di una delle particolarità dell'ottimizzazione vincolata. L'imposizione delle condizioni analitiche di ottimalità non garantisce di individuare tutti i punti di minimo locale a meno che i vincoli che definiscono la regione ammissibile non soddisfino alcune peculiari

condizioni. Poiché il problema nasce dal fatto che non sempre lo spazio definito dall'approssimazione lineare delle funzioni h_j (e come vedremo in seguito, g_j) costituisce una buona approssimazione locale di tali funzioni, occorre identificare in quali casi ciò accade. Le condizioni in esame sono dette condizioni di *qualificazione dei vincoli* e devono valere in un punto di ottimo $\boldsymbol{x}^*$, sia per i vincoli di uguaglianza sia per i vincoli di disuguaglianza che sono soddisfatti come uguaglianza in $\boldsymbol{x}^*$.

Condizione 4 Qualificazione dei vincoli *In un punto $\boldsymbol{x}^*$ diciamo che sono soddisfatte le condizioni di qualificazione dei vincoli se esiste un vettore $\boldsymbol{h}$ tale che $\nabla g_j(\boldsymbol{x}^*)^T\boldsymbol{h} < 0$, in corrispondenza di tutti gli indici j tali che $g_j(\boldsymbol{x}^*) = 0$, $\nabla h_j(\boldsymbol{x}^*)^T\boldsymbol{h} = 0$ con $j = 1, 2, \ldots, h$ ed i vettori $\nabla h_j(\boldsymbol{x}^*)$ con $j = 1, 2, \ldots, h$ sono linearmente indipendenti.*

Le condizioni di qualificazione dei vincoli sono sempre soddisfatte se:

- i gradienti dei vincoli di uguaglianza e dei vincoli attivi in $\boldsymbol{x}^*$ sono fra loro linearmente indipendenti, ovvero se lo Jacobiano dei vincoli di uguaglianza e dei vincoli attivi ha rango massimo in $\boldsymbol{x}^*$;
- se tutti i vincoli sono lineari e
- se tutti i vincoli sono convessi e la regione ammissibile contiene almeno un punto interno.

Definizione 15 *Si dice* punto regolare *un punto $\boldsymbol{x}^*$ che soddisfa le condizioni di qualificazione dei vincoli.*

Richiamiamo infine che nel caso di funzioni convesse le condizioni introdotte nel Teorema 16 diventano *condizioni sufficienti*

Teorema 17 *Sono date una funzione convessa $f(\boldsymbol{x})$ ed h funzioni convesse $h_j(\boldsymbol{x})$, con $j = 1, \ldots, h$ di classe C^1. Condizioni sufficienti, nell'ipotesi che la matrice Jacobiana delle funzioni $h_j(\boldsymbol{x})$, calcolate nel punto $\boldsymbol{x}^*$, sia di rango h, affinché $\boldsymbol{x}^*$ sia un minimo locale del problema con vincoli di uguaglianza è che esista $\boldsymbol{\lambda}^*$ tale che $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ sia un punto stazionario della funzione $L(\boldsymbol{x}, \boldsymbol{\lambda})$.*

9.1.1 Funzione obiettivo quadratica e vincoli di uguaglianza lineari

Vediamo ora come si applica il metodo lagrangiano alla minimizzazione di una funzione obiettivo quadratica *definita positiva*, soggetta a vincoli di uguaglianza lineari

$$\begin{array}{ll} \min & f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T Q\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x} \\ t.c. & A\boldsymbol{x} = \boldsymbol{d} \end{array}$$

Per ipotesi Q è definita positiva ed A è una matrice $(h \times n)$ di rango pieno con $h < n$. Poiché i vincoli sono lineari essi soddisfano le *condizioni di qualificazione*. In questo caso la funzione lagrangiana è

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{x}^T Q\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x} + \boldsymbol{\lambda}^T(\boldsymbol{d} - A\boldsymbol{x})$$

e le condizioni necessarie per l'esistenza di un minimo vincolato in $\boldsymbol{x}^*$ è che esista un vettore $\boldsymbol{\lambda}^*$ tale che:

$$\begin{aligned} \nabla_x L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) &= Q\boldsymbol{x}^* - \boldsymbol{b} - A^T\boldsymbol{\lambda}^* = \mathbf{0} \\ \nabla_\lambda L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) &= A\boldsymbol{x}^* - \boldsymbol{d} = \mathbf{0} \end{aligned}$$

Tali condizioni possono essere riscritte nella forma:

$$\begin{bmatrix} Q & -A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{x}^* \\ \boldsymbol{\lambda}^* \end{bmatrix} = \begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{d} \end{bmatrix}$$

la cui soluzione è

$$\begin{bmatrix} \boldsymbol{x}^* \\ \boldsymbol{\lambda}^* \end{bmatrix} = \begin{bmatrix} Q & -A^T \\ A & 0 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{d} \end{bmatrix}$$

9.1.2 Da vincoli di disuguaglianza a vincoli di uguaglianza

Riprendiamo ora il problema generale

$$\begin{array}{lll} \min & f(\boldsymbol{x}) & \\ g_j(\boldsymbol{x}) & \leq 0 & j = 1, \ldots, k; \\ h_j(\boldsymbol{x}) & = 0 & j = 1, \ldots, h \end{array}$$

Una prima tecnica per la soluzione del problema generale lo riconduce ad un problema con soli vincoli di uguaglianza mediante l'introduzione di variabili ausiliarie. Ciascun vincolo di disuguaglianza viene trasformato in un vincolo di uguaglianza mediante l'aggiunta di una variabile ausiliaria. Si passa da $g_i(\boldsymbol{x}) \leq 0$ a $g_i(\boldsymbol{x}) + \theta_i^2 = 0$. A differenza dell'analoga trasformazione che permette di passare dalla forma generale alla forma standard nell'ambito della programmazione lineare, in questo caso le variabili ausiliarie vengono elevate al quadrato. Elevare al quadrato le variabili ausiliarie elimina la necessità di introdurre le condizioni di non negatività $\theta_i \geq 0$, cioè altre disuguaglianze. Il problema diventa quindi:

$$\begin{array}{lll} \min & f(\boldsymbol{x}) & \\ g_j(\boldsymbol{x}) + \theta_j^2 & = 0 & j = 1, \ldots, k; \\ h_j(\boldsymbol{x}) & = 0 & j = 1, \ldots, h \end{array}$$

a cui corrisponde il modello lagrangiano:

$$L(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \sum_{j=1}^{k} \lambda_j (g_j(\boldsymbol{x}) + \theta_j^2) + \sum_{j=1}^{h} \mu_j h_j(\boldsymbol{x})$$

Le condizioni necessarie perché un punto $\boldsymbol{x}$ sia di minimo sono

$$\begin{aligned} \frac{\partial L}{\partial x_i} &= \frac{\partial f(\boldsymbol{x})}{\partial x_i} + \sum_{j=1}^{k} \lambda_j \frac{\partial g_j(\boldsymbol{x})}{\partial x_i} + \sum_{j=1}^{h} \mu_j \frac{\partial h_j(\boldsymbol{x})}{\partial x_i} = 0, \quad i = 1, 2, \ldots, n \\ \frac{\partial L}{\partial \theta_j} &= 2\lambda_j \theta_j = 0, \quad j = 1, 2, \ldots, k \\ \frac{\partial L}{\partial \lambda_j} &= g_j(\boldsymbol{x}) + \theta_j^2 = 0, \quad j = 1, 2, \ldots, k \\ \frac{\partial L}{\partial \mu_j} &= h_j(\boldsymbol{x}) = 0, \quad j = 1, 2, \ldots, h \end{aligned}$$

Tali condizioni rappresentano un sistema di $n + 2k + h$ equazioni in $n + 2k + h$ incognite, la cui (eventuale) soluzione individua i punti candidati ad essere soluzione del problema di ottimizzazione.

Si osservi che le k relazioni $2\lambda_j\theta_j = 0$, con $j = 1, 2, \ldots, k$, sono relazioni di scarto complementare, in quanto impongono che il moltiplicatore λ_j sia nullo ogniqualvolta il vincolo $g_j(\boldsymbol{x}) \leq 0$ è soddisfatto come disuguaglianza stretta, e che il vincolo sia soddisfatto come uguaglianza ogniqualvolta λ_j è diverso da zero.

9.2 Il caso generale: le condizioni KKT

Partendo dal problema generale

$$\begin{array}{lll} \min & f(\boldsymbol{x}) & \\ g_j(\boldsymbol{x}) & \leq 0 & j = 1, \ldots, k; \\ h_j(\boldsymbol{x}) & = 0 & j = 1, \ldots, h \end{array} \tag{33}$$

e dal relativo modello lagrangiano:

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \sum\nolimits_{j=1}^{k} \lambda_j g_j(\boldsymbol{x}) + \sum\nolimits_{j=1}^{h} \mu_j h_j(\boldsymbol{x})$$

Karush (1939) e Kuhn e Tucker (1951) derivarono indipendentemente alcune condizioni necessarie affinché un punto $\boldsymbol{x}$ sia un punto di minimo locale. Tali condizioni sono ora note come *condizioni KKT*.

Teorema 18 *Sia dato un problema in forma generale (33), dove le funzioni $f(\boldsymbol{x})$, $g_j(\boldsymbol{x})$ e $h_j(\boldsymbol{x})$ sono tutte di classe C^1. Se $\boldsymbol{x}^*$ è un minimo locale e in $\boldsymbol{x}^*$ valgono le condizioni di qualificazione dei vincoli di uguaglianza e di quelli di disuguaglianza attivi, allora esistono moltiplicatori di Lagrange $\boldsymbol{\lambda}^*$ e $\boldsymbol{\mu}^*$, tali che le seguenti condizioni sono soddisfatte,*

$$\begin{array}{lll} \frac{\partial f(\boldsymbol{x}^*)}{\partial x_i} + & \sum_{j=1}^{k} \lambda_j^* \frac{\partial g_j(\boldsymbol{x}^*)}{\partial x_i} + \sum_{j=1}^{h} \mu_j^* \frac{\partial h_j(\boldsymbol{x}^*)}{\partial x_i} = 0, & i = 1, \ldots, n \\ g_j(\boldsymbol{x}^*) & \leqslant 0, & j = 1, \ldots, k \\ \lambda_j^* g_j(\boldsymbol{x}^*) & = 0, & j = 1, \ldots, k \\ h_j(\boldsymbol{x}^*) & = 0, & j = 1, \ldots, h \\ \lambda_j^* & \geq 0, & j = 1, \ldots, k \end{array}$$

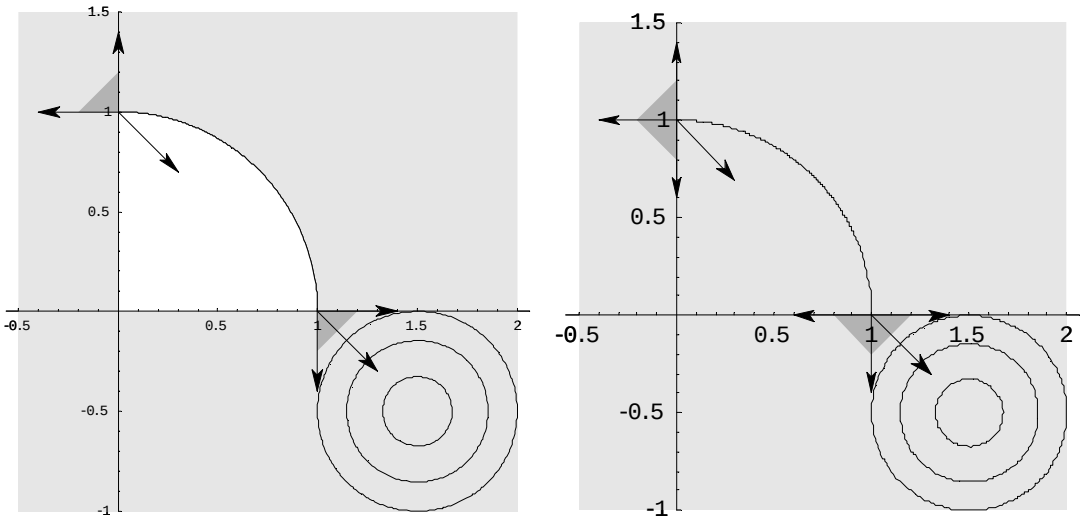
Si osservi che le k relazioni $\lambda_j^* g_j(\boldsymbol{x}^*) = 0$, con $j = 1, 2, \ldots, k$, sono relazioni di scarto complementare, in quanto impongono che $\lambda_j^* = 0$ ogniqualvolta vale $g_j(\boldsymbol{x}^*) > 0$, e che valga $g_j(\boldsymbol{x}^*) = 0$, cioè il vincolo sia attivo ogniqualvolta vale $\lambda_j > 0$. Indichiamo con $I \subseteq \{1, 2, \ldots, k\}$ il sottoinsieme degli indici $1, 2, \ldots, k$, che corrispondono ai vincoli di disuguaglianza attivi. Tenendo a mente queste relazioni il primo gruppo di n condizioni coincide con

$$-\nabla f(\boldsymbol{x}^*) = \sum_{j \in I} \lambda_j^* \nabla g_j(\boldsymbol{x}^*) + \sum\nolimits_{j=1}^{h} \mu_j^* \nabla h_j(\boldsymbol{x}^*) \tag{34}$$

che esprime la

richiesta che nel punto di ottimo $\boldsymbol{x}^*$ l'antigradiente della funzione obiettivo si possa ottenere come combinazione lineare non negativa dei gradienti dei vincoli di disuguaglianza attivi e come combinazione lineare dei vincoli di uguaglianza.

Abbiamo già incontrato questa relazione nell'ambito della programmazione lineare, ma mentre nel contesto della PL siamo sempre in grado di calcolare i valori delle variabili duali, qui è vero solo se sono soddisfatte le condizioni di qualificazione dei vincoli.



(a) Solo vincoli di disuguaglianza

(b) Vincoli di disuguaglianza ed uguaglianza

Figura 17: Condizioni di ottimalità con vincoli di disuguaglianza.

Esempio. Consideriamo il problema di Figura 17 a sinistra

$$\begin{aligned} \min \quad & f(x,y) = (x-1.5)^2 + (y+0.5)^2 \\ & g_1(x,y) = -x \leq 0; \\ & g_2(x,y) = -y \leq 0; \\ & g_3(x,y) = x^2 + y^2 - 1 \leq 0. \end{aligned}$$

Il punto di ottimo è $(1,0)$. In tale punto sono attivi i vincoli g_2 e g_3, i vettori $\nabla g_2(x,y)$ e $\nabla g_3(x,y)$ sono $(0,-1)^T$ e $(2,0)^T$, rispettivamente, e sono linearmente indipendenti. Nello stesso punto l'antigradiente della funzione obiettivo, $\nabla f(x,y)$, vale $(1,-1)^T$ ed il vettore $\boldsymbol{\lambda}^*$ vale quindi $(0,1,1/2)^T$. Come si vede l'antigradiente giace all'interno del cono generato dalle combinazioni lineari non negative dei gradienti dei vincoli attivi nel punto.
Consideriamo ora il punto *non* ottimo $(0,1)$. In tale punto sono attivi i vincoli g_1 e g_3, i vettori $\nabla g_1(x,y)$ e $\nabla g_3(x,y)$ sono $(-1,0)^T$ e $(0,2)^T$, rispettivamente, e sono linearmente indipendenti. Nello stesso punto l'antigradiente della funzione obiettivo, $\nabla f(x,y)$, vale $(3,-3)^T$ e l'unico modo di generarlo mediante combinazione lineare dei gradienti dei vincoli attivi nel punto è per mezzo del vettore $\boldsymbol{\lambda}^T = (-3,0,-3/2)^T$ non ammissibile. Come si vede l'antigradiente giace all'esterno del cono generato dalle combinazioni lineari *non negative* dei gradienti dei vincoli attivi nel punto.

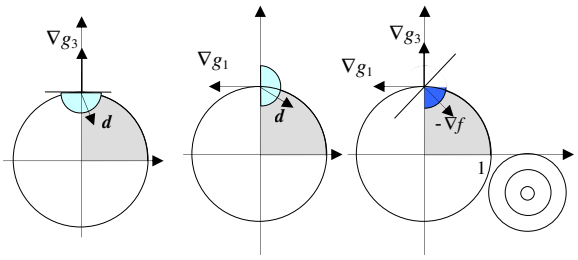


Figura 18: Condizioni di non ottimalità con vincoli di disuguaglianza.

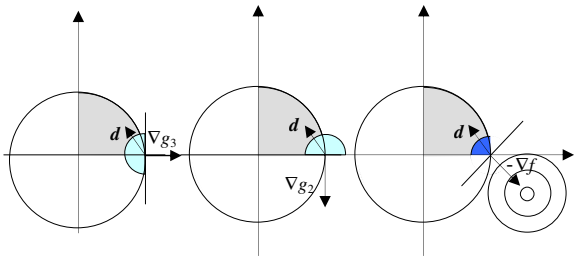


Figura 19: Condizioni di ottimalità con vincoli di disuguaglianza.

Consideriamo ora il problema di Figura 17 a destra

$$\begin{aligned} \min \quad & f(x,y) = (x-1.5)^2 + (y+0.5)^2 \\ & g_1(x,y) = -x \leq 0; \\ & g_2(x,y) = -y \leq 0; \\ & h_1(x,y) = x^2 + y^2 - 1 = 0; \end{aligned}$$

Il punto di ottimo è ancora $(1,0)$ e valgono le considerazioni fatte per il problema precedente, salvo che ora risultano accettabili anche valori negativi del moltiplicatore μ_1 (che prende il posto di λ_3), visto che il relativo vincolo è di uguaglianza.

Esempio continua. In Figura 18 vediamo l'interpretazione geometrica delle condizioni di ottimalità in presenza di vincoli di disuguaglianza. Qui abbiamo rappresentato in grigio la regione ammissibile. Consideriamo il punto *non* ottimo $(0,1)$. Poiché i vincoli di disugualglianza sono posti in forma di $\leq$, i loro gradienti puntano verso l'esterno della regione ammissibile e le direzioni $\boldsymbol{d}$ tali che $\nabla g_j(\boldsymbol{x})^T \boldsymbol{d} \leq 0$ sono direzioni seguendo le quali, almeno per un infinitesimo spostamento, si rimane all'interno della regione ammissibile. Tali direzioni sono evidenziate in figura mediante un semicerchio colorato per entrambi i vincoli attivi $g_1(x,y)$ e $g_3(x,y)$. Le direzioni ammissibili di spostamento sono quindi date dall'itersezione dei diversi semispazi, intersezione evidenziata nella figura di destra, ed un punto non è ottimo se tale intersezione contiene delle direzioni di discesa, come nel caso evidenziato in figura.

In Figura 19 (anche qui abbiamo rappresentato in grigio la regione ammissibile) vediamo come nel caso del punto di ottimo $(1,0)$ invece, l'intersezione dei semispazi che definiscono direzioni ammissibili per i diversi vincoli attivi nel punto, non contenga direzioni di discesa.

Definiamo ora formalmente il concetto di *direzione ammissibile* appena introdotto.

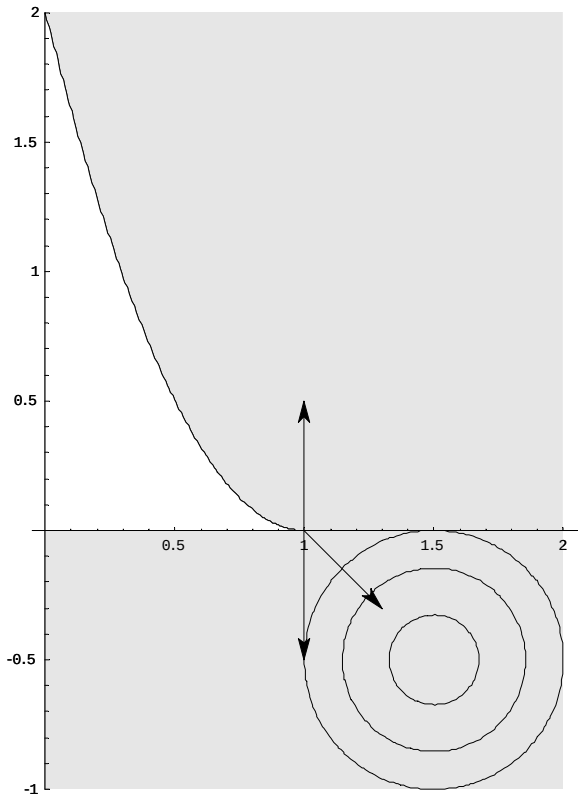


Figura 20: Esempio di gradienti di vincoli attivi linearmente dipendenti.

Definizione 16 *Dato un punto ammissibile* $\boldsymbol{x}$ *e l'insieme* $I \subseteq \{1, 2, \ldots, k\}$ *degli indici che corrispondono a vincoli di disuguaglianza attivi in* $\boldsymbol{x}$*, chiamiamo insieme delle direzioni ammissibili, l'insieme*

$$F(\boldsymbol{x}) = \{\boldsymbol{d} \mid \nabla h_j(\boldsymbol{x})^T \boldsymbol{d} = 0,\ j = 1, \ldots, h;\ \nabla g_j(\boldsymbol{x})^T \boldsymbol{d} \leq 0,\ j \in I\}.$$

Esempio. Consideriamo ora il problema di Figura 20

$$\begin{array}{ll} \min & f(x,y) = (x-1.5)^2 + (y+0.5)^2 \\ & g_1(x,y) = -2(x-1)^3 + y \leq 0; \\ & g_2(x,y) = -y \leq 0. \end{array}$$

Il punto di ottimo è $\boldsymbol{x}^* = (1, 0)$. In tale punto sono attivi entrambi i vincoli g_1 e g_2, i vettori $\nabla g_1(x, y)$ e $\nabla g_2(x, y)$ sono $(0, 1)^T$ e $(0, -1)^T$, rispettivamente, e sono linearmente *dipendenti*. In questo esempio $F(\boldsymbol{x}^*) = \{(d, 0)^T \mid d \in \mathbb{R}\}$. In $\boldsymbol{x}^*$ l'antigradiente della funzione obiettivo, $\nabla f(x, y)$, vale $(1, -1)^T$ e non può esistere alcun vettore $\boldsymbol{\lambda}^*$.

Dal punto di vista operativo, la determinazione di un punto di ottimo per mezzo delle condizioni analitiche nel caso di un problema in forma generale (33) diventa un problema di tipo *combinatorico*. Da un punto di vista puramente teorico si tratta di generare tanti sistemi di equazioni non lineari quanti sono i sottoinsiemi di vincoli di disuguaglianza, dove per ciascun sistema, il relativo insieme di vincoli di disuguaglianza è imposto *attivo* (cioè soddisfatto come

uguaglianza). La maggior parte di tali sistemi non avrà soluzioni ammissibili nello spazio delle variabili $\boldsymbol{x}$ (coinvolgendo vincoli g_j o h_j che non hanno punti in comune nelle rispettive frontiere), mentre altri non saranno ammissibili nello spazio dei moltiplicatori λ_j, poichè alcuni di essi risulteranno negativi. Naturalmente ci riferiamo a casi ideali, poiché come abbiamo visto trattando i problemi non vincolati, quasi mai è possibile determinare i punti stazionari imponendo la soluzione dei sitemi di equazioni derivanti dall'imposizione delle condizioni analitiche. Nel caso dei problemi vincolati c'è l'ulteriore complicazione derivante dall'esistenza di punti non regolari che, come evidenziato dagli esempi nelle Figure 16 e 20, possono essere punti di ottimo globale non ricavabili dalla soluzione di alcuno dei possibili sistemi di equazioni non lineari.

Richiamiamo infine che nel caso di funzioni convesse le condizioni KKT sono sufficienti

Teorema 19 *Sia dato un problema in forma generale (33), dove le funzioni $f(\boldsymbol{x})$, $g_j(\boldsymbol{x})$ e $h_j(\boldsymbol{x})$ sono tutte di classe C^1. Se le funzioni $f(\boldsymbol{x})$, $g_j(\boldsymbol{x})$ e $h_j(\boldsymbol{x})$ sono convesse allora le condizioni KKT sono condizioni sufficienti.*

9.3 Condizioni di ottimalità del secondo ordine

Nel definire le condizioni che un punto di minimo vincolato deve soddisfare abbiamo usato fino ad ora solo condizioni del primo ordine. Le informazioni che ci provengono dall'uso del secondo termine nello sviluppo in serie di Taylor, permettono di trattare lo status delle direzioni ammissibili che sono ortogonali alla direzione del gradiente della funzione obiettivo. Iniziamo definendo l'insieme delle direzioni *critiche*.

Definizione 17 *Dato un punto ammissibile $\boldsymbol{x}^*$ e due vettori di moltiplicatori $\boldsymbol{\lambda}^*$ e $\boldsymbol{\mu}^*$ che soddisfano le condizioni KKT, con $I \subseteq \{1, 2, \ldots, k\}$ a denotare l'insieme degli indici che corrispondono a vincoli di disuguaglianza attivi in $\boldsymbol{x}^*$, chiamiamo insieme delle direzioni critiche, l'insieme*

$$C(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \{\boldsymbol{d} \in F(\boldsymbol{x}^*) \mid \nabla g_j(\boldsymbol{x}^*)^T \boldsymbol{d} = 0,\ j \in I,\ \textit{con } \lambda_j^* > 0\}.$$

Per definizione valgono quindi le relazioni

$$\lambda_j^* \nabla g_j(\boldsymbol{x}^*)^T \boldsymbol{d} = 0, \quad \forall \boldsymbol{d} \in C(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*), \quad j \in I,$$

$$\mu_j^* \nabla h_j(\boldsymbol{x}^*)^T \boldsymbol{d} = 0, \quad \forall \boldsymbol{d} \in C(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*), \quad j = \ldots, h.$$

Dalla relazione 34 possiamo perciò ricavare

$$-\nabla f(\boldsymbol{x}^*)^T \boldsymbol{d} = \sum_{j \in I} \lambda_j^* \nabla g_j(\boldsymbol{x}^*)^T \boldsymbol{d} + \sum\nolimits_{j=1}^{h} \mu_j^* \nabla h_j(\boldsymbol{x}^*)^T \boldsymbol{d} = 0 \ \ \forall \boldsymbol{d} \in C(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*).$$

Le direzioni critiche sono perciò ortogonali alla direzione del gradiente della funzione obiettivo.

In Figura 21 (la regione ammissibile è in grigio) vediamo il problema

$$\begin{aligned} \min \quad & f(x,y) = (x-1.5)^2 + y^2 \\ & g_1(x,y) = -x \leq 0; \\ & g_2(x,y) = -y \leq 0; \\ & g_3(x,y) = x^2 + y^2 - 1 \leq 0; \end{aligned}$$

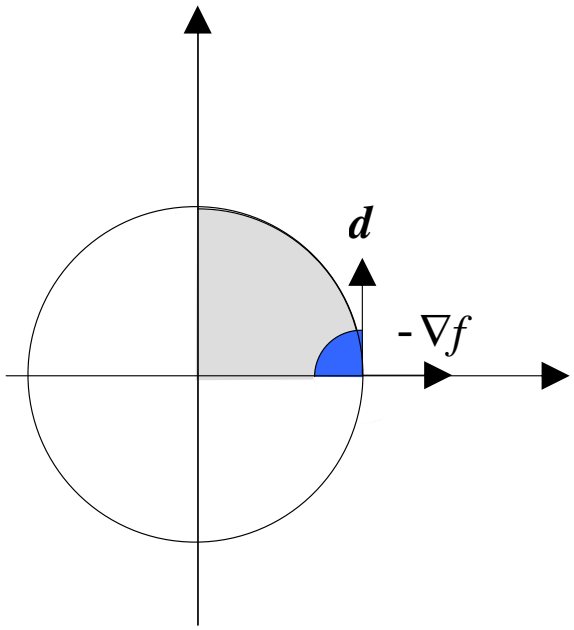


Figura 21: Direzioni critiche.

dove l'insieme delle direzioni critiche è $C(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \{(1, d)^T \mid d \geq 0\}$.

Vediamo ora delle condizioni necessarie del secondo ordine, definite nei termini delle direzioni critiche.

> **Teorema 20** *Sia dato un problema in forma generale (33), dove le funzioni $f(\boldsymbol{x})$, $g_j(\boldsymbol{x})$ e $h_j(\boldsymbol{x})$ sono tutte di classe C^2. Se $\boldsymbol{x}^*$ è un minimo locale e in $\boldsymbol{x}^*$ valgono le condizioni di qualificazione dei vincoli di uguaglianza e di quelli di disuguaglianza attivi, e $\boldsymbol{\lambda}^*$ e $\boldsymbol{\mu}^*$ sono vettori di moltiplicatori che soddisfano le condizioni KKT, allora vale*
>
> $$\boldsymbol{d}^T \nabla^2_{\boldsymbol{x},\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^* \boldsymbol{\mu}^*) \boldsymbol{d} \geq 0 \quad \forall \boldsymbol{d} \in C(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*).$$

Nel teorema appena enunciato la matrice hassiana della funzione lagrangiana è data da

$$\nabla^2_{\boldsymbol{x},\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^* \boldsymbol{\mu}^*) = H(\boldsymbol{x}^*) + \sum_{j \in I} \lambda_j^* \nabla^2 g_j(\boldsymbol{x}^*) + \sum\nolimits_{j=1}^{h} \mu_j^* \nabla^2 h_j(\boldsymbol{x}^*).$$

In pratica si richiede che la matrice hessiana della funzione lagrangiana sia semi definita positiva nell'insieme delle direzioni critiche. Se tale matrice è definita positiva in questo insieme le condizioni risultano sufficienti, e non è più necessario soddisfare le condizioni di qualificazione dei vincoli.

> **Teorema 21** *Sia dato un problema in forma generale (33), dove le funzioni $f(\boldsymbol{x})$, $g_j(\boldsymbol{x})$ e $h_j(\boldsymbol{x})$ sono tutte di classe C^2. Se $\boldsymbol{x}^*$ è un punto ammissibile e $\boldsymbol{\lambda}^*$ e $\boldsymbol{\mu}^*$ sono vettori di moltiplicatori che soddisfano le condizioni KKT, e vale*
>
> $$\boldsymbol{d}^T \nabla^2_{\boldsymbol{x},\boldsymbol{x}} L(\boldsymbol{x}^*, \boldsymbol{\lambda}^* \boldsymbol{\mu}^*) \boldsymbol{d} \ > \ 0 \quad \forall \boldsymbol{d} \in C(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$$
>
> *allora $\boldsymbol{x}^*$ è un minimo locale in senso stretto del problema (33).*

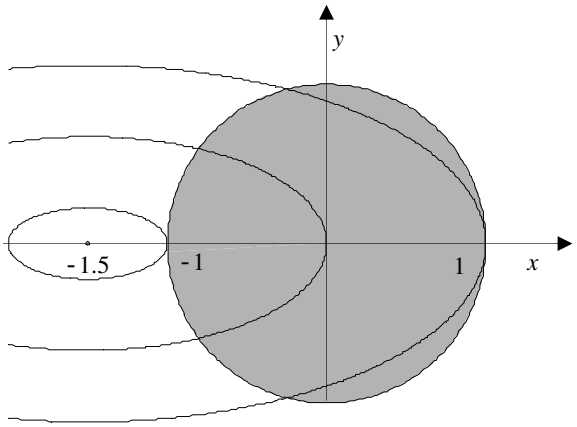


Figura 22: Esempio sulle condizioni del secondo ordine.

Esempio In Figura 22 (la regione ammissibile è in chiaro) vediamo il problema

$$\begin{array}{ll} \min & f(x,y) = 2(x+1.5)^2 + 10y^2 \\ & g_1(x,y) = 1 - x^2 - y^2 \leq 0; \end{array}$$

che ammette un punto di minimo globale in $(-1.5, 0)^T$ dove g_1 non è attivo e $\lambda^* = 0$, ed un punto di minimo locale in senso stretto $\tilde{\boldsymbol{x}} = (1,0)$ dove invece $g_1(1,0) = 0$. Verifichiamo questa seconda affermazione. In $\tilde{\boldsymbol{x}}$ valgono le condizioni KKT con

$$\begin{pmatrix} 4(x+1.5) \\ 20y \end{pmatrix} - \lambda_1 \begin{pmatrix} 2x \\ 2y \end{pmatrix} = \begin{pmatrix} 10 \\ 0 \end{pmatrix} - \lambda_1 \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

con $\lambda_1^* = 5$. La matrice hessiana della funzione lagrangiana risulta

$$\nabla^2_{\boldsymbol{xx}} L(\tilde{\boldsymbol{x}}, \lambda_1^*) = \begin{pmatrix} 4 & 0 \\ 0 & 20 \end{pmatrix} - \lambda_1^* \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 4 - 2\lambda_1^* & 0 \\ 0 & 20 - 2\lambda_1^* \end{pmatrix} = \begin{pmatrix} -6 & 0 \\ 0 & 10 \end{pmatrix}$$

Il gradiente del solo vincolo attivo in $\tilde{\boldsymbol{x}}$ è $\nabla g_1(\tilde{\boldsymbol{x}}) = (2,0)^T$ e l'insieme delle direzioni critiche ad esso ortogonali è $C(\tilde{\boldsymbol{x}}, \lambda_1^*) = \{(0,d)^T \mid d \in \mathbb{R}\}$. Quindi abbiamo,

$$\boldsymbol{d}^T \nabla^2_{\boldsymbol{x},\boldsymbol{x}} L(\tilde{\boldsymbol{x}}, \lambda^*) \boldsymbol{d} = \begin{pmatrix} 0 \\ d \end{pmatrix}^T \begin{pmatrix} -6 & 0 \\ 0 & 10 \end{pmatrix} = \begin{pmatrix} 0 \\ d \end{pmatrix} = 10d^2 > 0.$$

Quindi nel punto $\tilde{\boldsymbol{x}}$ sono soddisfatte le condizioni sufficienti del secondo ordine ed esso è un punto di minimo locale in senso stretto.

9.4 Punti di sella e dualità

Richiamiamo qui brevemente alcuni risultati relativi alla teoria della dualità per problemi di programmazione non lineare. Considereremo solo il caso di problemi con soli vincoli di disuguaglianza, che nel seguito indicheremo come problema primale (P), e nel quale sia la funzione obiettivo che i vincoli sono funzioni convesse.

$$\begin{array}{lll} \min & f(\boldsymbol{x}) & \\ & g_i(\boldsymbol{x}) \leq 0 & i = 1, \ldots, k; \end{array} \tag{35}$$

Il relativo modello lagrangiano è:

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{j=1}^{k} \lambda_j g_j(\boldsymbol{x})$$

Innanzitutto introduciamo la nozione di punto di sella della funzione lagrangiana.

Definizione 18 *Chiamiamo punto di sella della Lagrangiana un punto* $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ *tale che:*

$$L(\boldsymbol{x}^*, \boldsymbol{\lambda}) \leq L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \leq L(\boldsymbol{x}, \boldsymbol{\lambda}^*) \quad \textit{per ogni } x \textit{ e per ogni } \boldsymbol{\lambda} \geq \mathbf{0}$$

In pratica in un punto di sella, fissato $\boldsymbol{\lambda}^*$, il punto $\boldsymbol{x}^*$ è un minimo della funzione $L(\boldsymbol{x}, \boldsymbol{\lambda}^*)$, mentre, fissato $\boldsymbol{x}^*$, il punto $\boldsymbol{\lambda}^*$ è un massimo della funzione $L(\boldsymbol{x}^*, \boldsymbol{\lambda})$.
Osserviamo ora che uno degli aspetti problematici delle condizioni KKT è che, da un lato sono generalmente solo condizioni necessarie, e dell'altro esse si applicano solo se le funzioni f, e g_j sono di classe C^1, sono cioè differenziabili con continuità. Tali richieste possono essere rimosse formulando le condizioni di KKT nei termini di proprietà di punto di sella della funzione lagrangiana.

Teorema 22 *Un punto* $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ *con* $\boldsymbol{\lambda}^* \geq \mathbf{0}$ *è un punto di sella della funzione lagrangiana del problema primale se e solo se valgono le seguenti condizioni:*

$$\begin{array}{lll} \boldsymbol{x}^* & = \arg\min L(\boldsymbol{x}, \boldsymbol{\lambda}^*) & \\ g_j(\boldsymbol{x}^*) & \leq 0 & j = 1, \ldots, k \\ \lambda_j g_j(\boldsymbol{x}^*) & = 0 & j = 1, \ldots, k \end{array}$$

Teorema 23 *Se il punto* $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ *con* $\boldsymbol{\lambda}^* \geq \mathbf{0}$ *è un punto di sella della funzione lagrangiana del problema primale allora* $\boldsymbol{x}^*$ *è la soluzione ottima del problema primale.*

Il principale vantaggio di questi due teoremi è che forniscono condizioni necessarie per problemi di ottimizzazione che non sono né convessi né differenziabili. Qualunque tecnica di ricerca può essere usata per minimizzare $L(\boldsymbol{x}, \boldsymbol{\lambda}^*)$ al variare di $\boldsymbol{x}$. Naturalmente rimane il problema di conoscere il vettore dei moltiplicatori ottimi $\boldsymbol{\lambda}^*$. Dal punto di vista pratico si possono ottenere stime di $\boldsymbol{\lambda}^*$ usando tecniche iterative o risolvendo il cosiddetto problema duale. A partire dalla funzione lagrangiana introduciamo ora il modello duale del problema primale (P). In primo luogo definiamo la funzione duale

$$h(\boldsymbol{\lambda}) = \min_{x} L(\boldsymbol{x}, \boldsymbol{\lambda})$$

Va notato che, per un dato vettore di moltiplicatori $\boldsymbol{\lambda}$ la soluzione ottima $\boldsymbol{x}^*(\boldsymbol{\lambda})$ del problema $\min_{x} L(\boldsymbol{x}, \boldsymbol{\lambda})$ non necessariamente soddisfa $\boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0}$, e addirittura un minimo $\boldsymbol{x}^*(\boldsymbol{\lambda})$ potrebbe non esistere per ogni valore di $\boldsymbol{\lambda}$.
Definiamo quindi l'insieme, D, dei vettori di moltiplicatori $\boldsymbol{\lambda}$ per i quali un minimo $\boldsymbol{x}^*(\boldsymbol{\lambda})$ esiste:

$$D = \{ \boldsymbol{\lambda} \mid \exists\, h(\boldsymbol{\lambda}),\ \boldsymbol{\lambda} \geqslant \mathbf{0} \}$$

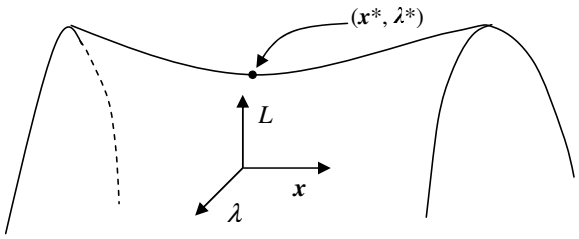


Figura 23: Punto di sella della funzione Lagrangiana

Definizione 19 *Il problema*

$$\max_{\boldsymbol{\lambda} \in D} h(\boldsymbol{\lambda}) = \max_{\boldsymbol{\lambda} \in D} \left\{ \min_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}) \right\}$$

è detto duale del problema (P).

In analogia a quanto visto nell'ambito della programmazione lineare, che, per inciso, diventa una caso particolare della teoria qui presentata, è possibile ricavare relazioni di dualità debole e forte fra il problema primale ed il suo duale.

Teorema 24 *La funzione duale $h(\boldsymbol{\lambda})$ soddisfa la relazione $h(\boldsymbol{\lambda}) \leq f(\boldsymbol{x})$ per tutti i punti $\boldsymbol{x}$ che soddisfano $\boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0}$, e per tutti i $\lambda \in D$.*

Dimostrazione

$$h(\boldsymbol{\lambda}) = \min_{x} L(\boldsymbol{x}, \boldsymbol{\lambda}), \quad \boldsymbol{\lambda} \in D \tag{36}$$

$$h(\boldsymbol{\lambda}) \leq \min_{x} L(\boldsymbol{x}, \boldsymbol{\lambda}), \quad \boldsymbol{\lambda} \in D, \quad \boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0} \tag{37}$$

$$h(\boldsymbol{\lambda}) \leq f(\boldsymbol{x}) + \sum_{j=1}^{k} \lambda_j g_j(\boldsymbol{x}), \quad \boldsymbol{\lambda} \in D, \quad \boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0} \tag{38}$$

$$h(\boldsymbol{\lambda}) \leq f(\boldsymbol{x}), \quad \boldsymbol{\lambda} \in D, \quad \boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0} \tag{39}$$

□

Quindi la funzione duale restituisce un limite inferiore al valore della soluzione ottima del problema primale. Il limite inferiore massimo corrisponde al valore della soluzione ottima del problema duale.

Teorema 25 *Il punto $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ con $\boldsymbol{\lambda}^* \geq \mathbf{0}$ è un punto di sella della funzione lagrangiana del problema primale se e solo se:*
- $\boldsymbol{x}^*$ è una soluzione del problema primale
- $\boldsymbol{\lambda}^*$ è una soluzione del problema duale
- $f(\boldsymbol{x}^*) = h(\boldsymbol{\lambda}^*)$.

La Figura 23 illustra visivamente le condizioni di ottimalità descritte nel teorema 25.
Le implicazioni algoritmiche del teorema di dualità sono che il problema primale può essere risolto eseguendo i seguenti passi:

1. Risolvi il problema di ottimizzazione non vincolato duale $\max_{\boldsymbol{\lambda} \in D} h(\boldsymbol{\lambda})$ per ricavare $\boldsymbol{\lambda}^* \geq \mathbf{0}$.

2. Noto $\boldsymbol{\lambda}^* \geq \mathbf{0}$, risolvi il problema primale $\min_x L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ per ricavare $\boldsymbol{x}^*(\boldsymbol{\lambda}^*)$.

3. Verifica se $\boldsymbol{x}^*, \boldsymbol{\lambda}^*$) soddisfa le condizioni di KKT.

9.5 Programmazione quadratica con vincoli di disuguaglianza lineari

Un problema con soli vincoli di disuguaglianza di estremo interesse è il caso di una funzione quadratica definita positiva, soggetta a vincoli lineari di disuguaglianza

$$\begin{array}{ll} \min & f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T Q \boldsymbol{x} - \boldsymbol{b}^T \boldsymbol{x} \\ t.c. & A\boldsymbol{x} \leq \boldsymbol{d} \end{array}$$

Rappresenta la generalizzazione della programmazione quadratica con vincoli lineari. La soluzione ottima, $\boldsymbol{x}^*$, potrebbe essere all'interno o sul confine della regione ammissibile. Se $\boldsymbol{x}^*$ si trova all'interno, allora nessun vincolo è attivo, e $\boldsymbol{x}^* = \boldsymbol{x}_0 = -Q^{-1}b$.
Altrimenti, almeno un vincolo lineare è attivo nella soluzione ottima. Se l'insieme di vincoli attivi in $\boldsymbol{x}^*$ è noto, allora il problema è enormemente semplificato. Supponiamo che tale insieme sia noto, allora posso rappresentare questo insieme *attivo* come: $A'\boldsymbol{x} = \boldsymbol{d}'$ e applicare la teoria di Lagrange (cfr. Sezione 9.1.1).
Nel risolvere problemi di programmazione quadratica, il lavoro più oneroso è rappresentato dall'identificazione dell'insieme dei vincoli attivi. Un metodo immediato, ma applicabile quando il numero di vincoli non è eccessivo, è quello di Theil e Van de Panne. Si tratta di identificare l'insieme dei vincoli attivi S partendo dall'insieme vuoto, $S = \emptyset$, risolvendo il problema nell'insieme S e verificando se la soluzione ottima viola qualche vincolo. Se non si violano vincoli il procedimento si arresta altrimenti si generano tutti gli insiemi $S_1, S_2, \ldots, S_r$ formati da un solo vincolo. Si risolvono gli r problemi risultanti e si verifica se esistono soluzioni che non violano alcun vincolo. In caso affermativo la migliore di esse è la soluzione ottima, altrimenti si generano tutti gli insiemi $S_1, S_2, \ldots, S_q$ formati da coppie, terne, ecc. di vincoli e si itera il processo.

E' importante richiamare che, come messo in evidenza nella Sezione 9.1.1, la tecnica restituisce non solo il valore della soluzione ottima $\boldsymbol{x}^*$ ma anche quella dei corrispondenti moltiplicatori.

In generale l'importanza dei problemi quadratici con vincoli di disuguaglianza lineari risiede nel fatto che esso risulta il sotto problema chiave da risolvere all'interno delle moderne tecniche di risoluzione dei problemi in forma generale (cfr. Sezione 9.10).

Poiché molto raramente è possibile utilizzare direttamente le condizioni di ottimalità per individuare in modo rapido un punto stazionario o, meglio ancora, un punto di minimo, sono stati introdotti numerosi algoritmi di tipo iterativo per approssimare le soluzioni del generico problema di ottimizzazione vincolata.

9.6 Metodi con funzione di penalità

Un primo tipo di approccio, si basa sull'idea di ricondurre la soluzione di un problema vincolato a quella di un problema non vincolato. Tale approccio è di tipo sequenziale, cioè è basato sulla soluzione di una successione di problemi non vincolati, in modo tale che le soluzioni ottime convergano a quella del problema vincolato. In tali problemi si fa uso di una funzione continua

$p(\boldsymbol{x})$, detta di penalità, tale che $p(\boldsymbol{x})$ è nulla nei punti $\boldsymbol{x}$ che rispettano i vincoli e maggiore di zero altrove. Consideriamo per primo un problema con soli vincoli di uguaglianza

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ h_j(\boldsymbol{x}) & = 0 \quad j = 1, \ldots, h; \end{array}$$

In tali modelli una tipica scelta per la funzione di penalità è

$$p(\boldsymbol{x}) = \sum_{j=1}^{h} h_j^2(\boldsymbol{x})$$

Tale scelta garantisce che la funzione di penalità regolare ed in particolare sia differenziabile nei punti in cui i vincoli sono soddisfatti. Il modello con funzione di penalità diviene

$$\min q(\boldsymbol{x}) = f(\boldsymbol{x}) + \alpha \sum_{j=1}^{h} h_j^2(\boldsymbol{x}).$$

Intuitivamente, maggiore è il valore del parametro α e maggiore è la probabilità che la soluzione ottima del modello penalizzato soddisfi i vincoli di uguaglianza. La tecnica prevede quindi di risolvere una successione di problemi penalizzati per valori crescenti di α, ottenendo così una successione di punti che convergono alla soluzione ottima del problema vincolato.
In questo caso, le condizioni necessarie del primo e del secondo ordine affinché $\boldsymbol{x}^*$ sia un punto di minimo del problema non vincolato sono

$$\nabla q(\boldsymbol{x}^*) = \nabla f(\boldsymbol{x}^*) + 2\alpha \sum_{j=1}^{h} h_j(\boldsymbol{x}^*) \nabla h_j(\boldsymbol{x}^*) = \mathbf{0},$$

e che la matrice hessiana di $q(\boldsymbol{x})$ in $\boldsymbol{x}^*$

$$\nabla^2 q(\boldsymbol{x}^*) = \nabla^2 f(\boldsymbol{x}^*) + 2\alpha \sum_{j=1}^{h} (h_j(\boldsymbol{x}^*) \nabla^2 h_j(\boldsymbol{x}^*) + \nabla h_j(\boldsymbol{x}^*) \nabla h_j(\boldsymbol{x}^*)^T)$$

sia semi definita positiva.
Si può dimostrare che facendo crescere α a infinito, la successione degli ottimi del problema penalizzato, $\boldsymbol{x}^*(\alpha)$, tende ad un minimo locale del problema vincolato, inoltre si ha

$$\lim_{\alpha \to \infty} 2\alpha h_j(\boldsymbol{x}^*(\alpha)) = \lambda_j^*$$

dove λ_j^* è il valore ottimo del moltiplicatore di Lagrange associato all'j-esimo vincolo.
Dalla condizione di ottimalità del secondo ordine si può osservare che la matrice hessiana della funzione obiettivo del problema penalizzato è data dalla somma di due parti. La prima parte è

$$\nabla^2 f(\boldsymbol{x}^*) + 2\alpha \sum_{j=1}^{h} h_j(\boldsymbol{x}^*) \nabla^2 h_j(\boldsymbol{x}^*)$$

che, per quanto appena detto, facendo crescere α a infinito, tende alla forma

$$\nabla^2 f(\boldsymbol{x}^*) + \sum_{j=1}^{h} \lambda_j^* \nabla^2 h_j(\boldsymbol{x}^*)$$

cioè alla matrice Hessiana della funzione Lagrangiana nel punto di ottimo.
La seconda parte è

$$\sum_{j=1}^{h} 2\alpha \nabla h_j(\boldsymbol{x}^*) \nabla h_j(\boldsymbol{x}^*)^T$$

che al tendere di α a infinito diventa illimitata in norma. La conseguenza di questo fatto è che, sebbene da un punto di vista teorico il metodo converga, da un punto di vista pratico l'Hessiana della funzione obiettivo penalizzata diviene sempre più malcondizionata man mano che ci si avvicina al punto ottimo $\boldsymbol{x}^*$.
Questa difficoltà può essere ovviata usando funzioni di penalità diverse, che non richiedano di far tendere α a infinito, ma in genere questo porta a perdere la differenziabilità della funzione $q(\boldsymbol{x})$, introducendo difficoltà di tipo diverso.

9.7 Metodi di barriera

Consideriamo ora un problema con soli vincoli di disuguaglianza

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ g_j(\boldsymbol{x}) & \leq 0 \quad j = 1, \ldots, k; \end{array}$$

Dividiamo la regione ammissibile del problema nell'insieme frontiera $S_f := \{\boldsymbol{x} \in \mathbb{R}^n \,|\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{0}\}$ e nell'insieme dei punti interni $S_{int} := \{\boldsymbol{x} \in \mathbb{R}^n \,|\boldsymbol{g}(\boldsymbol{x}) < \boldsymbol{0}\}$. I metodi a barriera sono applicabili nell'ipotesi che l'insieme S_{int} non sia vuoto. In tali metodi si fa uso di una *funzione barriera* $v(\boldsymbol{x})$ per l'insieme ammissibile del problema che è continua in S_{int}, e tale che $v(\boldsymbol{x}) \to \infty$ quando $\boldsymbol{x} \to S_f$. Il modello con funzione a barriera diviene quindi

$$\min\ b(\boldsymbol{x}) = f(\boldsymbol{x}) + \alpha v(\boldsymbol{x}).$$

E' un modello non vincolato dove si è creato un effetto barriera che impedisce a un punto che si trovi in S_{int} di uscire dalla regione ammissibile. L'effetto barriera cresce al crescere del parametro α.
Contrariamente al metodo che usa le funzioni di penalità, nel metodo a barriera si lavora con punti che sono sempre ammissibili. Si può dimostrare che sotto ipotesi abbastanza blande, per valori decrescenti del parametro α la successione delle soluzioni ottime dei problemi non vincolati converge a un minimo locale del problema vincolato.
La funzione barriera più usata è la funzione logaritmica

$$v(\boldsymbol{x}) = -\sum\nolimits_{i=1}^{k} \log(-g_j(\boldsymbol{x}))$$

Anche in questo caso il problema principale sta nel malcondizionamento della Hessiana della funzione $b(\boldsymbol{x})$ al decrescere di α. Un'ulteriore difficoltà è che questi metodi richiedono che il punto di partenza $\boldsymbol{x}_0$ sia ammissibile e questo non è semplice facile da ottenere.

Recentemente l'interesse per i metodi a barriera si è risvegliato grazie al fatto che si sono rivelati utili nella risoluzione di particolari problemi lineari.

9.8 Metodo del gradiente proiettivo

Questo metodo è dovuto a Rosen (1960, 1961). Consideriamo un problema con vincoli di uguaglianza lineari

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ t.c. & A\boldsymbol{x} = \boldsymbol{b} \end{array}$$

Si parte da una soluzione iniziale $\boldsymbol{x}'$ ammissibile, $A\boldsymbol{x}' = \boldsymbol{b}$, e si cerca $\boldsymbol{x} = \boldsymbol{x}' + \alpha\boldsymbol{d}$ (nuova soluzione migliorata). La direzione $\boldsymbol{d}$ deve soddisfare condizioni di normalizzazione, $||\boldsymbol{d}|| = 1$,

deve conservare l'ammissibilità, $A(\boldsymbol{x}' + \alpha\boldsymbol{d}) - \boldsymbol{b} = \boldsymbol{0}$, e soprattutto deve garantire il miglior decremento del valore della funzione obiettivo in $\boldsymbol{x}'$. La condizione di normalizzazione si può scrivere come $1 - \boldsymbol{d}^T\boldsymbol{d} = 0$, la condizione di ammissibilità si traduce in $\lambda A\boldsymbol{d} = \boldsymbol{0}$, cioè $A\boldsymbol{d} = \boldsymbol{0}$, mentre la richiesta di una direzione di massima discesa di $f(\boldsymbol{x})$ equivale a trovare una direzione $\boldsymbol{d}$ tale che sia minimizzata la derivata direzionale $\nabla f(\boldsymbol{x}')^T\boldsymbol{d}$ in $\boldsymbol{x}'$. Da queste condizioni ricaviamo un problema di ottimizzazione vincolata in cui le nuove variabili sono $\boldsymbol{d}$.

$$\begin{array}{ll} \min & \nabla f(\boldsymbol{x}')^T\boldsymbol{d} \\ t.c. & 1 - \boldsymbol{d}^T\boldsymbol{d} = 0 \\ & A\boldsymbol{d} = \boldsymbol{0} \end{array}$$

Applicando la tecnica lagrangiana per un problema soggetto a vincoli di uguaglianza ricaviamo:

$$L(\boldsymbol{d}, \boldsymbol{\lambda}, \lambda_0) = \nabla f(\boldsymbol{x}')^T\boldsymbol{d} + \boldsymbol{\lambda}^T A\boldsymbol{d} + \lambda_0(1 - \boldsymbol{d}^T\boldsymbol{d})$$

Le condizioni necessarie sono

$$\begin{aligned} \nabla_d L &= \nabla f(\boldsymbol{x}') + \boldsymbol{\lambda}^T A - 2\lambda_0\boldsymbol{d} = \boldsymbol{0} \\ \nabla_{\boldsymbol{\lambda}} L &= A\boldsymbol{d} = \boldsymbol{0} \\ \nabla_{\lambda_0} L &= (1 - \boldsymbol{d}^T\boldsymbol{d}) = 0 \end{aligned}$$

Dalla prima condizione ricaviamo:

$$\boldsymbol{d} = \frac{1}{2\lambda_0}(\nabla f(\boldsymbol{x}') + \boldsymbol{\lambda}^T A).$$

Sostituendo questo risultato nella terza condizione ricaviamo:

$$1 = \frac{1}{4\lambda_0^2}(\nabla f(\boldsymbol{x}') + \boldsymbol{\lambda}^T A)^T(\nabla f(\boldsymbol{x}') + \lambda^T A)$$

ed infine:

$$\lambda_0 = \pm\frac{1}{2}\left\|\nabla f(\boldsymbol{x}') + \lambda^T A\right\|$$

Sostituendo l'espressione per λ_0 nell'equazione che definisce la direzione $\boldsymbol{d}$ si ottiene

$$\boldsymbol{d} = \pm\frac{(\nabla f(\boldsymbol{x}') + \lambda^T A)}{\|\nabla f(\boldsymbol{x}') + \lambda^T A\|}$$

Fra i due versi $(+/-)$, si sceglie quello negativo che individua una direzione di decrescita di $\nabla f(\boldsymbol{x}')^T\boldsymbol{d}$. Questo garantisce inoltre che la matrice hessiana della funzione lagrangiana rispetto a $\boldsymbol{d}$, $\nabla^2_{\boldsymbol{d}} L = -2\lambda_0$, sia definita positiva, assicurandoci che la funzione di $\nabla f(\boldsymbol{x}')^T\boldsymbol{d}$ assume un valore minimo rispetto a $\boldsymbol{d}$.
Rimane da ricavare $\boldsymbol{\lambda}$. Dalla seconda condizione si ottiene:

$$A(\nabla f(\boldsymbol{x}') + \boldsymbol{\lambda}^T A) = \boldsymbol{0}$$

Così se $\boldsymbol{d} \neq \boldsymbol{0}$, si ricava:

$$AA^T\boldsymbol{\lambda} = -A\nabla f(\boldsymbol{x}')$$

con soluzione:

$$\boldsymbol{\lambda} = -(AA^T)^{-1}A\nabla f(\boldsymbol{x}')$$

La direzione $\boldsymbol{d}$, detta direzione del *gradiente proiettivo*, è così data da:

$$\boldsymbol{d} = -\frac{\left(I - A^T(AA^T)^{-1}A\right)\nabla f(\boldsymbol{x}')}{\|(I - A^T(AA^T)^{-1}A)\nabla f(\boldsymbol{x}')\|}$$

L'espressione trovata si può interpretare nel seguente modo. La direzione di massima decrescita per la funzione $f(\boldsymbol{x})$ in $\boldsymbol{x}'$ è la direzione dell'antigradiente, $-\nabla f(\boldsymbol{x}')$. La direzione trovata è la direzione dell'antigradiente **proiettata** nell'iperpiano $A\boldsymbol{x} = \boldsymbol{b}$. La matrice $P = \left(I - A^T(AA^T)^{-1}A\right)$ è detta perciò *matrice di proiezione*.

Nella pratica, $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha\boldsymbol{d}$, spesso si usa la direzione non normalizzata $\boldsymbol{d} = -P\nabla f(\boldsymbol{x}')$, e si determina il passo α con i metodi già visti in precedenza (es. Armijo, Fibonacci, Sezione Aurea, ecc.).

Consideriamo ora un problema con vincoli di uguaglianza non necessariamente lineari

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ & h_j(\boldsymbol{x}) = 0, \quad j = 1, \ldots, h \end{array}$$

Si linearizzano i vincoli nell'intorno della soluzione ammissibile corrente $\boldsymbol{x}'$, mediante lo sviluppo in serie di Taylor arrestato al primo ordine, ricavando

$$h_j(\boldsymbol{x}) = h_j(\boldsymbol{x}') + \nabla h_j(x')^T(\boldsymbol{x} - \boldsymbol{x}'),$$

che permette quindi la seguente approssimazione lineare (tenuto conto che $h_j(\boldsymbol{x}') = 0$) nell'intorno di $\boldsymbol{x}'$

$$\nabla h_j(\boldsymbol{x}')^T\boldsymbol{x} - \nabla h_j(\boldsymbol{x}')^T\boldsymbol{x}' = 0, \quad j = 1, \ldots, h.$$

Ora, ponendo $A = \left[\frac{\partial h(\boldsymbol{x}')}{\partial x}\right]^T$, e $\boldsymbol{b} = \left[\frac{\partial h(\boldsymbol{x}')}{\partial x}\right]^T \boldsymbol{x}'$, il problema linearizzato in $\boldsymbol{x}'$ diventa

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ t.c. & A\boldsymbol{x} = \boldsymbol{b} \end{array}$$

Si applica la tecnica della matrice di proiezione $P(\boldsymbol{x}') = \left(I - A^T(AA^T)^{-1}A\right)$, che ora però dipende da $\boldsymbol{x}'$ attraverso la matrice A, e si usa la direzione $\boldsymbol{d} = -P(\boldsymbol{x}')\nabla f(\boldsymbol{x}')$.
Poiché in generale, dato $\boldsymbol{x}_k = \boldsymbol{x}'$, per ogni scelta del passo $\alpha > 0$, il nuovo punto $\boldsymbol{x}'' = \boldsymbol{x}_k + \alpha\boldsymbol{d}$, non soddisfa necessariamente i vincoli, $\boldsymbol{h}(\boldsymbol{x}'') \neq \boldsymbol{0}$, occorre apportare un passo correttivo, $\boldsymbol{x}'' \to \boldsymbol{x}_{k+1}$. Tale passo è calcolato in modo che la sua proiezione in $\boldsymbol{x}_{k+1}$ sia nulla, cioè

$$P(\boldsymbol{x}_k)(\boldsymbol{x}_{k+1} - \boldsymbol{x}'') = \boldsymbol{0},$$

il che equivale a chiedere che la correzione sia ortogonale alla direzione di discesa calcolata in $\boldsymbol{x}_k$, e inoltre si vuole valga $\boldsymbol{h}(\boldsymbol{x}_{k+1}) = \boldsymbol{0}$.
Imponendo

$$\left(I - A^T(AA^T)^{-1}A\right)(\boldsymbol{x}_{k+1} - \boldsymbol{x}'') = \boldsymbol{0},$$

con A calcolata in $\boldsymbol{x}_k$, si ricava

$$\boldsymbol{x}_{k+1} - A^T(AA^T)^{-1}A\boldsymbol{x}_{k+1} = \boldsymbol{x}'' - A^T(AA^T)^{-1}A\boldsymbol{x}'',$$

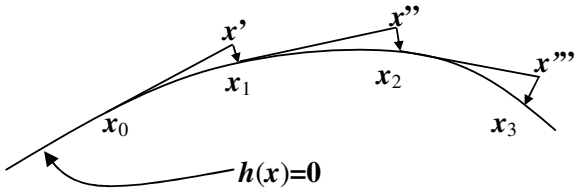


Figura 24: Passo di correzione nel gradiente proiettivo

cioè

$$\boldsymbol{x}_{k+1} - A^T(AA^T)^{-1}h(\boldsymbol{x}_{k+1}) \approx \boldsymbol{x}'' - A^T(AA^T)^{-1}h(\boldsymbol{x}''),$$

per l'approssimazione $h(\boldsymbol{x}) \approx A\boldsymbol{x} - \boldsymbol{b}$, che, imponendo $h(\boldsymbol{x}_{k+1}) = \mathbf{0}$, si riduce a

$$\boldsymbol{x}_{k+1} \approx \boldsymbol{x}'' - A^T(AA^T)^{-1}\boldsymbol{h}(\boldsymbol{x}'').$$

La Figura 24 illustra la dinamica dell'aggiornamento mediante il passo di correzione. Il passo di correzione viene ripetuto fino a quando $\boldsymbol{h}(\boldsymbol{x}_{k+1})$ è sufficientemente piccolo, mentre l'intero algoritmo viene fatto arrestare quando $P(\boldsymbol{x}')\nabla f(\boldsymbol{x}') \approx \mathbf{0}$.

9.9 Metodo dei lagrangiani aumentati

Questo metodo combina il classico metodo della funzione lagrangiana (cfr. Sezione 9.2) con la tecnica della funzione di penalità (cfr. Sezione 9.6). Come abbiamo visto nell'approccio lagrangiano il punto di minimo del problema vincolato coincide con un punto stazionario $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ della funzione lagrangiana che, in generale, è difficile trovare analiticamente. D'altro canto, nell'approccio con funzioni di penalità il minimo della funzione di penalità approssima il minimo vincolato ma al crescere dell'accuratezza richiesta cresce il malcondizionamento della matrice hessiana della funzione di penalità.
Nel *metodo dei moltiplicatori* (Bertsekas 1976) i due approcci vengono combinati per dare un problema non vincolato e non malcondizionato. L'idea è quella di approssimare i moltiplicatori di Lagrange.
Consideriamo il caso di un problema con soli vincoli di uguaglianza.

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ & h_j(\boldsymbol{x}) = 0, \quad j = 1, \ldots, h \end{aligned}$$

e introduciamo la corrispondente funzione dei **langrangiani aumentati**:

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \rho) = f(\boldsymbol{x}) + \sum_{j=1}^{h} \lambda_j h_j(\boldsymbol{x}) + \rho \sum_{j=1}^{h} h_j^2(\boldsymbol{x})$$

Con tutti i moltiplicatori $\lambda_j = 0$ la funzione $\mathcal{L}$ si riduce alla funzione di penalità, mentre se sono noti i valori ottimi λ_j^* è possibile dimostrare (Fletcher 1987) che per ogni valore $\rho > 0$ la minimizzazione di $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \rho)$ rispetto a $\boldsymbol{x}$ fornisce la soluzione ottima $\boldsymbol{x}^*$ del problema.
Se il vettore $\boldsymbol{\lambda}^k$ è una buona approssimazione di $\boldsymbol{\lambda}^*$, allora è possibile approssimare l'ottimo attraverso la minimizzazione non vincolata di $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^k, \rho)$ senza richiedere valori di ρ eccessivamente grandi.
Il valore di ρ deve essere abbastanza grande da garantire che la funzione $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^k, \rho)$ abbia un minimo locale rispetto a $\boldsymbol{x}$ e non solo un punto di stazionarietà.

Per comprendere la tecnica dei lagrangiani aumentati è necessario confrontare le condizioni di stazionarietà di L ed $\mathcal{L}$ in $\boldsymbol{x}^*$.
Per $\mathcal{L}$:

$$\frac{\partial \mathcal{L}}{\partial x_i} = \frac{\partial f}{\partial x_i} + \sum_{j=1}^{h} (\lambda_j^k + 2\rho h_j) \frac{\partial h_j}{\partial x_i} = 0, \quad i = 1, \ldots, n.$$

Per L :

$$\frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i} + \sum_{j=1}^{h} \lambda_j^k \frac{\partial h_j}{\partial x_i} = 0, \quad i = 1, \ldots, n.$$

Il confronto indica che al tendere del punto di minimo di $\mathcal{L}$ a $\boldsymbol{x}^*$, allora:

$$\lambda_j^k + 2\rho h_j \to \lambda_j^*$$

Questa osservazione ha suggerito il seguente schema per approssimare $\boldsymbol{\lambda}^*$. Dato un vettore di moltiplicatori approssimato $\boldsymbol{\lambda}^k$, con $k = 1, 2, \ldots$, e un valore $\rho > 0$, si risolve $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^k, \rho)$ in funzione di $\boldsymbol{x}$, con un qualsiasi approccio iterativo per l'ottimizzazione non vincolata, ricavando $\boldsymbol{x}_k^*$. I valori della nuova stima di $\boldsymbol{\lambda}^*$ sono calcolati come

$$\lambda_j^{k+1} := \lambda_j^k + 2\rho h_j(\boldsymbol{x}_k^*)$$

In letteratura sono stati proposti anche schemi per variare iterativamente anche il parametro ρ.

9.10 SQP (Sequential Quadratic Programming)

Il metodo SQP si basa sull'applicazione del metodo di Newton per determinare il punto $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ dalle condizioni di KKT del problema di ottimizzazione vincolato. Si può dimostrare (Bazaraa et al.1993) che la determinazione del passo di Newton è equivalente alla soluzione di un problema di programmazione quadratica. Consideriamo il problema generale

$$\begin{array}{lll} \min & f(\boldsymbol{x}) & \\ g_i(\boldsymbol{x}) & \leq 0 & i = 1, \ldots, k; \\ h_j(\boldsymbol{x}) & = 0 & j = 1, \ldots, h \end{array} \tag{40}$$

ed il relativo modello lagrangiano:

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \sum_{j=1}^{k} \lambda_j g_j(\boldsymbol{x}) + \sum_{j=1}^{h} \mu_j h_j(\boldsymbol{x})$$

E' data una stima $(\boldsymbol{x}_k, \boldsymbol{\lambda}_k, \boldsymbol{\mu}_k)$, con $\boldsymbol{\lambda}_k \geq 0$, $k = 1, 2, \ldots$, del valore della soluzione e dei relativi moltiplicatori lagrangiani ottimi e si è ricavata la matrice hessiana della funzione lagrangiana

$$\nabla^2 L(\boldsymbol{x}_k) = H(\boldsymbol{x}_k) + \sum_{j=1}^{k} \lambda_j^k \nabla^2 g_j(\boldsymbol{x}_k) + \sum_{j=1}^{h} \mu_j^k \nabla^2 h_j(\boldsymbol{x}_k).$$

Si può dimostrare che il valore del passo di Newton $\boldsymbol{d}$ che fornisce $\boldsymbol{x}_{k+1}$ a partire da $\boldsymbol{x}_k$,

$$\boldsymbol{x}_{k+1} := \boldsymbol{x}_k + \boldsymbol{d}_k$$

è dato dalla soluzione del seguente problema quadratico con vincoli lineari di uguaglianza e disuguaglianza:

$$\begin{aligned} \min\ \phi(\boldsymbol{d}) &= f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)^T \boldsymbol{d} + \frac{1}{2}\boldsymbol{d}^T \nabla^2 L(\boldsymbol{x}_k)\boldsymbol{d} \\ \boldsymbol{g}(\boldsymbol{x}_k) &+ \left[\frac{\partial \boldsymbol{g}(\boldsymbol{x}_k)}{\partial \boldsymbol{x}}\right]^T \boldsymbol{d} \leqslant \boldsymbol{0}, \\ \boldsymbol{h}(\boldsymbol{x}_k) &+ \left[\frac{\partial \boldsymbol{h}(\boldsymbol{x}_k)}{\partial \boldsymbol{x}}\right]^T \boldsymbol{d} = \boldsymbol{0} \end{aligned}$$

Si osservi cha la soluzione del problema quadratico non solo restituisce $\boldsymbol{d}$ e quindi $\boldsymbol{x}_{k+1}$, ma anche i vettori di moltiplicatori lagrangiani $\boldsymbol{\lambda}_{k+1}$ e $\boldsymbol{\mu}_{k+1}$ (cfr. Sezione 9.5). In questo modo è possibile immediatamente formulare il modello quadratico alla successiva iterazione.
La condizione di arresto è l'approssimarsi di $\boldsymbol{d}$ al vettore nullo.

In pratica il modello può non convergere se si parte da punti lontani dai minimi. Di conseguenza si preferisce adottare un modello del tipo

$$\boldsymbol{x}_{k+1} := \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$$

nel quale il valore del passo α_k si ricava minimizzando una cosiddetta *funzione di merito* mediante tecniche di ricerca monodimensionale.
Una funzione di merito frequentemente adottata è

$$f_m(\boldsymbol{x}_k) = f(\boldsymbol{x}_k) + \gamma \left(\sum_{j=1}^{k} \max\{0, g_j(\boldsymbol{x}_k)\} + \sum_{j=1}^{h} |h_j(\boldsymbol{x}_k)| \right)$$

dove $\gamma = \max \lambda_1, \lambda_2, \ldots, \lambda_k, |\mu_1|, |\mu_2|, \ldots, |\mu_h|$. Poiché tale funzione non è differenziabile si adottano come tecniche di ottimizzazione monodimensionale metodi stabili quali la Sezione Aurea o Fibonacci.

10 Appendice

Definizione 20 *Dato un punto $\boldsymbol{x} \in \mathbb{R}^n$ e una funzione $f(\boldsymbol{x}) : X \to \mathbb{R}^n$ di classe $C^1(\boldsymbol{x})$, definiamo vettore gradiente di $f(\boldsymbol{x})$ in $\boldsymbol{x}$ (o semplicemente gradiente), il vettore:*

$$\nabla f(\boldsymbol{x}) = \left[\begin{array}{cccc} \frac{\partial f(\boldsymbol{x})}{\partial x_1} & \frac{\partial f(\boldsymbol{x})}{\partial x_2} & \dots & \frac{\partial f(\boldsymbol{x})}{\partial x_n} \end{array} \right]^T$$

Definizione 21 *Dato un punto $\boldsymbol{x} \in \mathbb{R}^n$ e una funzione $f(\boldsymbol{x}) : X \to \mathbb{R}^n$ di classe $C^2(\boldsymbol{x})$, definiamo matrice Hessiana di $f(\boldsymbol{x})$ in $\boldsymbol{x}$, la matrice:*

$$\nabla^2 f(\boldsymbol{x}) = H(\boldsymbol{x}) = \left[\begin{array}{cccc} \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1^2} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2^2} & \dots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_n} \\ & \vdots & & \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n^2} \end{array} \right]^T$$

Definizione 22 *Una funzione $f(\boldsymbol{x}) : X \to \mathbb{R}^n$, con $X \subseteq \mathbb{R}^n$, si dice lipschitziana con costante di Lipschitz L se $|f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq L||\boldsymbol{x} - \boldsymbol{y}||, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in X$*

Definizione 23 *Chiamiamo curva di livello di valore C, di una funzione $f(\boldsymbol{x}) : \mathbb{R} \to \mathbb{R}^n$, il luogo dei punti*

$$\{\boldsymbol{x} : f(\boldsymbol{x}) = C\}.$$

Proprietà 21 *Dato un punto $\boldsymbol{x} \in \mathbb{R}^n$ e una funzione $f(\boldsymbol{x}) : \mathbb{R} \to \mathbb{R}^n$ di classe $C^1(\boldsymbol{x})$ il vettore gradiente calcolato in $\boldsymbol{x}$ è ortogonale al piano tangente in $\boldsymbol{x}$ alla curva di livello passante per $\boldsymbol{x}$.*

Definizione 24 *Data una funzione $f : \mathrm{X} \to \mathbb{R}$ con $\mathrm{X} \subseteq \mathbb{R}^n$, un punto $\boldsymbol{x}^* = [\boldsymbol{y}_0, \boldsymbol{z}_0]^T \in \mathrm{X}$ è un punto di sella della funzione f se esiste $\varepsilon > 0$ tale che $\forall \boldsymbol{y} \in I(\boldsymbol{y}_0, \varepsilon)$ e $\forall \boldsymbol{z} \in I(\boldsymbol{z}_0, \varepsilon)$ vale la relazione*

$$f(\boldsymbol{y}, \boldsymbol{z}_0) \leq f(\boldsymbol{y}_0, \boldsymbol{z}_0) \leq f(\boldsymbol{y}_0, \boldsymbol{z})$$

Lemma 1 *Se l'algoritmo del gradiente viene applicato ad un problema di programmazione quadratica convesso la forma esatta della riduzione dell'errore di approssimazione tra iterazioni successive misurata con la norma pesata $|| \cdot ||_Q$ è*

$$\frac{\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|_Q}{\|\boldsymbol{x}_k - \boldsymbol{x}^*\|_Q} = \left\{ 1 - \frac{((Q\boldsymbol{x}_k - \boldsymbol{b})^T (Q\boldsymbol{x}_k - \boldsymbol{b}))^2}{((Q\boldsymbol{x}_k - \boldsymbol{b})^T Q (Q\boldsymbol{x}_k - \boldsymbol{b}))((Q\boldsymbol{x}_k - \boldsymbol{b})^T Q^{-1} (Q\boldsymbol{x}_k - \boldsymbol{b}))} \right\}^{\frac{1}{2}}$$

Dimostrazione La riduzione dell'errore di approssimazione tra iterazioni successive del metodo del gradiente applicato a funzioni quadratiche può essere riscritta nel seguente modo

$$\frac{E_{k+1}}{E_k} = \left\{ 1 - \frac{(\boldsymbol{g}_k^T \boldsymbol{g}_k)^2}{(\boldsymbol{g}_k^T Q \boldsymbol{g}_k)(\boldsymbol{g}_k^T Q^{-1} \boldsymbol{g}_k)} \right\}$$

dove abbiamo adottato la seguente notazione:

$$\boldsymbol{g}_k = \nabla f(\boldsymbol{x}_k), \quad E_k = ||\boldsymbol{x}_k - \boldsymbol{x}^*||_Q^2, \quad E_{k+1} = ||\boldsymbol{x}_{k+1} - \boldsymbol{x}^*||_Q^2$$

e abbiamo utilizzato la relazione $\boldsymbol{g}_k = Q\boldsymbol{x}_k - \boldsymbol{b} = Q\boldsymbol{x}_k - Q\boldsymbol{x}^* = Q(\boldsymbol{x}_k - \boldsymbol{x}^*)$. Iniziamo a calcolare

$$E_{k+1} = \frac{1}{2}(\boldsymbol{x}_k - \alpha\boldsymbol{g}_k - \boldsymbol{x}^*)^T Q(\boldsymbol{x}_k - \alpha\boldsymbol{g}_k - \boldsymbol{x}^*) = E_k + \frac{1}{2}\alpha^2 \boldsymbol{g}_k^T Q\boldsymbol{g}_k - \alpha\boldsymbol{g}_k^T(Q\boldsymbol{x}_k - Q\boldsymbol{x}^*)$$

Sostituendo ad α il valore del passo ottimo, $\frac{\boldsymbol{g}_k^T\boldsymbol{g}_k}{\boldsymbol{g}_k^T Q\boldsymbol{g}_k}$ si ricava

$$E_{k+1} = E_k + \frac{1}{2}\frac{(\boldsymbol{g}_k^T\boldsymbol{g}_k)^2}{(\boldsymbol{g}_k^T Q\boldsymbol{g}_k)^2}\boldsymbol{g}_k^T Q\boldsymbol{g}_k - \frac{\boldsymbol{g}_k^T\boldsymbol{g}_k}{\boldsymbol{g}_k^T Q\boldsymbol{g}_k}\boldsymbol{g}_k^T\boldsymbol{g}_k = E_k - \frac{1}{2}\frac{(\boldsymbol{g}_k^T\boldsymbol{g}_k)^2}{\boldsymbol{g}_k^T Q\boldsymbol{g}_k}.$$

Ricaviamo ora

$$E_k = \frac{1}{2}(\boldsymbol{x}_k - \boldsymbol{x}^*)^T Q(\boldsymbol{x}_k - \boldsymbol{x}^*) = \frac{1}{2}\boldsymbol{g}_k^T Q^{-1}\boldsymbol{g}_k$$

dove abbiamo di nuovo fatto uso di $\boldsymbol{g}_k = Q(\boldsymbol{x}_k - \boldsymbol{x}^*)$ e dell'identità $QQ^{-1} = I$. Possiamo ora raccogliere E_k a fattor comune

$$E_{k+1} = E_k - \frac{1}{2}\frac{(\boldsymbol{g}_k^T\boldsymbol{g}_k)^2}{\boldsymbol{g}_k^T Q\boldsymbol{g}_k}\frac{E_k}{E_k} = E_k\left\{1 - \frac{(\boldsymbol{g}_k^T\boldsymbol{g}_k)^2}{\boldsymbol{g}_k^T Q\boldsymbol{g}_k \boldsymbol{g}_k^T Q^{-1}\boldsymbol{g}_k}\right\}$$

da cui segue l'espressione cercata. □

Definizione 25 *Una matrice T è ortogonale se valgono le relazioni*

$$TT^T = T^T T = I.$$

In altre parole, l'inversa di una matrice ortogonale è la sua trasposta. Vale inoltre $\det T = \det T^T = \pm 1$.

Proprietà 22 *Sia data una matrice Q simmetrica e definita positiva. Q può essere riscritta mediante la seguente* decomposizione spettrale

$$Q = \sum_{i=1}^{n} \lambda_i t_i t_i^T,$$

dove $\lambda_1, \lambda_2, \ldots, \lambda_n$ sono gli n autovalori di Q e $t_1, t_2, \ldots, t_n$ sono i corrispondenti n autovettori. In termini matriciali tale decoposizione può essere riscritta come

$$Q = T\Lambda T^T$$

dove

$$\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_n), \quad T = [t_1|t_2|\ldots|t_n]$$

e la matrice T è una matrice ortogonale. In altre parole, la matrice Q può essere diagonalizzata nel sequente modo

$$T^{-1}QT = \Lambda.$$

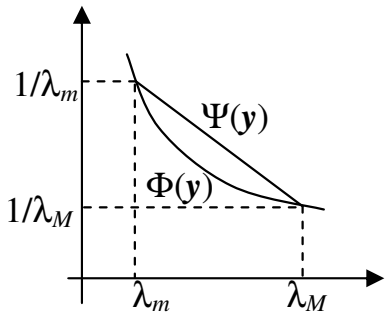


Figura 25: Andamento dei grafici di $\Phi(\boldsymbol{y})$ e $\Psi(\boldsymbol{y})$.

Teorema 9 *Data una matrice Q definita positiva vale la seguente relazione*

$$\frac{(\boldsymbol{x}^T\boldsymbol{x})^2}{(\boldsymbol{x}^TQ\boldsymbol{x})(\boldsymbol{x}^TQ^{-1}\boldsymbol{x})} \geq \frac{4\lambda_m\lambda_M}{(\lambda_m+\lambda_M)^2}$$

valida per ogni $\boldsymbol{x} \in \mathbb{R}^n$ e dove λ_M e $\lambda_m > 0$ sono l'autovalore massimo e minimo, rispettivamente, di Q.

Dimostrazione Si consideri la *decomposizione spettrale* di Q, $Q = T\Lambda T^T$ e si effettui la trasformazione $T^{-1}\boldsymbol{x} = \boldsymbol{z}$, cioè $\boldsymbol{x} = T\boldsymbol{z}$. Possiamo riscrivere

$$\frac{(\boldsymbol{x}^T\boldsymbol{x})^2}{(\boldsymbol{x}^TQ\boldsymbol{x})(\boldsymbol{x}^TQ^{-1}\boldsymbol{x})} = \frac{(\boldsymbol{z}^TT^TT\boldsymbol{z})^2}{(\boldsymbol{z}^TT^{-1}QT\boldsymbol{z})(\boldsymbol{z}^TT^{-1}Q^{-1}T\boldsymbol{z})}$$

che, sfruttando le proprietà delle matrici ortogonali si riduce a

$$\frac{(\boldsymbol{z}^T\boldsymbol{z})^2}{(\boldsymbol{z}^T\Lambda\boldsymbol{z})(\boldsymbol{z}^T\Lambda^{-1}\boldsymbol{z})} = \frac{(\sum_{i=1}^n z_i^2)^2}{(\sum_{i=1}^n \lambda_i z_i^2)(\sum_{i=1}^n z_i^2/\lambda_i)}.$$

Dividendo per $(\sum_{i=1}^n z_i^2)^2$ otteniamo

$$\frac{1}{(\sum_{i=1}^n \lambda_i(z_i^2/\sum_{i=1}^n z_i^2))(\sum_{i=1}^n(z_i^2/\sum_{i=1}^n z_i^2)/\lambda_i)}.$$

Effettuiamo ora un cambiamento di variabili definendo $y_i = \frac{z_i^2}{\sum_{i=1}^n z_i^2}$, con $0 \leq y_i \leq 1$ e $\sum_{i=1}^n y_i = 1$, e riscriviamo l'espressione come

$$\frac{1}{(\sum_{i=1}^n \lambda_i y_i)(\sum_{i=1}^n y_i/\lambda_i)} = \frac{1/\sum_{i=1}^n \lambda_i y_i}{\sum_{i=1}^n y_i/\lambda_i} = \frac{\Phi(\boldsymbol{y})}{\Psi(\boldsymbol{y})}.$$

Le funzioni $\Phi(\boldsymbol{y})$ e $\Psi(\boldsymbol{y})$ sono ottenute da combinazioni lineari di λ_i e $1/\lambda_i$, e ogni λ_i e $1/\lambda_i$ può essere generato come combinazione lineare convessa dei valori y_i. In Figura 25 si vede l'andamento delle funzioni $\Phi(\boldsymbol{y})$ e $\Psi(\boldsymbol{y})$ per fissati valori degli autovalori λ_i. La funzione $\Phi(\boldsymbol{y})$ può essere minorata dall'iperbole passante per i punti $(\lambda_m, 1/\lambda_m)$ e $(\lambda_M, 1/\lambda_M)$, mentre la funzione $\Psi(\boldsymbol{y})$ può essere maggiorata dal segmento di retta compreso fra gli stessi punti. Possiamo quindi scrivere

$$\frac{\Phi(\boldsymbol{y})}{\Psi(\boldsymbol{y})} \geq \min_{\lambda_m \leq \lambda \leq \lambda_M} F(\lambda) = \min_{\lambda_m \leq \lambda \leq \lambda_M} \frac{1/\lambda}{(\lambda_m+\lambda_M-\lambda)/\lambda_m\lambda_M}.$$

Derivando ed uguagliando a zero $F(\lambda)$ si perviene all'espressione

$$\frac{1}{\lambda_m \lambda_M \lambda} = \frac{\lambda_m + \lambda_M - \lambda}{\lambda_m \lambda_M \lambda^2} \quad \text{da cui si ricava} \quad \lambda^* = \frac{\lambda_m + \lambda_M}{2},$$

che sostituito in $F(\lambda)$ ci dà $F(\lambda^*) = \frac{4\lambda_m \lambda_M}{(\lambda_m + \lambda_M)^2}$, il che dimostra la tesi. □

Riferimenti bibliografici

[1] Francesco Maffioli. *Elementi di Programmazione matematica.* Volume secondo. Ed. Masson, 1991.

[2] Jan A. Snyman. *Practical Mathematical Optimization.* Ed. Springer, 2005.

[3] J. Nocedal, S.J. Wright. *Numerical Optimization.* Ed. Springer, 2006.