



UNIVERSITÀ DI SIENA 1240

RICERCA OPERATIVA

Appunti delle Lezioni

Corso di Laurea in
INGEGNERIA INFORMATICA E DELL'INFORMAZIONE

No part of this publication may be reproduced or used in any manner without the prior written permission of the owner. To request permissions, contact the publisher at info@sbuch.it

Indice

1	Introduzione	7
1.1	Ottimizzazione di una funzione	7
1.2	Punto di minimo	8
1.2.1	Punto di minimo globale	8
1.2.2	Punto di minimo globale stretto	8
1.2.3	Punto di minimo locale	8
1.2.4	Punto di minimo locale stretto	8
1.3	Osservazioni	8
2	Richiami di Analisi e Algebra	9
2.1	Derivata direzionale	9
2.1.1	Derivata direzionale di una funzione monodimensionale	9
2.1.2	Definizione di derivata direzionale	9
2.2	Gradiente	10
2.3	Matrice Hessiana	10
2.4	Formula di Taylor	10
2.4.1	Formula di Taylor di una funzione in una variabile	10
2.4.2	Formula di Taylor di una funzione in due variabili	10
2.4.3	Formula di Taylor di una funzione in più variabili	11
2.5	Teorema 1	11
2.5.1	Dimostrazione	11
2.6	Convessità	12
2.6.1	Combinazione convessa di due punti	12
2.6.2	Insieme convesso	12
2.6.3	Funzione convessa	12
2.6.4	Funzione concava	13
2.6.5	Teorema 2	13
2.6.5.1	Dimostrazione	13
2.6.6	Teorema 3	14
2.6.6.1	Dimostrazione	14
2.7	Matrici definite/semidefinite positive o negative	15
2.7.1	Matrice definita positiva	15
2.7.2	Matrice semidefinita positiva	15
2.7.3	Matrice definita negativa	15
2.7.4	Matrice definita negativa	15
2.7.5	Teoremi di Sylvester	15
2.7.5.1	Teorema 4	15
2.7.5.2	Teorema 5	15

2.7.6	Teorema 6	16
2.7.7	Teorema 7	16
2.7.7.1	Dimostrazione	16
2.7.7.2	Caso particolare	17
2.8	Esistenza di un minimo	17
2.8.1	Insieme chiuso e limitato	17
2.8.2	Teorema 8	17
2.8.3	Insieme di livello	18
2.8.4	Superficie di livello	18
2.8.5	Teorema 9	18
2.9	Angolo tra due vettori e direzioni di discesa	18
2.9.1	Prodotto scalare	18
2.9.2	Angolo tra due vettori	19
2.9.3	Direzione di discesa	19
2.9.4	Direzione di salita	19
2.9.5	Teorema 10	19
2.9.5.1	Dimostrazione	20
2.9.6	Teorema 11	20
2.9.6.1	Dimostrazione	20
2.9.7	Teorema 12	20
2.9.7.1	Dimostrazione	21
2.9.8	Osservazioni	21
3	Ottimizzazione Non Vincolata	22
3.1	Caratterizzazione dei punti di minimo	22
3.1.1	Teorema 13	22
3.1.1.1	Dimostrazione	22
3.1.2	Teorema 14	22
3.1.2.1	Dimostrazione	23
3.1.3	Teorema 15	23
3.1.3.1	Dimostrazione	24
3.2	Caratterizzazione dei punti di massimo	24
3.2.1	Teorema 16	24
3.2.2	Teorema 17	25
3.2.3	Teorema 18	25
3.2.4	Punto di sella	25
3.3	Caso convesso	25
3.3.1	Teorema 19	25
3.3.1.1	Dimostrazione	25
3.3.1.2	Osservazioni	26
3.3.2	Teorema 20	26
3.3.2.1	Dimostrazione	27
3.3.2.2	Osservazione	27
3.3.3	Teorema 21	27
3.3.3.1	Esempio	27
3.3.3.2	Dimostrazione	27
3.3.4	Osservazioni	28
3.4	Esercizio 1	28

3.4.1	Soluzione	28
3.4.1.1	Condizioni del primo ordine	28
3.4.1.2	Condizioni del secondo ordine	29
3.4.1.3	Esistenza di un minimo globale	29
3.5	Esercizio 2	30
3.5.1	Soluzione	30
4	Ottimizzazione Vincolata	31
4.1	Introduzione	31
4.1.1	Esempio	31
4.2	Condizioni di ottimalità	32
4.2.1	Esempio	32
4.2.2	Vincoli attivi	32
4.2.3	Matrice jacobiana	33
4.3	Un vincolo di uguaglianza	33
4.3.1	Teorema 22	34
4.3.1.1	Dimostrazione	34
4.3.1.2	Interpretazione geometrica in due dimensioni	35
4.3.2	Esempio	36
4.4	Un vincolo di disuguaglianza	38
4.4.1	Caso 1	39
4.4.2	Caso 2	39
4.5	Più vincoli di disuguaglianza	42
4.5.1	Definizione di cono	42
4.5.2	Definizione di punto regolare	42
4.5.3	Due vincoli di disuguaglianza	43
4.5.3.1	Caso 1	43
4.5.3.2	Caso 2	43
4.5.3.3	Caso 3	43
4.5.3.4	Caso 4	43
4.6	Le condizioni di Karush-Kuhn-Tucker	46
4.6.1	Funzione lagrangiana di un problema di PNL	47
4.6.2	Teorema 23	47
4.6.3	Caso particolare di complementarità	48
4.6.4	Condizioni necessarie del secondo ordine (solo equazioni)	48
4.6.5	Teorema 24	48
4.6.6	Punti candidati ad essere minimi locali	48
4.7	Sensibilità alla variazione dei parametri	49
4.8	Cenni sugli algoritmi di ottimizzazione vincolata	50
4.8.1	Funzioni di penalità quadratiche	51
4.8.2	Metodi di barriera	52
5	Programmazione Lineare	54
5.1	Introduzione	54
5.1.1	Teorema 25	54
5.1.1.1	Dimostrazione	54
5.2	Problema primale in forma standard	56
5.2.1	Funzione lagrangiana	56

5.2.2	Condizioni KKT del problema primale	57
5.3	Problema duale	58
5.3.1	Funzione lagrangiana	58
5.3.2	Condizioni KKT del problema duale	58
5.4	Condizioni KKT del problema primale e duale	59
5.5	Dualità debole	60
5.5.1	Dimostrazione	60
5.6	Dualità forte	61
5.6.1	Dimostrazione 1	61
5.6.1.1	Minimo globale per il problema primale	61
5.6.1.2	Minimo globale per il problema duale	61
5.6.2	Dimostrazione 2	61
5.7	Soluzioni ottime per il primale e per il duale	62
5.7.1	Teorema 26	62
5.7.1.1	Dimostrazione	62
5.8	Problema primale in forma canonica	63
5.8.1	Condizioni KKT del problema primale	63
5.8.2	Condizioni KKT del problema duale	63
5.9	Problema duale associato a un problema primale	64
5.9.1	Esempio	65
5.10	Trasformazione dei vincoli	65
5.11	Problema di allocazione di risorse	66
5.11.1	Esempio	67
5.12	Problema di miscelazione	68
5.12.1	Esempio	69
5.13	Problema di trasposto	70
6	Programmazione Lineare con Vertici	71
6.1	Introduzione	71
6.1.1	Poliedro	71
6.1.2	Politopo	72
6.1.3	Vertice o punto estremo di un poliedro	72
6.2	Teorema di Minkowski-Weyl	72
6.2.1	Dimostrazione	72
6.3	Teorema 27	73
6.3.1	Dimostrazione	73
6.4	Sistema ammette un'unica soluzione	73
6.5	Base di una matrice	74
6.6	Soluzione di base e soluzione di base ammissibile	74
6.7	Teorema 28	75
6.7.1	Dimostrazione $\Leftarrow$	75
6.7.2	Dimostrazione $\Rightarrow$	77
6.8	Riassumendo	78
7	Metodo del Simplexso	79
7.1	Introduzione	79
7.2	Test di ottimalità per una SBA	80
7.2.1	Teorema 29	80

	7.2.1.1	Dimostrazione	81
7.3		Operazione di Pivot	81
	7.3.1	Caso 1	83
		7.3.1.1 Caso particolare	83
		7.3.1.2 Caso particolare: valore massimo ϑ	83
	7.3.2	Caso 2	84
	7.3.3	Teorema 30	84
7.4		Base ammissibile per il metodo del simplesso	84
	7.4.1	Metodo del Big-M	84
	7.4.2	Metodo delle 2 fasi	85
		7.4.2.1 Fase 1	85
		7.4.2.2 Fase 2	86
	7.4.3	Basi e SBA degeneri	87
	7.4.4	Esercizio	87
	7.4.5	Regola di Bland (regola anticiclo)	90
7.5		Metodo del simplesso e teoria della dualità	90
7.6		Analisi di sensibilità della soluzione ottima	91
	7.6.1	alla variazione del vettore dei termini noti	91
	7.6.2	alla variazione del vettore dei costi	92

Capitolo 1

Introduzione

1.1 Ottimizzazione di una funzione

Studiamo i concetti introduttivi relativi all'ottimizzazione di una funzione f di n variabili su un insieme X . La funzione f prende il nome di funzione obiettivo e l'insieme X si chiama insieme o regione ammissibile. Ciascun punto

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in X$$

costituisce una soluzione ammissibile. Il problema di Programmazione Lineare (PL) consiste nel determinare un punto x^* appartenente all'insieme X tale da rendere minima la funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}$ quindi il problema è definito come

$$\begin{cases} \min f(x) \\ x \in X \end{cases}$$

Segue che

- se $X \equiv \emptyset \implies$ Problema è non ammissibile o inammissibile
- se $\forall M > 0, \exists \bar{x} \in X : f(\bar{x}) < -M \implies$ Problema è illimitato
- se $X \subseteq \mathbb{R}^n \implies$ Problema di ottimizzazione continuo
- se $X \subseteq \mathbb{Z}^n \implies$ Problema di ottimizzazione discreto
- se $X \equiv \mathbb{R}^n \implies$ Problema di ottimizzazione non vincolato
- se $X \subset \mathbb{R}^n \implies$ Problema di ottimizzazione vincolato

Per convenzione, i vettori sono pensati come vettori colonna e dunque un vettore riga sarà rappresentato con un vettore colonna trasposto. Inoltre possiamo trasformare un problema di massimo in un problema di minimo e viceversa, segue che

$$\max f(x) = -\min -f(x)$$

1.2 Punto di minimo

1.2.1 Punto di minimo globale

Un punto $x^* \in X$ è un punto di minimo globale di f su X se

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

1.2.2 Punto di minimo globale stretto

Un punto $x^* \in X$ è un punto di minimo globale stretto di f su X se

$$f(x^*) < f(x) \quad \forall x \in X$$

1.2.3 Punto di minimo locale

Un punto $x^* \in X$ è un punto di minimo locale di f su X se $\exists$ un intorno circolare $I(x^*, \varepsilon)$ di x^* , avente raggio $\varepsilon > 0$ tale che

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

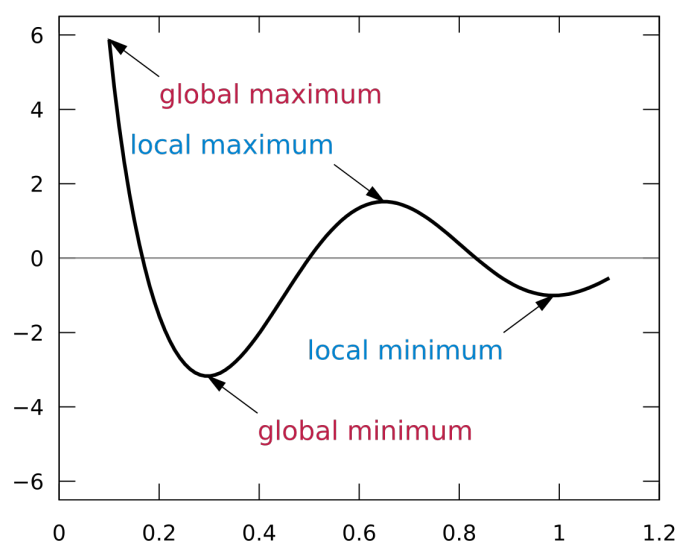
1.2.4 Punto di minimo locale stretto

Un punto $x^* \in X$ è un punto di minimo locale di f su X se $\exists$ un intorno circolare $I(x^*, \varepsilon)$ di x^* , avente raggio $\varepsilon > 0$ tale che

$$f(x^*) < f(x) \quad \forall x \in X \cap I(x^*, \varepsilon)$$

1.3 Osservazioni

Ovviamente un punto di minimo globale è anche locale, il viceversa può non essere vero. Un punto di massimo globale è anche locale, il viceversa può non essere vero.



Il problema di programmazione matematica consiste nel determinare, se esiste, un punto di minimo globale oppure se questo risultasse eccessivamente oneroso dal punto di vista risolutivo, almeno un punto di minimo locale.

Capitolo 2

Richiami di Analisi e Algebra

2.1 Derivata direzionale

2.1.1 Derivata direzionale di una funzione monodimensionale

Si consideri una funzione $f : \mathbb{R} \rightarrow \mathbb{R}$. Sia $x \in \mathbb{R}$ un punto in cui la funzione f è definita. Se esiste il limite

$$\lim_{\Delta \rightarrow 0} \frac{f(x + \Delta) - f(x)}{\Delta}$$

allora tale limite è la derivata di $f : \mathbb{R} \rightarrow \mathbb{R}$ nel punto x . In $\mathbb{R}^1$ la variabile indipendente può variare lungo la retta, in $\mathbb{R}^n$ si può avere la variazione di x in una qualsiasi direzione.

2.1.2 Definizione di derivata direzionale

Si consideri una funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}$, un vettore $d \in \mathbb{R}^n$ e uno scalare $\lambda \in \mathbb{R}$. Sia $x \in \mathbb{R}^n$ un punto in cui la funzione f è definita. Se esiste il limite

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

allora tale limite è la derivata direzionale della funzione f nel punto x lungo la direzione

$$d = \begin{bmatrix} d_1 & d_2 & \cdots & d_n \end{bmatrix}^T$$

Si noti nel caso in cui

$$d = \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix}^T$$

ossia se d è il vettore costituito da tutti zeri tranne 1 in i -esima posizione, la derivata direzionale coincide con la derivata parziale di f rispetto alla variabile x_i , indicata con

$$\frac{\partial f}{\partial x_i}$$

La derivata parziale seconda rispetto a due variabili nell'ordine x_i e x_j è

$$\frac{\partial^2 f}{\partial x_i \partial x_j}$$

2.2 Gradiente

Si consideri $f : \mathbb{R}^n \rightarrow \mathbb{R}$ e un punto $x \in \mathbb{R}^n$. Se in x esistono le n derivate parziali, definiamo gradiente della funzione f in x il vettore $\nabla f(x) \in \mathbb{R}^n$ avente come componenti le derivate parziali cioè

$$\nabla f(x) = \left[\frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \quad \cdots \quad \frac{\partial f}{\partial x_n} \right]^T \in \mathbb{R}^n$$

2.3 Matrice Hessiana

Sia $f : \mathbb{R}^n \rightarrow \mathbb{R}$ due volte continuamente differenziabile in $x \in \mathbb{R}^n$. Definiamo matrice Hessiana della funzione f in x la matrice

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \in \mathbb{R}^{n \times n}$$

Si noti che nelle ipotesi di derivate parziali seconde continue, per il Teorema di Schwarz, la matrice Hessiana è simmetrica

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i} \quad \forall i, j = 1, 2, \dots, n$$

2.4 Formula di Taylor

2.4.1 Formula di Taylor di una funzione in una variabile

Sia $f : \mathbb{R} \rightarrow \mathbb{R}$ una funzione in una variabile con derivata continua in un intorno di un punto x e si considera il punto $x + h$, appartenente a tale intorno, allora l'incremento di f risulta essere

$$f(x + h) = f(x) + f'(x)h + \beta_1(x, h)$$

dove $\beta_1(x, h)$ è un infinitesimo di ordine superiore rispetto ad h cioè

$$\lim_{h \rightarrow 0} \frac{\beta_1(x, h)}{h} = 0$$

2.4.2 Formula di Taylor di una funzione in due variabili

Sia $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ una funzione in una variabile con derivata continua seconda in un intorno di un punto x e si considera il punto $x + h$, appartenente a tale intorno, allora l'incremento di f risulta essere

$$f(x + h) = f(x) + f'(x)h + \frac{1}{2}f''(x)h^2 + \beta_2(x, h)$$

dove $\beta_2(x, h)$ è un infinitesimo di ordine superiore rispetto ad h^2 cioè

$$\lim_{h \rightarrow 0} \frac{\beta_2(x, h)}{h^2} = 0$$

La formula di Taylor permette di approssimare il valore di una funzione in un punto "incrementato" $x + h$ usando i valori delle derivate nel punto x , tale approssimazione è tanto migliore quanto meno ci spostiamo dal punto iniziale, ossia quanto più piccolo è h . In più dimensioni, il significato e la struttura della formula di Taylor sono molto simili.

2.4.3 Formula di Taylor di una funzione in più variabili

Sia $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una funzione in più variabili sufficientemente derivabili.

Sia $x \in \mathbb{R}^n, h \in \mathbb{R}^n, \nabla f(x) \in \mathbb{R}^n, \nabla^2 f(x) \in \mathbb{R}^{n \times n}$. Segue che

$$f(x + h) = f(x) + \nabla f(x)^T h + \beta_1(x, h)$$

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

dove $\beta_1(x, h)$ è un infinitesimo di ordine superiore rispetto alla norma dell'incremento h e $\beta_2(x, h)$ è un infinitesimo di ordine superiore rispetto al quadrato della norma di h cioè

$$\lim_{h \rightarrow 0} \frac{\beta_1(x, h)}{\|h\|} = 0 \quad \text{e} \quad \lim_{h \rightarrow 0} \frac{\beta_2(x, h)}{\|h\|^2} = 0$$

La norma di h è definita come

$$\|h\| = \sqrt{h_1^2 + h_2^2 + \dots + h_n^2}$$

2.5 Teorema 1

La derivata direzionale di f nel punto x lungo d è $\nabla f(x)^T d$

2.5.1 Dimostrazione

Consideriamo la seguente formula di Taylor

$$f(x + h) = f(x) + \nabla f(x)^T h + \beta_1(x, h)$$

e poniamo $h = \lambda d$, dove $\lambda \in \mathbb{R}$ è uno scalare. Segue che

$$f(x + \lambda d) = f(x) + \lambda \cdot \nabla f(x)^T d + \beta_1(x, \lambda, d)$$

Poi portiamo $f(x)$ a sinistra e dividendo per λ , si ha

$$\frac{f(x + \lambda d) - f(x)}{\lambda} = \nabla f(x)^T d + \frac{\beta_1(x, \lambda, d)}{\lambda}$$

da cui passando al limite si ottiene

$$\lim_{\lambda \rightarrow 0^+} \frac{f(x + \lambda d) - f(x)}{\lambda} = \lim_{\lambda \rightarrow 0^+} \left(\nabla f(x)^T d + \frac{\beta_1(x, \lambda, d)}{\lambda} \right) = \lim_{\lambda \rightarrow 0^+} \nabla f(x)^T d = \nabla f(x)^T d$$

2.6 Convessità

Dati K punti $x^{(1)}, x^{(2)}, \dots, x^{(K)} \in \mathbb{R}^n$ ed K scalari $\lambda_1, \lambda_2, \dots, \lambda_K$ si ha che $\sum_{i=1}^K \lambda_i \cdot x^{(i)}$

1. è detta combinazione lineare dei punti per generici coefficienti λ_i
2. è detta combinazione conica dei punti se $\lambda_i \geq 0 \quad \forall i = 1, \dots, K$
3. è detta combinazione affine dei punti se $\sum_{i=1}^K \lambda_i = 1$
4. è detta combinazione convessa dei punti se vale (2) e (3)

2.6.1 Combinazione convessa di due punti

Dati due punti $x, y \in \mathbb{R}^n$ e considerato uno scalare $\lambda \in [0, 1]$ allora si dice combinazione convessa di x e y un qualunque punto ottenuto come

$$z = \lambda x + (1 - \lambda)y$$

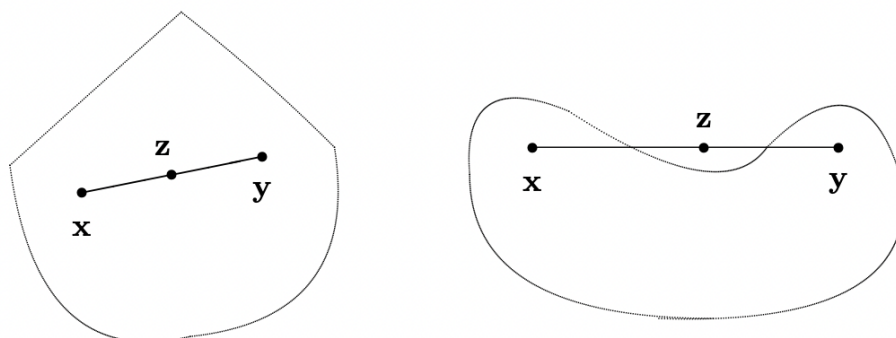
Inoltre, al variare di λ tra 0 e 1, si ottiene il segmento in $\mathbb{R}^n$ che unisce x e y . Segue che

$$\lambda = 1 \quad \implies \quad z = x$$

$$\lambda = 0 \quad \implies \quad z = y$$

2.6.2 Insieme convesso

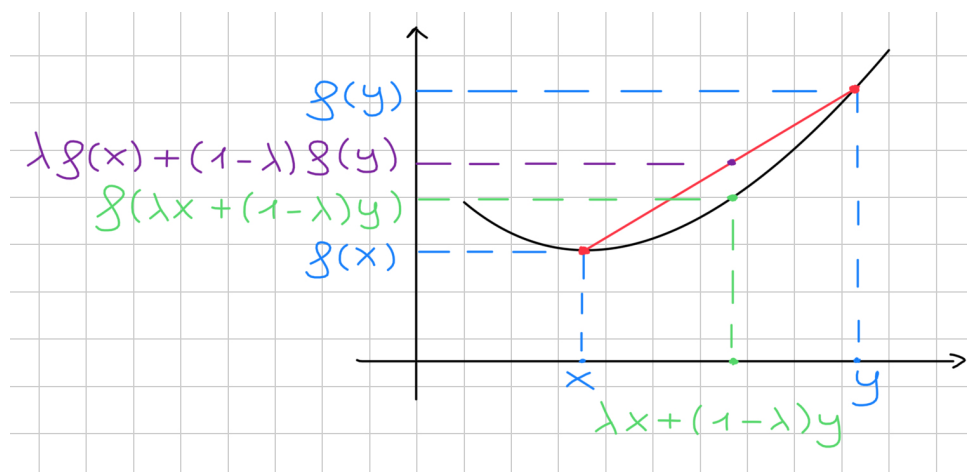
Un insieme $X \subset \mathbb{R}^n$ si dice convesso se, presi comunque due punti $x, y \in X$, il segmento che li unisce è interamente contenuto in X . Inoltre se un insieme non è convesso è concavo.



2.6.3 Funzione convessa

Una funzione f definita su un insieme convesso X si dice convessa se

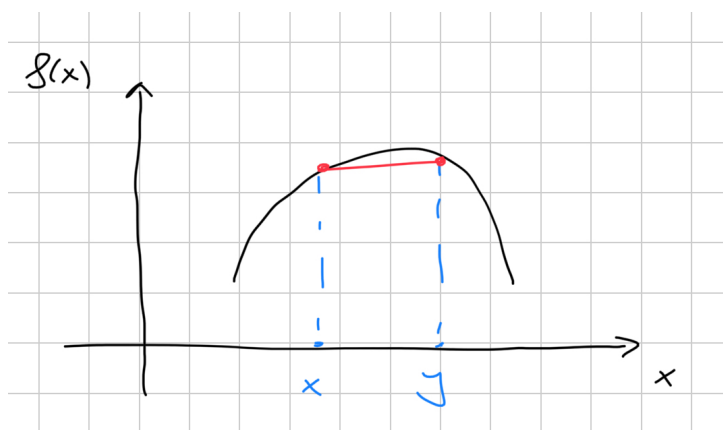
$$\lambda \cdot f(x) + (1 - \lambda)f(y) \geq f(\lambda x + (1 - \lambda)y) \quad \forall x, y \in X \quad \lambda \in [0, 1]$$



2.6.4 Funzione concava

Una funzione f definita su un insieme convesso X si dice concava se

$$\lambda \cdot f(x) + (1 - \lambda)f(y) \leq f(\lambda x + (1 - \lambda)y) \quad \forall x, y \in X \quad \lambda \in [0, 1]$$



Inoltre una funzione f tale che $-f$ è convessa è concava.

2.6.5 Teorema 2

Sia f una funzione continuamente differenziabile in $\mathbb{R}^n$ e convessa. Si ha che

$$f(y) - f(x) \geq \nabla f(x)^T (y - x) \quad \forall x, y \in \mathbb{R}^n$$

2.6.5.1 Dimostrazione

Dato che la funzione f è convessa, vale la relazione

$$\lambda \cdot f(x) + (1 - \lambda)f(y) \geq f(\lambda x + (1 - \lambda)y)$$

e ponendo $\varepsilon = 1 - \lambda$ possiamo riscrivere la relazione come

$$(1 - \varepsilon)f(x) + \varepsilon \cdot f(y) \geq f((1 - \varepsilon)x + \varepsilon \cdot y) = f(x + \varepsilon \cdot (y - x))$$

In questa disuguaglianza possiamo interpretare x come un punto e $x + \varepsilon \cdot (y - x)$ come un punto incrementato, ottenuto muovendosi nella direzione $(y - x)$ di ε .

Possiamo riscrivere la seguente formula di Taylor

$$f(x + h) = f(x) + \nabla f(x)^T h + \beta_1(x, h)$$

come

$$f(x + \varepsilon \cdot (y - x)) = f(x) + \varepsilon \cdot \nabla f(x)^T (y - x) + \beta_1(x, y, \varepsilon)$$

Dalla condizione di convessità e dalla formula di Taylor si ottiene

$$(1 - \varepsilon) \cdot f(x) + \varepsilon \cdot f(y) \geq f(x) + \varepsilon \cdot \nabla f(x)^T (y - x) + \beta_1(x, y, \varepsilon)$$

e semplificando $f(x)$ da entrambi le parti si ha

$$\varepsilon \cdot (f(y) - f(x)) \geq \varepsilon \cdot \nabla f(x)^T (y - x) + \beta_1(x, y, \varepsilon)$$

da cui, dividendo per ε e passando al limite

$$\lim_{\varepsilon \rightarrow 0^+} (f(y) - f(x)) \geq \lim_{\varepsilon \rightarrow 0^+} \left(\nabla f(x)^T (y - x) + \frac{\beta_1(x, y, \varepsilon)}{\varepsilon} \right) = \lim_{\varepsilon \rightarrow 0^+} \nabla f(x)^T (y - x)$$

si ha la tesi

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

2.6.6 Teorema 3

Sia f una funzione continuamente differenziabile in X . Se

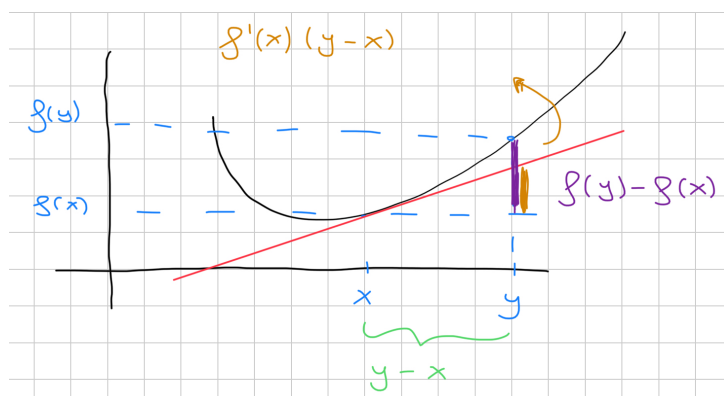
$$f(y) - f(x) \geq \nabla f(x)^T (y - x) \quad \forall x, y \in \mathbb{R}^n$$

allora la funzione f è convessa in X .

2.6.6.1 Dimostrazione

Dimostriamo graficamente il caso in $\mathbb{R}^1$ cioè

$$f(y) - f(x) \geq f'(x)(y - x)$$



Nel caso monodimensionale,

$$f(x) + f'(x)(y - x)$$

è l'ordinata del punto y sulla retta tangente alla curva in x . Se f è convessa allora la curva della funzione si trova sempre al di sopra di una retta tangente in un suo punto.

2.7 Matrici definite/semidefinite positive o negative

2.7.1 Matrice definita positiva

Data una matrice A quadrata di ordine n , essa si dice definita positiva se

$$x^T A x > 0 \quad \forall x \in \mathbb{R}^n, x \neq 0$$

2.7.2 Matrice semidefinita positiva

Data una matrice A quadrata di ordine n , essa si dice semidefinita positiva se

$$x^T A x \geq 0 \quad \forall x \in \mathbb{R}^n, x \neq 0$$

2.7.3 Matrice definita negativa

Una matrice A quadrata di ordine n si dice definita negativa se $-A$ è definita positiva.

2.7.4 Matrice definita negativa

Una matrice A quadrata di ordine n è semidefinita negativa se $-A$ è semidefinita positiva.

2.7.5 Teoremi di Sylvester

2.7.5.1 Teorema 4

Una matrice A simmetrica è definita positiva se e solo se i determinanti di tutti i minori principali della matrice A sono positivi. Inoltre se la matrice A non è simmetrica, la suddetta condizione è necessaria ma non sufficiente.

2.7.5.2 Teorema 5

Una matrice A simmetrica è semidefinita positiva se e solo se i determinanti di tutti i minori principali della matrice A sono non negativi. I minori principali di una matrice sono le n sottomatrici quadrate formate dall'intersezione delle prime i righe e delle prime j colonne. Consideriamo la matrice

$$M = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

Il minori principali della matrice M sono

$$M_1 = [2] \quad \implies \quad \det(M_1) = 2 > 0$$

$$M_2 = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \quad \implies \quad \det(M_2) = 3 > 0$$

$$M_3 = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & 0 \\ -1 & 0 & 1 \end{bmatrix} \implies \det(M_3) = 1 > 0$$

Quindi la matrice M è definita positiva.

2.7.6 Teorema 6

Tutti gli autovalori di una matrice A simmetrica definita positiva sono positivi e ciò implica che una matrice definita positiva è non singolare, dal momento che altrimenti avrebbe tra i suoi autovalori anche lo 0.

2.7.7 Teorema 7

Data una funzione f due volte continuamente differenziabile in $\mathbb{R}^n$, sia x un punto in cui $\nabla^2 f(x)$ è definita positiva. Allora, almeno in un intorno di x , la funzione f è convessa.

2.7.7.1 Dimostrazione

Consideriamo f con derivate seconde continue in $\mathbb{R}^n$, dati due punti $x, y \in \mathbb{R}^n$ e un incremento $h = y - x$ possiamo scrivere la formula di Taylor come

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

Supponiamo che la matrice Hessiana calcolata nel punto x sia definita positiva, segue che

$$(y - x)^T \nabla^2 f(x)(y - x) > 0$$

Essendo $\beta_2(x, y)$ un infinitesimo di ordine superiore rispetto al termine di secondo grado, si ha che il segno di quest'ultimo prevale rispetto a quello di $\beta_2(x, y)$, almeno in un intorno di x . Questo significa che, se eliminiamo a destra della formula di Taylor la quantità positiva $\beta_2(x, y)$ e anche

$$\frac{1}{2}(y - x)^T \nabla^2 f(x)(y - x)$$

che è positiva, ottengo una disuguaglianza

$$f(y) \geq f(x) + \nabla f(x)^T(y - x)$$

e portando $f(x)$ a sinistra si ha

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

Dunque, invocando il Teorema 3, risulta dimostrato il seguente Teorema 7.

2.7.7.2 Caso particolare

Un caso particolare di notevole interesse è quello relativo a funzioni la cui Hessiana è costante. Queste sono le funzioni del tipo

$$f(x) = \frac{1}{2}x^T A x + b^T x + c$$

con $A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$, $c \in \mathbb{R}$ ed esse sono dette funzioni quadratiche. In questo caso si può verificare che

$$\begin{cases} \nabla f(x) = Ax + b \\ \nabla^2 f(x) = A \end{cases}$$

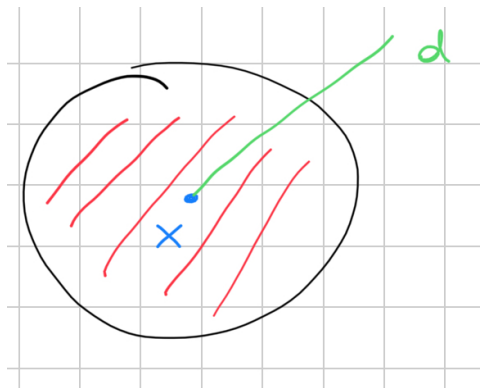
indipendentemente dal punto in cui essa è calcolata.

2.8 Esistenza di un minimo

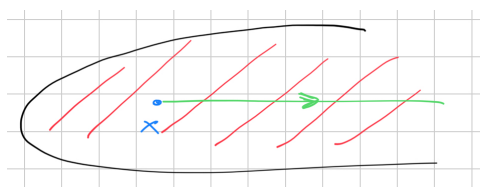
2.8.1 Insieme chiuso e limitato

Un insieme è chiuso se contiene la sua frontiera ed è limitato se non esistono direzioni di spostamento a partire da un certo punto x tale per cui se ci muoviamo, in modo indefinito, lungo questa direzione rimaniamo sempre nell'insieme X .

- Esempio di insieme chiuso e limitato



- Esempio di insieme chiuso e non limitato



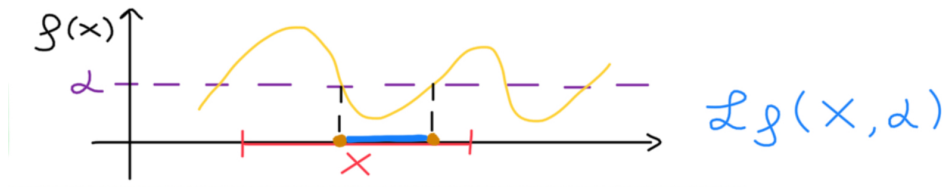
2.8.2 Teorema 8

Se una funzione f è continua su un insieme X chiuso e limitato, allora la funzione f ammette un punto di minimo globale in X .

2.8.3 Insieme di livello

Data una funzione f definita in un insieme X e sia $\alpha \in \mathbb{R}$ si definisce insieme di livello di f su X come l'insieme di tutti i punti x in cui il valore della funzione non eccede α ossia

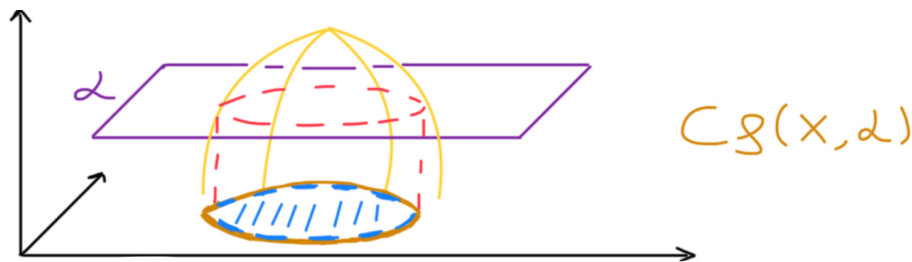
$$L_f(X, \alpha) = \{x \in X : f(x) \leq \alpha\}$$



2.8.4 Superficie di livello

Data una funzione f definita in un insieme X e sia $\alpha \in \mathbb{R}$ si definisce superficie di livello di f su X come l'insieme di tutti i punti x in cui il valore della funzione vale α ossia

$$C_f(X, \alpha) = \{x \in X : f(x) = \alpha\}$$



2.8.5 Teorema 9

Se una funzione f è continua su un insieme X chiuso, e almeno uno dei suoi insiemi di livello è chiuso e limitato, allora la funzione f ammette un punto di minimo in X .

2.9 Angolo tra due vettori e direzioni di discesa

Per introdurre i prossimi concetti conviene far riferimento al caso non vincolato, in cui X coincide con tutto $\mathbb{R}^n$, così che tutti i punti x sono ammissibili.

2.9.1 Prodotto scalare

Dati due punti $x, y \in \mathbb{R}^n$, il prodotto scalare è

$$x^T y = \sum_{i=1}^n x_i y_i$$

2.9.2 Angolo tra due vettori

Dati due vettori $x, y \in \mathbb{R}^n$, l'angolo tra essi è il numero ϑ (compreso tra 0 e π) tale che

$$\cos \vartheta = \frac{x^T y}{\|x\| \cdot \|y\|}$$

Se il prodotto scalare di due vettori è 0, i due vettori si dicono ortogonali e in questo caso risulta $\vartheta = 0$. Se i due vettori hanno la stessa direzione, i due vettori si dicono paralleli e in questo caso

$$\cos \vartheta = \frac{\alpha \cdot \|x\|^2}{|\alpha| \cdot \|x\|} = \begin{cases} 1 & \text{se } \alpha > 0 \\ -1 & \text{se } \alpha < 0 \end{cases} \implies \vartheta = \pm \pi$$

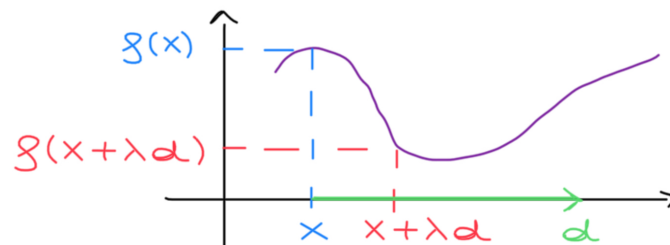
Consideriamo un punto x e una funzione f , da questo punto x possiamo pensare di muoverci seguendo una direzione $d \in \mathbb{R}^n$. Muovendosi di una quantità $\lambda \in \mathbb{R}$ lungo la direzione d , raggiungiamo il punto $x + \lambda d$. Siccome il nostro scopo ultimo è quello di trovare un punto di minimo della funzione f , risulta chiaramente di interesse stabilire se, muovendosi lungo la direzione d , la funzione cresce o decresce.

2.9.3 Direzione di discesa

Data una funzione f allora $d \in \mathbb{R}^n$ si dice direzione di discesa per f in $x \in \mathbb{R}^n$ se

$$\exists \bar{\lambda} > 0 : f(x + \lambda d) < f(x) \quad \forall 0 < \lambda < \bar{\lambda}$$

Quindi una direzione d è di discesa se, spostandosi da x lungo quella direzione, la funzione decresce purché gli spostamenti siano sufficientemente piccoli.



2.9.4 Direzione di salita

Data una funzione f allora $d \in \mathbb{R}^n$ si dice direzione di salita per f in $x \in \mathbb{R}^n$ se

$$\exists \bar{\lambda} > 0 : f(x + \lambda d) > f(x) \quad \forall 0 < \lambda < \bar{\lambda}$$

Quindi una direzione d è di salita se, spostandosi da x lungo quella direzione, la funzione cresce purché gli spostamenti siano sufficientemente piccoli.

2.9.5 Teorema 10

Si consideri una funzione f con gradiente continuo in $\mathbb{R}^n$. Dati $x \in \mathbb{R}^n$, una direzione $d \in \mathbb{R}^n$ e se la derivata direzionale di f in x rispetto alla direzione d è negativa allora d è una direzione di discesa per la funzione f in x .

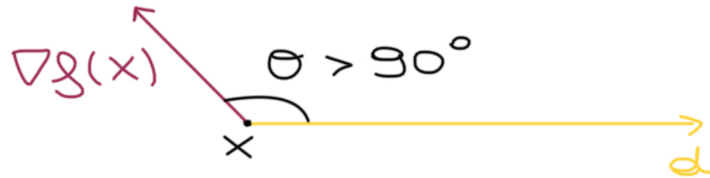
2.9.5.1 Dimostrazione

Dal Teorema 1, si ha

$$\lim_{\lambda \rightarrow 0^+} \frac{f(x + \lambda d) - f(x)}{\lambda} = \nabla f(x)^T d$$

e

$$\nabla f(x)^T d < 0 \quad \Rightarrow \quad \cos(\vartheta) < 0 \quad \Rightarrow \quad \vartheta > \frac{\pi}{2}$$



Quindi, per λ sufficientemente piccolo, si ha

$$f(x + \lambda d) - f(x) < 0$$

2.9.6 Teorema 11

Si consideri una funzione f con gradiente continuo in $\mathbb{R}^n$. Dati $x, d \in \mathbb{R}^n$ e se la derivata direzionale di f in x rispetto alla direzione d è positiva allora d è una direzione di salita per la funzione f in x .

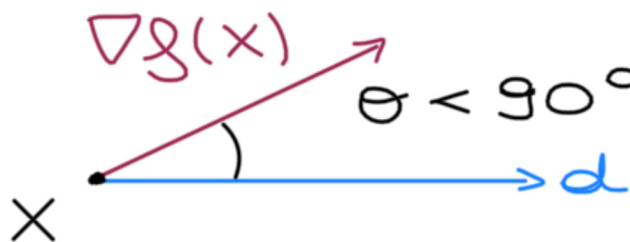
2.9.6.1 Dimostrazione

Se scegliessimo una direzione d per la quale

$$\nabla f(x)^T d > 0$$

si ha che la funzione f crescerebbe e d sarebbe una direzione di salita, allora

$$\cos(\vartheta) > 0 \quad \Rightarrow \quad \vartheta < \frac{\pi}{2}$$



Quindi, per λ sufficientemente piccolo, si ha

$$f(x + \lambda d) - f(x) > 0$$

2.9.7 Teorema 12

Si consideri una funzione f con gradiente continuo in $\mathbb{R}^n$. Dati $x, d \in \mathbb{R}^n$ e se la derivata direzionale di f in x rispetto alla direzione d è nulla allora non si può determinare nulla.

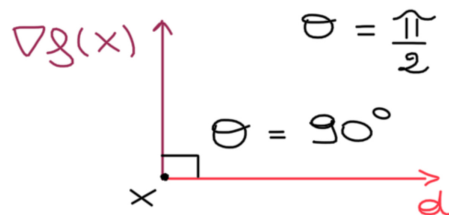
2.9.7.1 Dimostrazione

Se scegliessimo una direzione d per la quale

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

vuol dire che d è ortogonale al gradiente e non si può dire, in generale, se d è una direzione di discesa o meno quindi si ha

$$\vartheta = \frac{\pi}{2}$$

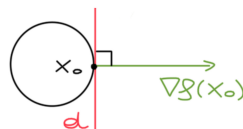


2.9.8 Osservazioni

Inoltre, osserviamo che se ci muoviamo lungo la superficie di livello

$$C_f(X, f(x_0)) = \{x \in X : f(x) = f(x_0)\}$$

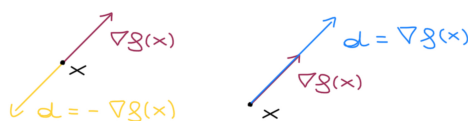
in ogni momento seguiamo una direzione d tangente alla superficie stessa. Ma non essendovi in quella direzione variazione della f , perché rimaniamo sulla stessa superficie di livello, la derivata direzionale è nulla. Questo indica che la direzione del gradiente di f in un punto x è sempre ortogonale alla superficie di livello passante per quel punto.



Si noti che il segno della derivata direzionale ci fornisce informazioni sull'angolo tra la direzione d e il gradiente. Se tale segno è negativo, ciò vuol dire che l'angolo (minore o uguale a π) fra d e $\nabla f(x)$ è ottuso. In particolare, se d è la direzione opposta a quella del gradiente, allora $d = -\nabla f(x)$ e l'angolo è piatto, allora

$$\cos(\vartheta) = \frac{-\nabla f(x)^T \cdot \nabla f(x)}{\|\nabla f(x)\|^2} = -1$$

e dunque l'antigradiente è sempre una direzione di discesa se $\nabla f(x) \neq 0$, mentre per lo stesso motivo il gradiente è sempre una direzione di salita.



Capitolo 3

Ottimizzazione Non Vincolata

3.1 Caratterizzazione dei punti di minimo

3.1.1 Teorema 13

Si consideri una funzione f con gradiente continuo in un punto $x^* \in \mathbb{R}^n$. La condizione necessaria affinché x^* sia un punto di minimo locale per f è che $\nabla f(x^*) = 0$

3.1.1.1 Dimostrazione

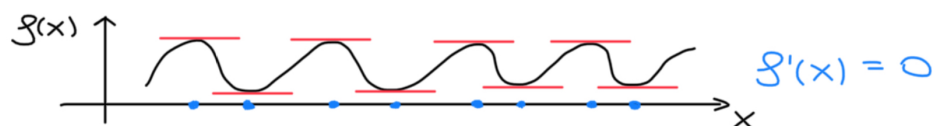
Per assurdo supponiamo che x^* sia un punto di minimo locale e che $\nabla f(x^*) \neq 0$ allora $d = -\nabla f(x^*)$ sarebbe una direzione di discesa ed inoltre, per il Teorema 10, esisterebbe

$$x^* - \lambda \cdot \nabla f(x^*)$$

tale che

$$f(x^* - \lambda \cdot \nabla f(x^*)) < f(x^*)$$

contraddicendo così l'ipotesi che x^* sia un punto di minimo locale. Dimostriamo graficamente il caso in cui $x^* \in \mathbb{R}^1$



Il Teorema 13 fornisce delle condizioni molto generali, dette condizioni del 1° ordine. Se la funzione f è due volte continuamente differenziabile, si possono enunciare anche delle condizioni del 2° ordine ed esse sono definite dal Teorema 14.

3.1.2 Teorema 14

Si consideri una funzione f con matrice Hessiana continua in un punto $x^* \in \mathbb{R}^n$. Le condizioni necessarie affinché x^* sia un punto di minimo locale per f sono che

$$\begin{cases} \nabla f(x^*) = 0 \\ \nabla^2 f(x^*) \text{ è semidefinita positiva} \end{cases}$$

3.1.2.1 Dimostrazione

La condizione

$$\nabla f(x^*) = 0$$

segue dal Teorema 13. Data una direzione $d \in \mathbb{R}^n$, poiché f è due volte differenziabile, possiamo scrivere la formula di Taylor

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

con riferimento a un punto incrementato $x^* + \lambda d$ dove d è una qualsiasi direzione quindi

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

Dalla prima condizione cioè $\nabla f(x^*) = 0$ si ha

$$f(x^* + h) = f(x^*) + \underbrace{\lambda \nabla f(x^*)^T d}_0 + \frac{1}{2} \lambda^2 d^T \nabla^2 f(x^*) d + \beta_2(x^*, \lambda, d)$$

poi, portando a sinistra $f(x^*)$ e dividendo per λ^2 , otteniamo

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

Per ipotesi x^* è un minimo locale, per λ sufficientemente piccolo, il termine di sinistra è sicuramente non negativo e quindi risulta

$$\frac{1}{2} d^T \nabla^2 f(x^*) d + \frac{\beta_2(x^*, \lambda, d)}{\lambda^2} \geq 0$$

da cui, passando al limite per $\lambda \rightarrow 0^+$, si ha

$$\lim_{\lambda \rightarrow 0^+} \frac{1}{2} d^T \nabla^2 f(x^*) d + \frac{\beta_2(x^*, \lambda, d)}{\lambda^2} = \frac{1}{2} d^T \nabla^2 f(x^*) d$$

Osservando che d è una direzione di discesa qualsiasi quindi

$$\frac{1}{2} d^T \nabla^2 f(x^*) d \geq 0$$

e allora $\nabla^2 f(x^*)$ è una matrice semidefinita positiva.

3.1.3 Teorema 15

Si consideri una funzione f con matrice Hessiana continua in un intorno di un punto $x^* \in \mathbb{R}^n$. Se valgono le seguenti condizioni

$$\begin{cases} \nabla f(x^*) = 0 \\ \nabla^2 f(x^*) \text{ è definita positiva} \end{cases}$$

allora x^* è un punto di minimo locale stretto per f .

3.1.3.1 Dimostrazione

La prima condizione

$$\nabla f(x^*) = 0$$

segue dal Teorema 13. Data una direzione $d \in \mathbb{R}^n$, poiché f è due volte differenziabile, possiamo scrivere la formula di Taylor

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

con riferimento a un punto incrementato $x^* + \lambda d$ dove d è una qualsiasi direzione quindi

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

Dalla prima condizione cioè $\nabla f(x^*) = 0$ si ha

$$f(x^* + h) = f(x^*) + \underbrace{\lambda \nabla f(x^*)^T d}_0 + \frac{1}{2} \lambda^2 d^T \nabla^2 f(x^*) d + \beta_2(x^*, \lambda, d)$$

poi, portando a sinistra $f(x^*)$ e dividendo per λ^2 , otteniamo

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

Per ipotesi x^* è un minimo locale, per λ sufficientemente piccolo, il termine di sinistra è sicuramente non negativo e quindi risulta

$$\frac{1}{2} d^T \nabla^2 f(x^*) d + \frac{\beta_2(x^*, \lambda, d)}{\lambda^2} \geq 0$$

da cui, passando al limite per $\lambda \rightarrow 0^+$, si ha

$$\lim_{\lambda \rightarrow 0^+} \frac{1}{2} d^T \nabla^2 f(x^*) d + \frac{\beta_2(x^*, \lambda, d)}{\lambda^2} = \frac{1}{2} d^T \nabla^2 f(x^*) d$$

Dalla seconda condizione cioè $\nabla^2 f(x^*)$ è definita positiva abbiamo che per qualunque d e, per λ sufficientemente piccolo, segue che

$$f(x^* + \lambda d) - f(x^*)$$

e quindi x^* è punto di minimo locale stretto.

3.2 Caratterizzazione dei punti di massimo

3.2.1 Teorema 16

Si consideri una funzione f con gradiente continuo in un punto $x^* \in \mathbb{R}^n$. La condizione necessaria affinché x^* sia un punto di massimo locale per f è che $\nabla f(x^*) = 0$

3.2.2 Teorema 17

Si consideri una funzione f con matrice Hessiana continua in un punto $x^* \in \mathbb{R}^n$. Le condizioni necessarie affinché x^* sia un punto di massimo locale per f sono

$$\begin{cases} \nabla f(x^*) = 0 \\ \nabla^2 f(x^*) \text{ è semidefinita negativa} \end{cases}$$

3.2.3 Teorema 18

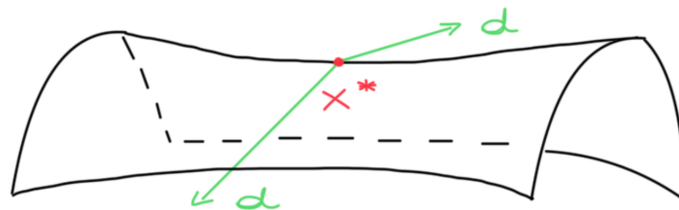
Si consideri una funzione f con matrice Hessiana continua in un intorno di un punto $x^* \in \mathbb{R}^n$. Se valgono le seguenti condizioni

$$\begin{cases} \nabla f(x^*) = 0 \\ \nabla^2 f(x^*) \text{ è definita negativa} \end{cases}$$

allora x^* è un punto di massimo locale stretto per f .

3.2.4 Punto di sella

Se in un punto x^* si annulla il gradiente, ma la matrice Hessiana è indefinita allora possiamo escludere che x^* sia punto di minimo o massimo. In tal caso, x^* si può muovere sia lungo direzioni di discesa che di salita ed è detto punto di sella.



3.3 Caso convesso

Consideriamo il caso in cui f è una funzione convessa, è possibile dimostrare alcune proprietà molto forti della soluzione ottima del problema. Inoltre, sotto ipotesi molto generali, nel caso convesso la distinzione tra minimi locali e globali non sussiste.

3.3.1 Teorema 19

Si consideri una funzione f convessa in $\mathbb{R}^n$. Se x^* è un punto di minimo locale, è anche un punto di minimo globale.

3.3.1.1 Dimostrazione

Essendo x^* un punto di minimo locale, segue che

$$\exists I(x^*, \varepsilon) : f(x^*) \leq f(x) \quad \forall x \in I(x^*, \varepsilon)$$

Consideriamo un qualsiasi punto $z \notin I(x^*, \varepsilon)$ e sia

$$\tilde{x} = \lambda x^* + (1 - \lambda)z$$

un punto del segmento che unisce x^* e z con $0 < \lambda < 1$ tale che

$$\tilde{x} \in I(x^*, \varepsilon)$$

cioè scegliendo $\lambda \rightarrow 1$. Dato che la funzione f è convessa, segue che

$$f(x^*) \leq f(\tilde{x}) = f(\lambda x^* + (1 - \lambda)z) \leq \lambda f(x^*) + (1 - \lambda)f(z)$$

dato che

$$f(\tilde{x}) = f(\lambda x^* + (1 - \lambda)z)$$

è una combinazione convessa e

$$f(\tilde{x}) = f(\lambda x^* + (1 - \lambda)z) \leq \lambda f(x^*) + (1 - \lambda)f(z)$$

perché la funzione f è convessa. Semplificando, otteniamo

$$(1 - \lambda)f(x^*) \leq (1 - \lambda)f(z)$$

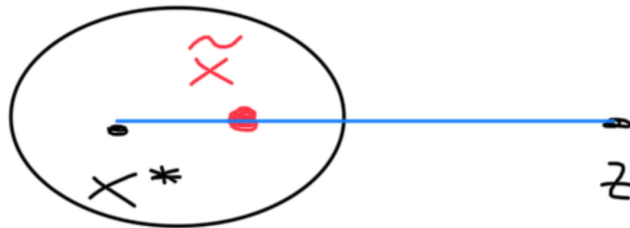
ed essendo $\lambda < 1$ si elimina $(1 - \lambda)$ ad entrambi i membri e si ha

$$f(x^*) \leq f(z)$$

ma

$$z \notin I(x^*, \varepsilon)$$

che è un assurdo.



3.3.1.2 Osservazioni

Il Teorema 19 vale in ipotesi del tutto generali, non abbiamo nemmeno supposto che la funzione f sia differenziabile. Se lo è, vediamo ora che la convessità consente di dare una caratterizzazione dei punti di minimo più forte di quanto visto finora. Infatti, in generale, il soddisfacimento delle condizioni necessarie del primo e del secondo ordine non basta a determinare la natura del punto in questione. Invece, nel caso particolare che la funzione f sia convessa, le sole condizioni del primo ordine diventano necessarie e sufficienti.

3.3.2 Teorema 20

Si consideri una funzione f con gradiente continuo e convessa in $\mathbb{R}^n$. Allora condizione necessaria e sufficiente affinché x^* sia un punto di minimo globale per f è che $\nabla f(x^*) = 0$.

3.3.2.1 Dimostrazione

La necessità della condizione deriva dal Teorema 13. Per quanto riguarda la sufficienza, consideriamo l'ipotesi che la funzione f sia convessa in $\mathbb{R}^n$ infatti dal Teorema 2 si ha

$$f(y) - f(x^*) \geq \nabla f(x^*)^T (y - x) \quad \forall x, y \in \mathbb{R}^n$$

per cui, se $\nabla f(x^*) = 0$ segue che

$$f(y) \geq f(x^*) \quad \forall x, y \in \mathbb{R}^n$$

quindi x^* è un punto di minimo globale per la funzione f .

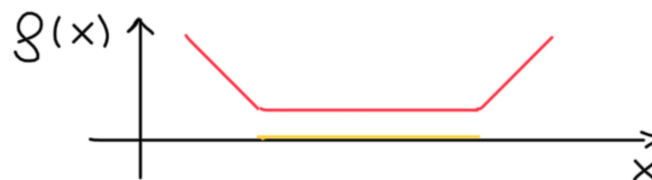
3.3.2.2 Osservazione

Dunque, nel caso convesso trovare un minimo locale equivale a trovare un minimo globale, e un punto è di minimo se e solo se soddisfa le condizioni del primo ordine.

3.3.3 Teorema 21

Si consideri una funzione f convessa in $\mathbb{R}^n$. Allora l'insieme dei punti di minimo della funzione f è convesso.

3.3.3.1 Esempio



3.3.3.2 Dimostrazione

Consideriamo due punti di minimo distinti x^*, x^{**} e sia $\bar{f}$ il valore ottimo di f . Il generico punto sul segmento che li unisce è

$$\lambda x^* + (1 - \lambda)x^{**} \quad \lambda \in [0, 1]$$

Per la convessità della funzione f

$$f(\lambda x^* + (1 - \lambda)x^{**}) \leq \lambda f(x^*) + (1 - \lambda)f(x^{**}) \quad \lambda \in [0, 1]$$

ma essendo sia x^* che x^{**} punti di minimo,

$$f(x^*) = f(x^{**}) = \bar{f}$$

e allora

$$f(\lambda x^* + (1 - \lambda)x^{**}) = \bar{f}$$

Quindi tutti i punti del segmento sono punti di minimo.

3.3.4 Osservazioni

Le condizioni di ottimalità che abbiamo visto possono essere utilizzate direttamente per calcolare la soluzione ottima di un problema di ottimizzazione non vincolata, ma spesso la forma della funzione obiettivo e/o il numero di variabili possono essere tali da rendere di fatto impossibili i calcoli in forma chiusa.

3.4 Esercizio 1

Consideriamo il problema

$$\begin{cases} \min x_1^2 + x_2^2 + x_1 x_2^2 \\ x \in \mathbb{R}^2 \end{cases}$$

Determinare i punti che soddisfano le condizioni di minimo locale.

3.4.1 Soluzione

Calcoliamo il gradiente della funzione obiettivo

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 + x_2^2 \\ 2x_2 + 2x_1 x_2 \end{bmatrix}$$

e la matrice Hessiana

$$\nabla^2 f(x) = \begin{bmatrix} 2 & 2x_2 \\ 2x_2 & 2 + 2x_1 \end{bmatrix}$$

che è simmetrica.

3.4.1.1 Condizioni del primo ordine

Studiamo i punti che soddisfano l'annullamento del gradiente

$$\begin{cases} 2x_1 + x_2^2 = 0 \\ 2x_2 + 2x_1 x_2 = 0 \end{cases} \implies 2x_2(1 + x_1) = 0$$

Se consideriamo

- $x_2 = 0$ allora

$$2x_1 = 0 \implies x_1 = 0 \implies A = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

- $x_1 = -1$ allora

$$-2 + x_2^2 = 0 \implies x_2 = \pm\sqrt{2} \implies B = \begin{bmatrix} -1 \\ \sqrt{2} \end{bmatrix} \quad C = \begin{bmatrix} -1 \\ -\sqrt{2} \end{bmatrix}$$

I punti A, B, C soddisfano le condizioni necessarie del primo ordine di minimo locale.

3.4.1.2 Condizioni del secondo ordine

Studiamo i punti che soddisfano l'annullamento della matrice Hessiana. Consideriamo

- punto A

$$\nabla^2 f(A) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

Dal teorema di Sylvester, abbiamo che la matrice è definita positiva e quindi A è un minimo locale stretto di f .

- punto B

$$\nabla^2 f(B) = \begin{bmatrix} 2 & 2\sqrt{2} \\ 2\sqrt{2} & 0 \end{bmatrix}$$

che non è semidefinita/definita positiva, possiamo studiare

$$-\nabla^2 f(B) = \begin{bmatrix} -2 & -2\sqrt{2} \\ -2\sqrt{2} & 0 \end{bmatrix}$$

che non è semidefinita/definita positiva allora non è semidefinita/definita negativa quindi la matrice Hessiana è indefinita. Segue che B è un punto di sella.

- punto C

$$\nabla^2 f(C) = \begin{bmatrix} 2 & -2\sqrt{2} \\ -2\sqrt{2} & 0 \end{bmatrix}$$

che non è semidefinita/definita positiva, possiamo studiare

$$\nabla^2 f(C) = \begin{bmatrix} -2 & 2\sqrt{2} \\ 2\sqrt{2} & 0 \end{bmatrix}$$

che non è semidefinita/definita positiva allora non è semidefinita/definita negativa quindi la matrice Hessiana è indefinita. Segue che C è un punto di sella.

3.4.1.3 Esistenza di un minimo globale

Considerando la funzione

$$f(x) = x_1^2 + x_2^2 + x_1 x_2^2$$

e supponiamo di porre $x_1 = -2$. Allora

$$f(x) = 4 + x_2^2 - 2x_2^2 = 4 - x_2^2 \quad x \in \mathbb{R}^2$$

Se

$$x_2 \rightarrow \pm\infty \implies f(x) \rightarrow -\infty$$

quindi la funzione è illimitata inferiormente, abbiamo una direzione

$$d = \begin{bmatrix} -2 \\ x_1 \end{bmatrix}$$

tale che la funzione tende a $-\infty$ cioè non esiste un minimo globale per f .

3.5 Esercizio 2

Consideriamo il problema

$$\begin{cases} \min x_2^2 - 3x_1x_2 + 3x_1^2 - x_1 + x_2 \\ x \in \mathbb{R}^2 \end{cases}$$

Determinare i punti che soddisfano le condizioni di minimo locale.

3.5.1 Soluzione

Notare che la funzione è quadratica, essa può essere espressa come

$$f(x) = x^T Ax + bx + c$$

con $A \in \mathbb{R}^{n \times n}$ è la matrice Hessiana, $b \in \mathbb{R}^n$ e $c \in \mathbb{R}$. Se la matrice Hessiana fosse semidefinita/definita positiva allora la funzione è convessa. Calcoliamo

$$\nabla f(x) = \begin{bmatrix} -3x_2 + 6x_1 - 1 \\ 2x_2 - 3x_1 + 1 \end{bmatrix} = 0 \quad \nabla^2 f(x) = \begin{bmatrix} 6 & -3 \\ -3 & 2 \end{bmatrix}$$

Abbiamo che la matrice Hessiana è definita positiva, allora la funzione f è convessa. Si ha che un punto di minimo locale è anche globale, condizione necessaria e sufficiente di minimo globale è l'annullamento del gradiente. Segue che

$$\begin{cases} -3x_2 + 6x_1 - 1 = 0 \\ 2x_2 - 3x_1 + 1 = 0 \end{cases}$$

Se sottraggo la prima equazione con la seconda moltiplicata per 2, otteniamo

$$x_2 + 1 = 0 \quad \implies \quad x_2 = -1 \quad \implies \quad x_1 = -1/3$$

Si ha che

$$x^* = \begin{bmatrix} -1/3 \\ -1 \end{bmatrix}$$

è un minimo globale.

Capitolo 4

Ottimizzazione Vincolata

4.1 Introduzione

Un problema si dice di ottimizzazione vincolata se $X \subset \mathbb{R}^n$ e studieremo problemi in cui i vincoli sono un insieme di m equazioni e p disequazioni. Il problema è il seguente

$$\left\{ \begin{array}{l} \min f(x) \\ h_1(x) = 0 \\ h_2(x) = 0 \\ \vdots \\ h_m(x) = 0 \\ g_{m+1}(x) \geq 0 \\ g_{m+2}(x) \geq 0 \\ \vdots \\ g_{m+p}(x) \geq 0 \end{array} \right.$$

ed è chiamato problema di programmazione non lineare (PNL), esso può essere scritto in forma compatta come

$$\left\{ \begin{array}{l} \min f(x) \\ h_i(x) = 0 \\ g_j(x) \geq 0 \end{array} \right.$$

dove

$$f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$$

$$h_i(x) : \mathbb{R}^n \rightarrow \mathbb{R}$$

$$g_j(x) : \mathbb{R}^n \rightarrow \mathbb{R}$$

4.1.1 Esempio

Consideriamo i seguenti esempi per comprendere che questa formulazione non è riduttiva.

$$X \equiv \{0, 1\}^n \equiv \left\{ \begin{array}{l} x_1^2 - x_1 = 0 \\ x_2^2 - x_2 = 0 \\ \vdots \\ x_n^2 - x_n = 0 \end{array} \right. \quad X \equiv \mathbb{Z} \equiv \left\{ \begin{array}{l} \sin(\pi x_1) = 0 \\ \sin(\pi x_2) = 0 \\ \vdots \\ \sin(\pi x_n) = 0 \end{array} \right.$$

4.2 Condizioni di ottimalità

Supponiamo che sia la funzione obiettivo f che le equazioni $h_i(x)$ e le disequazioni $g_j(x)$ siano almeno due volte differenziabili, notare che non vengono poste particolari restrizioni alla forma della regione ammissibile X . Infatti, benché la frontiera della regione ammissibile possa presentare "irregolarità" (ad esempio salti o punti angolosi), essa spesso è ancora esprimibile come intersezione di varie regioni, ciascuna avente frontiera regolare.

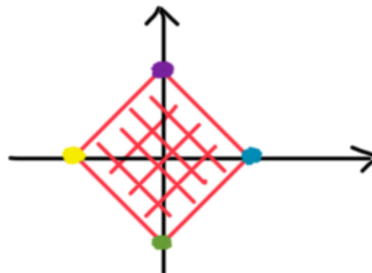
4.2.1 Esempio

Supponiamo che X consista in tutti e soli i punti che soddisfano

$$g(x_1, x_2) = -|x_1| - |x_2| + 1 \geq 0$$

La funzione $g(x_1, x_2)$ non è differenziabile nei quattro punti $(0, 1)$, $(0, -1)$, $(1, 0)$, $(-1, 0)$. Tuttavia, essa può anche essere equivalentemente espressa come

$$X \equiv \begin{cases} g_1(x_1, x_2) = x_1 + x_2 + 1 \geq 0 \\ g_2(x_1, x_2) = -x_1 - x_2 + 1 \geq 0 \\ g_3(x_1, x_2) = -x_1 + x_2 + 1 \geq 0 \\ g_4(x_1, x_2) = x_1 - x_2 + 1 \geq 0 \end{cases}$$



4.2.2 Vincoli attivi

Data una regione

$$X \equiv \begin{cases} h_i(x) = 0 & i = 1, \dots, m \\ g_j(x) \geq 0 & j = m+1, \dots, m+p \end{cases}$$

e un punto $x \in X$, l'insieme dei vincoli attivi $I_a(x)$ in x è costituito da tutti i vincoli soddisfatti in x all'uguaglianza ossia l'insieme di tutte le equazioni

$$h_i(x) = 0 \quad i = 1, \dots, m$$

e l'insieme di tutte le disequazioni

$$g_j(x) \geq 0 : g_j(x) = 0 \quad j = m+1, \dots, m+p$$

4.2.3 Matrice jacobiana

Sia

$$q = \begin{bmatrix} q_1 & q_2 & \cdots & q_m \end{bmatrix}$$

un vettore di funzioni di $x \in \mathbb{R}^n$. La matrice jacobiana è la matrice avente per colonne i gradienti delle m funzioni ossia

$$\frac{\partial q}{\partial x} = J(X) = \begin{bmatrix} \nabla q_1(x) & \nabla q_2(x) & \cdots & \nabla q_m(x) \end{bmatrix} = \begin{bmatrix} \frac{\partial q_1}{\partial x_1} & \frac{\partial q_2}{\partial x_1} & \cdots & \frac{\partial q_m}{\partial x_1} \\ \frac{\partial q_1}{\partial x_2} & \frac{\partial q_2}{\partial x_2} & \cdots & \frac{\partial q_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial q_1}{\partial x_n} & \frac{\partial q_2}{\partial x_n} & \cdots & \frac{\partial q_m}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{n \times m}$$

Ricordiamo che il gradiente di una funzione $f(x)$ in un punto x_0 è sempre ortogonale alla curva di livello passante per quel punto, ossia al luogo

$$f(x) = f(x_0)$$

Dunque, se consideriamo il luogo dei punti per cui

$$h_i(x) = 0 \quad i = 1, \dots, m$$

e calcoliamo il gradiente di $h_i(x)$, la direzione di $\nabla h_i(x)$ è ortogonale al vincolo.

4.3 Un vincolo di uguaglianza

Il caso più semplice che si possa presentare è quello per cui $m = 1$ e $p = 0$ ossia la regione ammissibile X è definita da un singolo vincolo di uguaglianza cioè

$$\begin{cases} \min f(x) \\ X \equiv \{h_1(x) = 0\} \end{cases}$$

dove

$$f(x) : \mathbb{R}^n \rightarrow \mathbb{R} \quad h_1(x) : \mathbb{R}^n \rightarrow \mathbb{R} \quad x \in \mathbb{R}^n$$

Sia $\bar{x}$ un punto appartenente a X quindi tale che $h_1(\bar{x}) = 0$ e supponiamo che $\nabla h_1(\bar{x})$ non sia nullo. Se andiamo a considerare un punto incrementato $\bar{x} + d$ con $d \in \mathbb{R}^n$, dalla formula di Taylor arrestata ai termini del primo ordine si ha

$$h_1(\bar{x} + d) = h_1(\bar{x}) + \nabla h_1(\bar{x})^T d + \beta(\bar{x}, d)$$

essendo $h_1(\bar{x}) = 0$, il nuovo punto $\bar{x} + d$ è ancora ammissibile se

$$h_1(\bar{x} + d) = 0$$

Ora, perché ciò accada, deve essere

$$\nabla h_1(\bar{x})^T d = 0$$

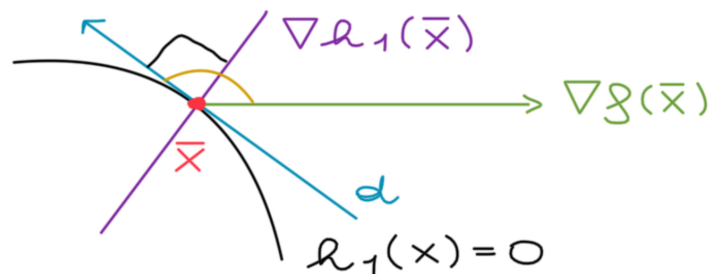
in quanto altrimenti, essendo $\beta(\bar{x}, d)$ un infinitesimo di ordine superiore, si avrebbe che per $\|d\|$ sufficientemente piccolo, $h_1(\bar{x} + d)$ risulterebbe pure diverso da zero. D'altro canto, il punto incrementato risulta migliore del precedente se d è una direzione di discesa, ossia, come sappiamo

$$\nabla f(\bar{x})^T d < 0$$

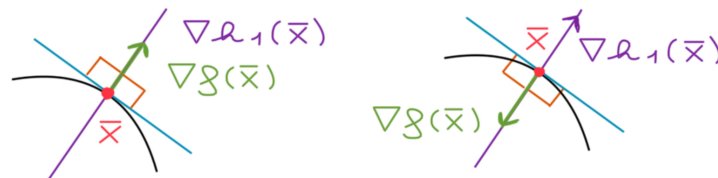
Dunque, se esiste un vettore d tale da soddisfare

$$\begin{cases} \nabla h_1(\bar{x})^T d = 0 & \text{direzione ammissibile} \\ \nabla f(\bar{x})^T d < 0 & \text{direzione di discesa} \end{cases}$$

possiamo sperare di trovare, a partire da $\bar{x}$, un punto ancora ammissibile e tale da migliorare la funzione obiettivo. Di conseguenza, una condizione necessaria affinché un punto $\bar{x}$ sia di minimo, è che non esista alcuna direzione d che sia ortogonale al gradiente del vincolo e allo stesso tempo che formi con il gradiente della funzione un angolo ottuso.



Se $\nabla h_1(\bar{x})$ e $\nabla f(\bar{x})$ sono paralleli, allora non riusciamo a trovare una direzione di discesa ammissibile e $\bar{x}$ può essere un punto di minimo.



4.3.1 Teorema 22

Date due funzioni $f(x)$ e $h_1(x)$, se in un punto x si ha che $\nabla f(x)$ e $\nabla h_1(x)$ non sono paralleli, allora esiste un vettore d che soddisfa

$$\begin{cases} \nabla h_1(x)^T d = 0 \\ \nabla f(x)^T d < 0 \end{cases}$$

4.3.1.1 Dimostrazione

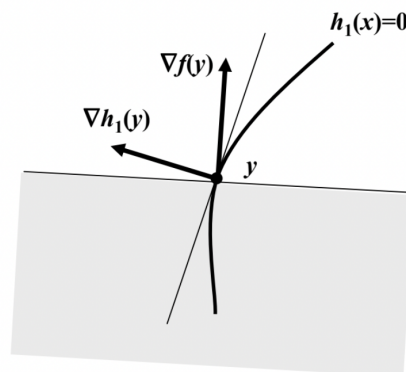
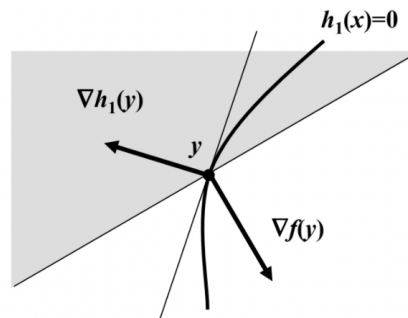
La dimostrazione è costruttiva e non particolarmente complessa: consiste nel mostrare che, se $\nabla f(x)$ e $\nabla h_1(x)$ non sono paralleli, il vettore

$$d = - \left(I - \frac{\nabla h_1(x) \nabla h_1(x)^T}{\|\nabla h_1(x)\|^2} \right) \nabla f(x)$$

soddisfa le condizioni del teorema.

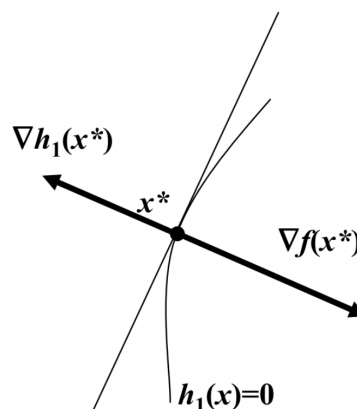
4.3.1.2 Interpretazione geometrica in due dimensioni

Se $\nabla f(x)$ e $\nabla h_1(x)$ non sono paralleli, esiste senz'altro una direzione contenuta nel semipiano indicato (e che dunque forma un angolo ottuso con $\nabla f(x)$) che è anche ortogonale a $\nabla h_1(x)$. Si noti che, avendo supposto $\nabla h_1(x) \neq 0$, l'ipotesi che $\nabla f(x)$ e $\nabla h_1(x)$ non siano paralleli implica che $\nabla f(x) \neq 0$.



Se $\nabla f(y)$ e $\nabla h_1(y)$ non sono paralleli, y non è un punto di minimo. Dal Teorema 22, abbiamo che all'ottimo senz'altro il gradiente della funzione e del vincolo devono essere paralleli, ossia, se x^* è un punto di minimo, deve esistere un valore di λ_1^* tale che

$$\nabla f(x^*) = \lambda_1^* \nabla h_1(x^*)$$



Abbiamo ∇f e ∇h_1 all'ottimo di un problema con un solo vincolo di uguaglianza. Tale condizione può essere espressa in un modo leggermente diverso, ma più utile per le estensioni che vedremo. Introduciamo la funzione lagrangiana

$$L(x, \lambda_1) = f(x) - \lambda_1 h_1(x)$$

e indichiamo con $\nabla_x L(x, \lambda_1)$ il gradiente calcolato rispetto al vettore delle variabili x

$$\nabla_x L(x, \lambda_1) = \left[\frac{\partial L}{\partial x_1} \quad \frac{\partial L}{\partial x_2} \quad \cdots \quad \frac{\partial L}{\partial x_n} \right]^T$$

Allora, la condizione

$$\nabla f(x^*) = \lambda_1^* \nabla h_1(x^*)$$

può riformularsi dicendo che deve esistere un $\lambda_1^* \in \mathbb{R}$ tale che

$$\nabla_x L(x^*, \lambda_1^*) = 0$$

Questa espressione suggerisce di cercare i punti di minimo del problema vincolato tra i punti stazionari della funzione lagrangiana. Il parametro λ_1 presente nella funzione prende il nome di moltiplicatore di Lagrange.

4.3.2 Esempio

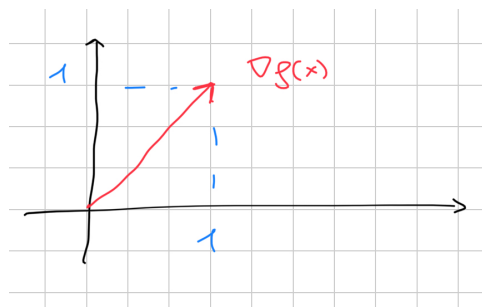
Applichiamo i concetti introdotti all'esempio

$$\begin{cases} \min x_1 + x_2 \\ h_1(x) = x_1^2 + x_2^2 - 2 = 0 \end{cases}$$

In questo esempio, X è costituita dalla circonferenza di raggio $\sqrt{2}$ e centro nell'origine. Evidentemente, il punto di minimo è $x^* = (-1, -1)$. Per ogni altro punto della circonferenza, è possibile muoversi in modo da mantenere l'ammissibilità (ossia, di rimanere sulla circonferenza) e contemporaneamente diminuire la funzione obiettivo. Ad esempio, dal punto $(-1, 1)$, la funzione migliora spostandosi lungo la circonferenza in senso antiorario. Si noti che il gradiente della funzione obiettivo è

$$\nabla f(x)^T = \begin{bmatrix} 1 & 1 \end{bmatrix}$$

indipendentemente dal punto ed ha direzione e verso come in figura



mentre

$$\nabla h_1(x)^T = \begin{bmatrix} 2x_1 & 2x_2 \end{bmatrix}$$

è dipendente da x . Si può osservare che nel punto x^* , il gradiente $\nabla f(x^*)$ e la normale al vincolo $\nabla h_1(x^*)$ sono paralleli, e

$$\nabla f(x^*) = -\frac{1}{2}\nabla h_1(x^*)$$

ossia vale

$$\nabla_x L(x^*, \lambda_1^*) = 0$$

con $\lambda_1^* = -1/2$. Si osservi che, benché la

$$\nabla_x L(x^*, \lambda_1^*) = 0$$

sia, come abbiamo visto, una condizione necessaria, essa non è in generale sufficiente affinché x^* sia un punto di minimo per la f nel problema vincolato. Infatti, in questo esempio, è immediato verificare che il punto $(1, 1)$ soddisfa la

$$\nabla_x L(x^*, \lambda_1^*) = 0$$

ma non è un punto di minimo (bensì è un punto di massimo).

Un'altra importante osservazione è che il segno del parametro λ_1^* non ha particolare significato. Infatti, nella formulazione del problema, avremmo potuto rappresentare il vincolo, anziché con

$$x_1^2 + x_2^2 - 2 = 0$$

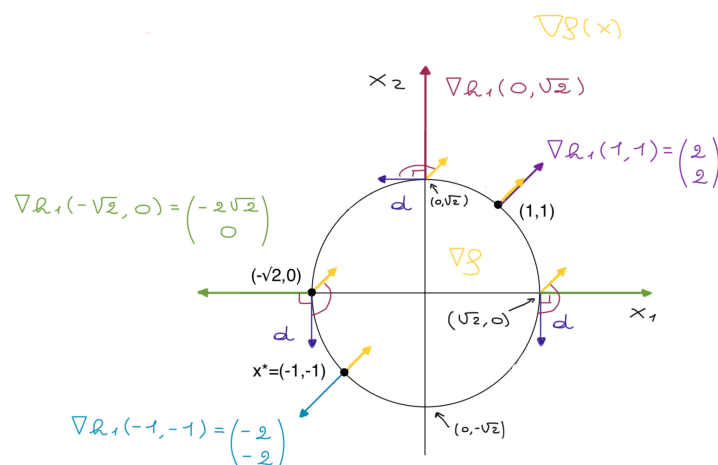
con

$$-x_1^2 - x_2^2 + 2 = 0$$

La soluzione del problema ovviamente non cambiava, ma per soddisfare la

$$\nabla_x L(x^*, \lambda_1^*) = 0$$

si sarebbe dovuto scegliere $\lambda_1^* = 1/2$. Dunque, se x^* è un punto di minimo di un problema con un solo vincolo di uguaglianza, $\nabla f(x^*)$ e $\nabla h_1(x^*)$ possono puntare nella stessa direzione o in direzioni opposte, ma l'essenziale è che siano paralleli.



Si ha che

$$\nabla h_1(1, 1) = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \quad \nabla h_1(-1, -1) = \begin{bmatrix} -2 \\ -2 \end{bmatrix} \quad \nabla h_1(-\sqrt{2}, 0) = \begin{bmatrix} -2\sqrt{2} \\ 0 \end{bmatrix}$$

Non riusciamo a trovare una direzione di discesa ammissibile nei punti $(1, 1)$ e $(-1, -1)$ perché in questi punti il gradiente della funzione e il gradiente del vincolo sono paralleli. Infatti abbiamo che il punto $(-1, -1)$ è il minimo globale e $(1, 1)$ è un punto di massimo, condizione necessaria e non sufficiente. Dimostriamo per $(-1, -1)$

$$\bar{x} = \begin{bmatrix} -1 \\ -1 \end{bmatrix} \quad \nabla f(\bar{x}) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\nabla h_1(\bar{x}) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \end{bmatrix} \quad \nabla f(\bar{x}) = \lambda_1 \nabla h_1(\bar{x}) = -\frac{1}{2} \begin{bmatrix} -2 \\ -2 \end{bmatrix}$$

sono paralleli.

4.4 Un vincolo di disuguaglianza

Consideriamo il caso in cui $m = 0$ e $p = 1$, ossia la regione ammissibile X è definita da una singola disequazione cioè

$$\begin{cases} \min & f(x) \\ & X \equiv \{g_1(x) \geq 0\} \end{cases}$$

dove

$$f(x) : \mathbb{R}^n \rightarrow \mathbb{R} \quad g_1(x) : \mathbb{R}^n \rightarrow \mathbb{R} \quad x \in \mathbb{R}^n$$

Sia $\bar{x}$ un punto appartenente a X , ossia tale che $g_1(\bar{x}) \geq 0$, e di nuovo supponiamo che $\nabla g_1(\bar{x}) \neq 0$. Seguendo lo stesso tipo di ragionamento svolto in precedenza, partendo da $\bar{x}$ vogliamo capire quali condizioni, se verificate, ci portano a escludere che $\bar{x}$ possa essere punto di minimo, e formulare così delle condizioni necessarie di ottimalità. Per quanto concerne la diminuzione della funzione obiettivo non cambia niente, ossia se non siamo al minimo deve esistere una direzione d ammissibile tale che

$$\nabla f(\bar{x})^T d < 0$$

Quello che è diverso è il modo in cui ora va affrontata la condizione di ottimalità, dalla formula di Taylor possiamo scrivere

$$0 \leq g_1(\bar{x} + d) \approx g_1(\bar{x}) + \nabla g_1(\bar{x})^T d$$

e dunque l'ammissibilità del punto $\bar{x} + d$ richiede che sia

$$g_1(\bar{x}) + \nabla g_1(\bar{x})^T d \geq 0$$

Per stabilire allora se esiste una direzione d tale da soddisfare

$$\begin{cases} \nabla f(\bar{x})^T d < 0 \\ g_1(\bar{x}) + \nabla g_1(\bar{x})^T d \geq 0 \end{cases}$$

distinguiamo il caso in cui $\bar{x}$ è nell'interno della regione ammissibile da quello in cui giace invece sulla frontiera.

4.4.1 Caso 1

Se $\bar{x}$ è all'interno di X , allora $g_1(\bar{x}) > 0$ cioè il vincolo non è attivo in $\bar{x}$. In tal caso,

$$g_1(\bar{x}) + \nabla g_1(\bar{x})^T d \geq 0$$

è verificata da qualunque vettore d , abbastanza piccolo in norma, tale che $\bar{x} + d$ sia ancora nella regione ammissibile. Si può verificare che se $\nabla f(\bar{x}) \neq 0$ possiamo sempre soddisfare

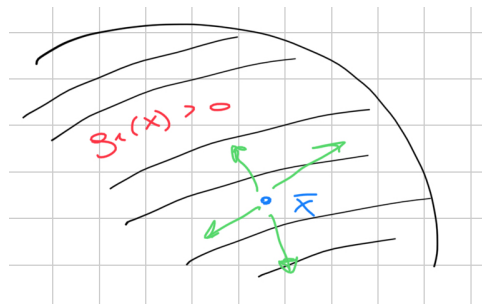
$$\begin{cases} \nabla f(\bar{x})^T d < 0 \\ g_1(\bar{x}) + \nabla g_1(\bar{x})^T d \geq 0 \end{cases}$$

considerando il vettore

$$\bar{d} = -g_1(\bar{x}) \frac{\nabla f(\bar{x})}{\|\nabla f(\bar{x})\| \cdot \|\nabla g_1(\bar{x})\|}$$

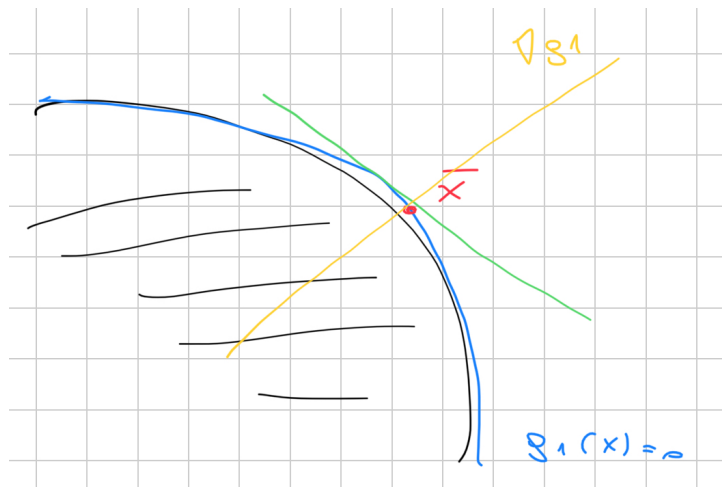
Dunque, in questo caso l'unica possibilità perché, a partire dal punto $\bar{x}$, non esista una direzione di discesa ammissibile è che sia

$$\nabla f(\bar{x}) = 0$$



4.4.2 Caso 2

Supponiamo che $\bar{x}$ appartenga alla frontiera di X cioè $g_1(\bar{x}) = 0$, il vincolo è attivo in $\bar{x}$.



Dato che $\bar{x} \rightarrow \bar{x} + d$ e quindi

$$g_1(\bar{x}) = 0 \implies g_1(\bar{x} + d) \geq 0$$

$$g_1(\bar{x} + d) = g_1(\bar{x}) + \nabla g_1(\bar{x})^T d + \beta_1(\bar{x}, d)$$

ma quando $\|d\| \rightarrow 0$ si ha

$$g_1(\bar{x} + d) \approx \nabla g_1(\bar{x})^T d \geq 0$$

quindi l'ammissibilità è verificata se

$$\nabla g_1(\bar{x}) \angle \leq 90^\circ \rightarrow \text{OK}$$

Le due condizioni

$$\begin{cases} \nabla f(\bar{x})^T d < 0 \\ g_1(\bar{x}) + \nabla g_1(\bar{x})^T d \geq 0 \end{cases}$$

divengono allora

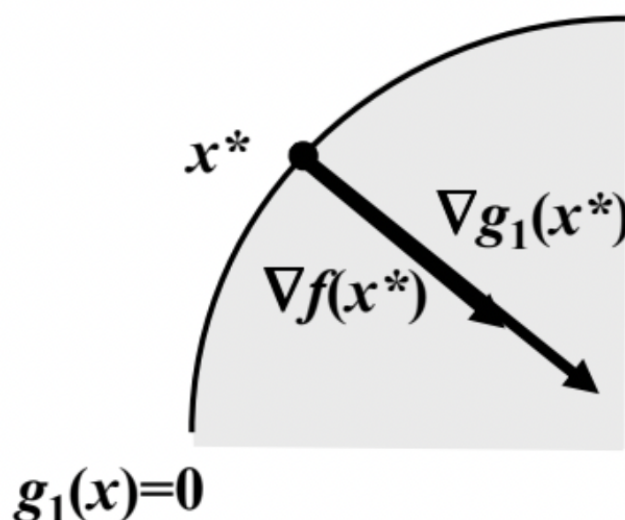
$$\begin{cases} \nabla f(\bar{x})^T d < 0 & \text{direzione di discesa} \\ \nabla g_1(\bar{x})^T d \geq 0 & \text{condizione di ammissibilità} \end{cases}$$

Queste due condizioni definiscono rispettivamente un semispazio aperto e uno chiuso. Se la loro intersezione non è vuota, è possibile individuare una direzione di discesa in cui è garantita ancora l'ammissibilità. Notiamo che l'unico caso in cui non esiste una direzione d che soddisfi

$$\begin{cases} \nabla f(\bar{x})^T d < 0 \\ \nabla g_1(\bar{x})^T d \geq 0 \end{cases}$$

è quello in cui $\nabla g_1(\bar{x})$ e $\nabla f(\bar{x})$ sono paralleli e puntano nello stesso verso, ossia esiste un $\lambda_1 \geq 0$ tale che

$$\nabla f(\bar{x}) = \lambda_1 \nabla g_1(\bar{x})$$



Si noti che stavolta il segno del moltiplicatore è importante. Se infatti

$$\nabla f(\bar{x}) = \lambda_1 \nabla g_1(\bar{x})$$

fosse soddisfatta con un moltiplicatore negativo, $\nabla g_1(\bar{x})$ e $\nabla f(\bar{x})$ punterebbero in direzioni opposte e i due semispazi definiti da

$$\begin{cases} \nabla f(\bar{x})^T d < 0 \\ \nabla g_1(\bar{x})^T d \geq 0 \end{cases}$$

verrebbero a coincidere (a meno della frontiera), e qualunque d in tale semispazio aperto soddisferebbe

$$\nabla g_1(\bar{x})^T d \geq 0$$

Introduciamo anche in questo caso la funzione lagrangiana

$$L(x, \lambda_1) = f(x) - \lambda_1 g_1(x)$$

e osserviamo che essa ci consente di unificare i due sotto-casi esaminati

$$\begin{cases} \nabla f(\bar{x}) = 0 \\ \nabla f(\bar{x}) = \lambda_1 \nabla g_1(\bar{x}) \end{cases}$$

Possiamo infatti concludere che se, nel punto x^* , non esiste una direzione di discesa ammissibile, risultano soddisfatte le due condizioni

$$\begin{cases} \nabla_x L(x^*, \lambda_1^*) = 0 \\ \lambda_1^* g_1(x^*) = 0 \end{cases} \quad \text{per qualche } \lambda_1^* \geq 0$$

La condizione

$$\lambda_1^* g_1(x^*) = 0$$

è nota come condizione di complementarità, e implica che il moltiplicatore di Lagrange λ_1^* può essere strettamente positivo solo se il vincolo è attivo. Infatti, se il vincolo non è attivo (caso 1), la condizione necessaria è l'annullamento del gradiente della f che si ottiene dalla

$$\nabla_x L(x^*, \lambda_1^*) = 0 \quad \text{per qualche } \lambda_1^* \geq 0$$

ponendo appunto $\lambda_1^* = 0$. Invece, se il vincolo è attivo (caso 2), la

$$\lambda_1^* g_1(x^*) = 0$$

è soddisfatta e rimane la

$$\nabla_x L(x^*, \lambda_1^*) = 0 \quad \text{per qualche } \lambda_1^* \geq 0$$

che coincide con la

$$\nabla f(\bar{x}) = \lambda_1 \nabla g_1(\bar{x})$$

Si noti che può anche accadere che $\lambda_1^* = 0$ pur essendo il vincolo attivo nel punto x^* . Riprendiamo l'esempio precedente, ma estendendo la regione X a tutto il cerchio delimitato dalla circonferenza. Il problema diviene

$$\begin{cases} \min x_1 + x_2 \\ g_1(x) = 2 - x_1^2 - x_2^2 \geq 0 \end{cases}$$

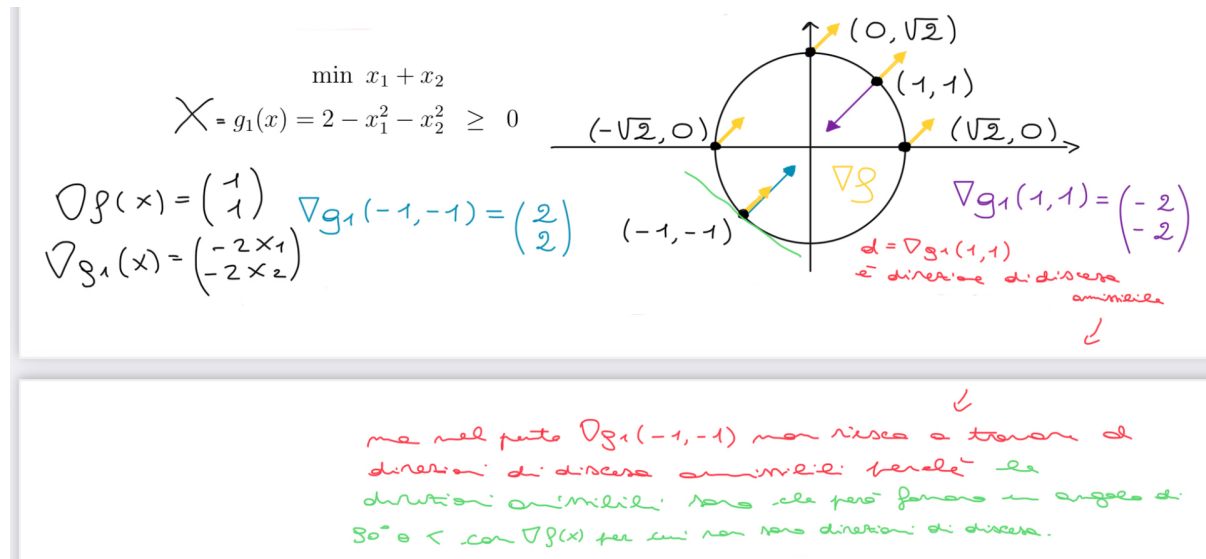
A differenza di quanto avveniva nel problema

$$\begin{cases} \min x_1 + x_2 \\ h_1(x) = x_1^2 + x_2^2 - 2 = 0 \end{cases}$$

Il vettore $\nabla g_1(x)$ calcolato in un punto x sulla circonferenza, punta verso l'interno del cerchio. Osserviamo che la soluzione ottima del problema è ancora data dal punto $x^* = (-1, -1)$ e che vale la condizione

$$\nabla f(x^*) = \lambda_1^*$$

con $\lambda_1^* = 1/2$.



4.5 Più vincoli di disuguaglianza

4.5.1 Definizione di cono

Dato un insieme di punti $x^1, x^2, \dots, x^k \in \mathbb{R}^n$, l'insieme di tutte le possibili combinazioni coniche di tali punti prende il nome di cono generato da $x^1, x^2, \dots, x^k$.

4.5.2 Definizione di punto regolare

Data una regione ammissibile

$$X \equiv \begin{cases} h_i(x) = 0 & i = 1, \dots, m \\ g_j(x) \geq 0 & j = m+1, \dots, m+p \end{cases}$$

un punto ammissibile x e il corrispondente insieme di vincoli attivi $I_a(x)$, si dice che i vincoli attivi soddisfano la condizione di qualificazione in x se i loro gradienti, calcolati in x , sono linearmente indipendenti. Un punto x per il quale vale la qualificazione dei vincoli attivi si dice un punto regolare. Ricordiamo che la condizione di qualificazione dei vincoli attivi in $\bar{x}$ equivale a richiedere che siano linearmente indipendenti le colonne della matrice jacobiana relativa alle funzioni

$$\{h_i(\bar{x}) : i = 1, \dots, m\} \cup \{g_j(\bar{x}) = 0 : m+1 \leq j \leq m+p\}$$

calcolata in $\bar{x}$. Si noti che il numero di vincoli attivi in un punto non supera la dimensione n dello spazio cioè $|I_a(\bar{x})| \leq n$, ciò corrisponde a richiedere che lo jacobiano dei vincoli attivi abbia rango pieno, mentre se $|I_a(\bar{x})| > n$ il punto non può essere regolare. Se in $\bar{x}$ il gradiente di un vincolo attivo è il vettore nullo allora $\bar{x}$ non è regolare quindi se $\bar{x}$ è un punto regolare, nessuno dei gradienti dei vincoli attivi in $\bar{x}$ può annullarsi.

4.5.3 Due vincoli di disuguaglianza

Consideriamo la situazione in cui abbiamo $p = 2$, ossia la regione ammissibile X è definita da due disequazioni, e sia $\bar{x}$ un punto ammissibile cioè tale che

$$X \equiv \begin{cases} g_1(\bar{x}) \geq 0 \\ g_2(\bar{x}) \geq 0 \end{cases}$$

Analizzeremo le condizioni sotto le quali, a partire da $\bar{x}$, esiste una direzione di discesa ammissibile. Nel caso in cui in $\bar{x}$ solo uno o nessuno dei due vincoli risulti attivo, si può ripetere lo stesso ragionamento fatto in precedenza mentre la situazione è diversa nel caso in cui in $\bar{x}$ ambedue i vincoli siano attivi.

4.5.3.1 Caso 1

Se $g_1(x)$ e $g_2(x)$ sono vincoli non attivi in $\bar{x}$ allora $\bar{x}$ è un punto interno, qualunque direzione $d \in \mathbb{R}^n$ è ammissibile e quindi d non è direzione di discesa se $\nabla f(\bar{x}) = 0$

4.5.3.2 Caso 2

Se $g_1(x)$ è attivo e $g_2(x)$ non è attivo in $\bar{x}$ allora non esiste una direzione di discesa ammissibile se

$$\nabla f(\bar{x}) = \lambda_1 \nabla g_1(\bar{x}) \quad \lambda_1 > 0$$

4.5.3.3 Caso 3

Se $g_1(x)$ non è attivo e $g_2(x)$ è attivo in $\bar{x}$ allora non esiste una direzione di discesa ammissibile se

$$\nabla f(\bar{x}) = \lambda_2 \nabla g_2(\bar{x}) \quad \lambda_2 > 0$$

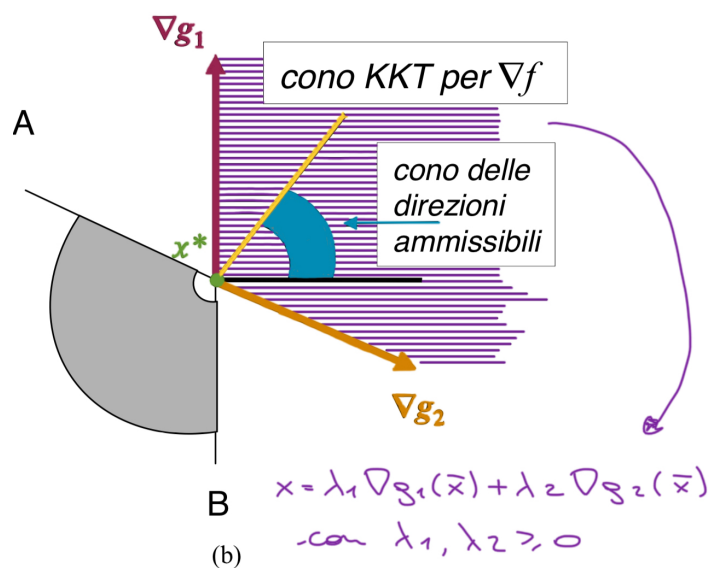
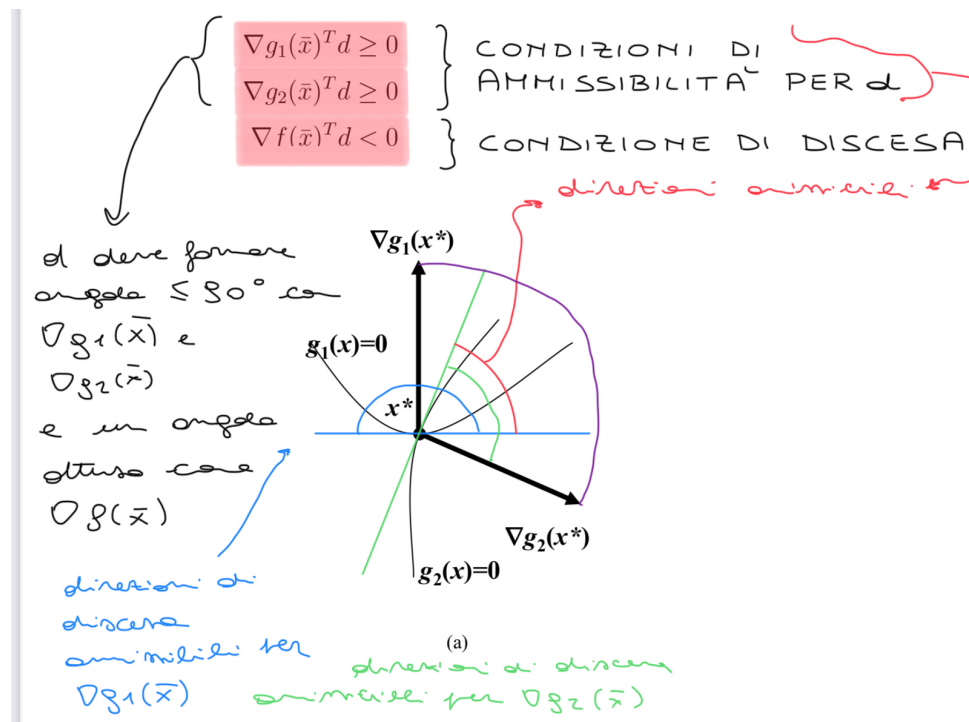
4.5.3.4 Caso 4

Supponiamo che $\bar{x}$ sia un punto regolare, in cui ambedue i vincoli sono attivi. Inoltre possiamo affermare che una direzione d è una direzione di discesa ammissibile se

$$\begin{cases} \nabla g_1(\bar{x})^T d \geq 0 & \text{condizione di ammissibilità per } g_1 \\ \nabla g_2(\bar{x})^T d \geq 0 & \text{condizione di ammissibilità per } g_2 \\ \nabla f(\bar{x})^T d < 0 & \text{condizione di direzione di discesa} \end{cases}$$

Consideriamo il cono in cui deve cadere ∇f se x^* è un punto di minimo. Le direzioni ammissibili sono racchiuse in un cono che è tanto più stretto quanto più grande è l'angolo (minore di 180 gradi) tra $\nabla g_1(\bar{x})$ e $\nabla g_2(\bar{x})$. Ora, se $\bar{x}$ è un punto di minimo, nessuna di queste direzioni può essere di discesa, il che implica che l'antigradiente deve essere tutto racchiuso nel cono che ha, come direzioni limite, quelle indicate in figura con A e B. Di

conseguenza, osservando la figura, si vede che se un punto x^* è un punto di minimo, allora il gradiente $\nabla f(x^*)$ deve essere contenuto nel cono generato da $\nabla g_1(x^*)$ e $\nabla g_2(x^*)$.



Le condizioni necessarie in una forma analoga a

$$\begin{cases} \nabla_x L(x^*, \lambda_1^*) = 0 \\ \lambda_1^* g_1(x^*) = 0 \end{cases} \quad \text{per qualche } \lambda_1^* \geq 0$$

definiamo la funzione lagrangiana come

$$L(x, \lambda) = f(x) - \lambda_1 g_1(x) - \lambda_2 g_2(x)$$

dove λ indica il vettore

$$\lambda = \begin{bmatrix} \lambda_1 & \lambda_2 \end{bmatrix}^T$$

dei moltiplicatori di Lagrange. Dunque, se x^* è un punto di minimo, deve esistere un vettore $\lambda^* \geq 0$ di moltiplicatori tale che

$$\begin{cases} \nabla_x L(x^*, \lambda^*) = 0 \\ \lambda_1^* g_1(x^*) = 0 & \text{condizione di complementarità per } g_1 \\ \lambda_2^* g_2(x^*) = 0 & \text{condizione di complementarità per } g_2 \end{cases}$$

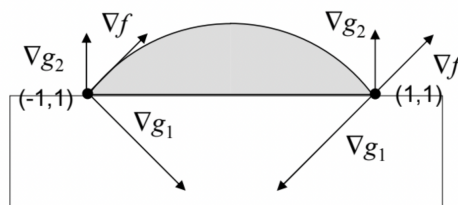
In particolare, la prima sta a indicare che

$$\nabla_x L(x^*, \lambda^*) = \nabla f(x^*) - \lambda_1^* g_1(x^*) - \lambda_2^* g_2(x^*) = 0$$

che esprime appunto il fatto che il gradiente di f è ottenibile come combinazione conica dei gradienti dei vincoli attivi in x^* . Riprendendo l'esempio precedente, introduciamo l'ulteriore vincolo $x_2 \geq 1$. Il problema diventa allora

$$\begin{cases} \min x_1 + x_2 \\ g_1(x) = 2 - x_1^2 - x_2^2 \geq 0 \\ g_2(x) = x_2 - 1 \geq 0 \end{cases}$$

e la regione ammissibile è



Si può verificare che la soluzione ottima è $x^* = (-1, 1)$, punto nel quale ambedue i vincoli sono attivi e che tale punto infatti soddisfa la

$$\nabla_x L(x^*, \lambda^*) = 0$$

scegliendo

$$\lambda^* = \begin{bmatrix} 1/2 \\ 2 \end{bmatrix}$$

mentre le condizioni di complementarità sono soddisfatte, essendo ambedue i vincoli attivi in x^* . Notiamo che in questo esempio tutti i punti ammissibili sono punti regolari. Consideriamo allora altri punti ammissibili che non sono di minimo per

$$\begin{cases} \min x_1 + x_2 \\ g_1(x) = 2 - x_1^2 - x_2^2 \geq 0 \\ g_2(x) = x_2 - 1 \geq 0 \end{cases}$$

e verifichiamo che in effetti le condizioni

$$\begin{cases} \nabla_x L(x^*, \lambda^*) = 0 \\ \lambda_1^* g_1(x^*) = 0 \\ \lambda_2^* g_2(x^*) = 0 \end{cases}$$

non sono soddisfatte. Prendiamo ad esempio il punto $\tilde{x} = (1, 1)$ e ambedue i vincoli sono attivi. Il gradiente della f non giace più nel cono individuato da

$$\begin{cases} \nabla g_1(\tilde{x})^T d \geq 0 \\ \nabla g_2(\tilde{x})^T d \geq 0 \end{cases}$$

Quindi, è possibile trovare una direzione di discesa ammissibile, come ad esempio è

$$d = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

Poi si può verificare che

$$\begin{cases} \nabla_x L(x^*, \lambda^*) = 0 \\ \lambda_1^* g_1(x^*) = 0 \\ \lambda_2^* g_2(x^*) = 0 \end{cases}$$

sono soddisfatte solo scegliendo

$$\lambda = \begin{bmatrix} -1/2 \\ 0 \end{bmatrix}$$

il che viola il requisito $\lambda \geq 0$. Consideriamo infine il punto $\bar{x} = (0, 1)$ nel quale solo g_2 è attivo. Poiché $g_1(\bar{x}) = 1$, si ha che la prima condizione di complementarità implica $\lambda_1 = 0$. Quindi, la condizione

$$\lambda_x L(\bar{x}, \lambda) = 0$$

si riduce a cercare un valore λ_2 tale che

$$\nabla f(\bar{x}) - \lambda_2 \nabla g_2(\bar{x}) = 0$$

ma evidentemente l'annullamento della prima componente di $\nabla_x L(\bar{x}, \lambda)$ richiederebbe che sia nullo ossia un tale valore λ_2 non esiste e dunque $\bar{x}$ non soddisfa le condizioni di ottimalità. Del resto, una direzione di discesa ammissibile è, ad esempio,

$$d = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

4.6 Le condizioni di Karush-Kuhn-Tucker

Dall'esempio illustrato, emerge il ruolo che alcune condizioni hanno nella caratterizzazione dei punti di minimo di un problema vincolato. Si tratta dell'annullamento del gradiente della funzione lagrangiana, la non negatività dei moltiplicatori e le condizioni di complementarità (le ultime due valgono solo relativamente ai vincoli espressi da disequazioni). Finora abbiamo supposto che nel punto di minimo sia verificata la qualificazione dei vincoli attivi. Ci si può chiedere se, venendo a mancare questa circostanza, le condizioni

$$\begin{cases} \nabla_x L(x^*, \lambda^*) = 0 \\ \lambda_1^* g_1(x^*) = 0 \\ \lambda_2^* g_2(x^*) = 0 \end{cases}$$

sono ancora necessarie. Consideriamo il seguente esempio

$$\begin{cases} \min -x_1 \\ g_1(x) = -(x_1 - 1)^3 - x_2^2 \geq 0 \\ g_2(x) = x_1 x_2 \geq 0 \end{cases}$$

Il punto $x^* = (1, 0)$ è un punto di minimo globale, in tal punto ambedue i vincoli sono attivi ma x^* non è un punto regolare dal momento che lo jacobiano dei vincoli attivi è

$$J(x^*) = \begin{bmatrix} \nabla g_1(x^*) & \nabla g_2(x^*) \end{bmatrix} = \begin{bmatrix} -3(x_1^* - 1)^2 & x_2^* \\ -2x_2^* & x_1^* \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Scrivendo la lagrangiana, si ha

$$L(x, \lambda) = -x_1 - \lambda_1(-(x_1 - 1)^3 - x_2^2) - \lambda_2 x_1 x_2$$

e dunque

$$\nabla_x L(x^*, \lambda^*) = 0$$

è

$$\frac{\partial L}{\partial x_1} = -1 + 3\lambda_1(x_1 - 1)^2 - \lambda_2 x_2 = 0$$

che, calcolata in $(1, 0)$, non è soddisfatta. Dunque, questo esempio mostra come

$$\begin{cases} \nabla_x L(x^*, \lambda^*) = 0 \\ \lambda_1^* g_1(x^*) = 0 \\ \lambda_2^* g_2(x^*) = 0 \end{cases}$$

siano necessarie per punti regolari. Si può osservare che i due vincoli sono attivi anche negli altri due punti $(0, 1)$ e $(0, -1)$ entrambi regolari, il secondo dei quali soddisfa anche le condizioni di KKT ed è un minimo locale. Possiamo generalizzare le condizioni precedenti solo per i punti regolari, diamo allora una definizione generale di funzione lagrangiana di un problema di PNL (programmazione non lineare).

4.6.1 Funzione lagrangiana di un problema di PNL

Dato un problema di programmazione non lineare, la funzione lagrangiana è definita come

$$L(x, \lambda) = f(x) - \sum_{i=1}^m \lambda_i h_i(x) - \sum_{j=m+1}^{m+p} \lambda_j g_j(x)$$

4.6.2 Teorema 23

Sia x^* un minimo locale di un problema di PNL e sia x^* un punto regolare. Allora esiste un vettore $\lambda^* \in \mathbb{R}^p$ tale che

$$\begin{cases} \nabla_x L(x^*, \lambda^*) = 0 \\ h_i(x^*) = 0 & i = 1, \dots, m \\ g_j(x^*) \geq 0 & j = m+1, \dots, m+p \\ \lambda_j^* \geq 0 & j = m+1, \dots, m+p \\ \lambda_j^* g_j(x^*) = 0 & j = m+1, \dots, m+p \end{cases}$$

Le precedenti condizioni prendono il nome di condizioni di Karush Kuhn Tucker, e per brevità chiameremo i punti che le soddisfano punti KKT. La prima condizione può anche esprimersi come

$$\nabla_x L(x^*, \lambda^*) = \nabla f(x^*) - \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) - \sum_{j=m+1}^{m+p} \lambda_j^* \nabla g_j(x^*) = 0$$

Si noti che, nel caso vi siano solo vincoli di disuguaglianza, e dal momento che i moltiplicatori λ_k^* relativi ai vincoli attivi devono essere non negativi, la precedente condizione richiede che, all'ottimo il gradiente della f sia contenuto nel cono individuato dai gradienti di tali vincoli. In definitiva, dato un problema di ottimizzazione vincolata (come stiamo supponendo f, h, g , sono tutte funzioni che ammettono derivate continue) che ammette un minimo globale, se indichiamo con K l'insieme dei punti KKT e con $\bar{\mathbb{R}}$ i punti non regolari, l'ottimo globale appartiene sicuramente all'insieme $K \cup \bar{\mathbb{R}}$.

4.6.3 Caso particolare di complementarità

Dato un problema di programmazione non lineare, un punto di minimo locale x^* e un vettore λ^* che soddisfa le condizioni KKT, vale la condizione di complementarità stretta se per ciascuna disequazione attiva in x^* si ha $\lambda_j^* > 0$. Quindi le condizioni di KKT possono essere soddisfatte da molti vettori λ^* in corrispondenza dello stesso punto x^* , se vale la stretta complementarità, allora λ^* è unico.

4.6.4 Condizioni necessarie del secondo ordine (solo equazioni)

Studiamo il corrispettivo, nei problemi di ottimizzazione vincolata, delle condizioni necessarie di ottimalità del secondo ordine limitandoci al caso di soli vincoli di uguaglianza. Le condizioni del secondo ordine riguardano la matrice Hessiana rispetto alle sole variabili x della funzione lagrangiana cioè $\nabla_{xx}^2 L(x, \lambda)$. Le condizioni enunciate in questo caso solo rispetto a punti in cui è soddisfatta la condizione di qualificazione dei vincoli attivi, riguardano il fatto che la matrice Hessiana sia semidefinita positiva. Tuttavia, questa condizione è meno restrittiva di quanto visto nell'ottimizzazione non vincolata.

4.6.5 Teorema 24

Sia x^* un minimo locale di un problema di PNL e in x^* valga la condizione di qualificazione dei vincoli attivi. Allora, per ogni vettore λ^* tale da soddisfare con x^* le condizioni di Karush Kuhn Tucker, si ha che

$$s^T \nabla_{xx}^2 L(x^*, \lambda^*) s \geq 0 \quad \forall s : J(x^*)s = 0$$

dove con $J(x^*)$ si è indicata la matrice jacobiana dei vincoli attivi in x^* calcolata in x^* . Le condizioni del secondo ordine richiedono che la matrice Hessiana del lagrangiano sia semidefinita positiva sullo spazio nullo della matrice jacobiana dei vincoli attivi in x^* .

4.6.6 Punti candidati ad essere minimi locali

In conclusione, si ha che i punti ad essere minimi locali sono

- punti $\bar{x} \in X$ che soddisfano le KKT
- punti non regolari

4.7 Sensibilità alla variazione dei parametri

Il moltiplicatore λ_i^* fornisce un'informazione legata al modo in cui il vincolo i influenza la funzione obiettivo. Nel seguito, ci occuperemo soltanto del caso in cui x^* è un punto di minimo locale regolare e faremo riferimento al caso in cui ci sono solo disequazioni. Se un certo vincolo g_j non è attivo in x^* ossia $g_j(x^*) > 0$, anche perturbando il vincolo di una piccola quantità (purché sufficientemente piccola) il vincolo sarà ancora non attivo e x^* continuerà ad essere minimo locale. In questo caso $\lambda_j^* = 0$.

Supponiamo invece che $g_j(x^*) = 0$ e perturbiamo il termine destro del vincolo di una piccola quantità $\varepsilon > 0$ cioè il vincolo da $g_j(x) \geq 0$ diviene

$$g_j(x) \geq -\varepsilon \|\nabla g_j(x^*)\|$$

ovvero, stiamo rendendo il vincolo leggermente meno restrittivo. Di conseguenza, anche il minimo si sposta in un nuovo punto, che indichiamo con $x^*(\varepsilon)$ e quindi il valore del minimo locale passa da $z^* = f(x^*)$ a $z^*(\varepsilon) = f(x^*(\varepsilon))$. Supponiamo che, a fronte di tale piccola perturbazione, l'insieme dei vincoli attivi non cambi (si può mostrare che l'insieme dei vincoli attivi non cambia se vale la stretta complementarità) cioè è lo stesso in x^* e in $x^*(\varepsilon)$. Si noti che

$$g_j(x^*(\varepsilon)) - g_j(x^*) = -\varepsilon \|\nabla g_j(x^*)\|$$

Allora, come sempre grazie alla formula di Taylor, possiamo scrivere

$$-\varepsilon \|\nabla g_j(x^*)\| = g_j(x^*(\varepsilon)) - g_j(x^*) \approx \nabla g_j(x^*)^T (x^*(\varepsilon) - x^*)$$

mentre per tutti gli altri vincoli attivi k diversi da j si ha

$$0 = g_k(x^*(\varepsilon)) - g_k(x^*) \approx \nabla g_k(x^*)^T (x^*(\varepsilon) - x^*)$$

D'altro canto abbiamo

$$z^*(\varepsilon) - z^* = f(x^*(\varepsilon)) - f(x^*) \approx \nabla f(x^*)^T (x^*(\varepsilon) - x^*)$$

che, per la

$$\nabla_x L(x^*, \lambda^*) = \nabla f(x^*) - \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) - \sum_{j=m+1}^{m+p} \lambda_j^* \nabla g_j(x^*) = 0$$

diviene

$$= \sum_{k=m+1}^{m+p} \lambda_k^* \nabla g_k(x^*)^T (x^*(\varepsilon) - x^*)$$

e sfruttando la

$$0 = g_k(x^*(\varepsilon)) - g_k(x^*) \approx \nabla g_k(x^*)^T (x^*(\varepsilon) - x^*)$$

rimane solo il termine relativo al vincolo j ossia

$$\approx \lambda_j^* \nabla g_j(x^*)^T (x^*(\varepsilon) - x^*)$$

e dunque dalla

$$-\varepsilon \|\nabla g_j(x^*)\| = g_j(x^*(\varepsilon)) - g_j(x^*) \approx \nabla g_j(x^*)^T (x^*(\varepsilon) - x^*)$$

in definitiva, si ha

$$z^*(\varepsilon) - z^* \approx -\varepsilon \|\nabla g_j(x^*)\| \lambda_j^*$$

di qui, dividendo per ε passando al limite per $\varepsilon \rightarrow 0$ si ha infine

$$\frac{dz^*(\varepsilon)}{d\varepsilon} = -\|\nabla g_j(x^*)\| \lambda_j^*$$

L'equazione precedente consente di effettuare un'analisi locale nell'intorno di x^* della sensibilità del valore ottimo locale della f a variazioni nel vincolo. Se $\|\nabla g_j(x^*)\| \lambda_j^*$ è grande, la sensibilità del valore ottimo a variazioni nel termine noto del vincolo è elevata mentre il contrario accade se $\|\nabla g_j(x^*)\| \lambda_j^*$ è piccola o nulla (almeno, a meno di infinitesimi di ordine superiore al primo).

L'analisi è valida anche per punti in cui non vale la complementarità stretta, considerando l'esempio

$$\begin{cases} \min -x_1 \\ g_1(x) = -(x_1 - 1)^3 - x_2^2 \geq 0 \\ g_2(x) = x_1 x_2 \geq 0 \end{cases}$$

si può verificare che il punto $A = (0, -1)$ è un minimo locale regolare, che soddisfa dunque le KKT, ma non vale la complementarità stretta cioè $\lambda_1 = 0$. Questo mostra che in effetti, se il primo vincolo fosse lievemente rilassato, diventando

$$-(x_1 - 1)^3 - x_2^2 \geq -\varepsilon \|\nabla g_1(A)\|$$

la funzione obiettivo non ne sarebbe influenzata, in quanto il punto A continuerebbe a essere un minimo locale. Si noti che infine che l'analisi di cui sopra è indipendente da fattori di scala. Se moltiplichiamo $g_j(x)$ per un fattore 10, il problema rimane ovviamente lo stesso ma si può vedere che il moltiplicatore di Lagrange ottimo diviene $\lambda_j^*/10$. Poiché al posto di $\|\nabla g_j(x^*)\|$ avremo

$$10 \cdot \|\nabla g_j(x^*)\|$$

il prodotto $\|\nabla g_j(x^*)\| \lambda_j^*$ non cambia. Se la funzione obiettivo viene moltiplicata per 10, dunque la sensibilità diviene 10 volte superiore.

4.8 Cenni sugli algoritmi di ottimizzazione vincolata

Le condizioni di ottimalità non sempre bastano a calcolare in modo rapido un punto stazionario o, meglio ancora, un punto di minimo. Gli algoritmi di ottimizzazione vincolata sono in genere complessi rispetto a quelli per il caso non vincolato. L'idea si basa sul concetto di ricondurre la soluzione di un problema vincolato a quella di un problema non vincolato. Un caso è più indicato quando i vincoli sono espressi da equazioni mentre l'altro quando i vincoli sono espressi da disequazioni. Tuttavia, con modifiche non particolarmente complicate, è possibile estendere ambedue gli approcci al caso generale. Gli approcci sono di tipo sequenziale cioè basati sulla soluzione di una successione di problemi non vincolati in modo tale che le soluzioni ottime convergano a quella del problema vincolato. Sono comunque diffusi approcci più sofisticati, quali quelli basati sulla programmazione quadratica o sui lagrangiani aumentati.

4.8.1 Funzioni di penalità quadratiche

In questo caso, consideriamo un problema con soli vincoli di uguaglianza cioè

$$\begin{cases} \min f(x) \\ h_1(x) = 0 \\ h_2(x) = 0 \\ \vdots \\ h_m(x) = 0 \end{cases}$$

L'idea alla base dei metodi basati sulle funzioni di penalità è concettualmente semplice, e consiste nel definire un opportuno problema non vincolato

$$\begin{cases} \min F(x) \\ x \in \mathbb{R}^n \end{cases}$$

Nella funzione obiettivo $F(x)$ è presente un termine che sparisce se i vincoli

$$h(x) = \begin{bmatrix} h_1(x) \\ h_2(x) \\ \vdots \\ h_m(x) \end{bmatrix} = 0$$

sono soddisfatti mentre altrimenti porta un contributo positivo alla funzione. L'approccio a questa soluzione diviene allora quello di riscrivere

$$\begin{cases} \min F(x) \\ x \in \mathbb{R}^n \end{cases}$$

ponendo

$$F(x) = f(x) + \rho \cdot p(h(x))$$

dove $\rho > 0$ è un opportuno coefficiente e $p(h(x))$ è la funzione di penalità definita come

$$p(h(x)) = \begin{cases} 0 & \text{se } h(x) = 0 \\ \alpha > 0 & \text{se } h(x) \neq 0 \iff x \notin X \end{cases}$$

Se ρ è molto grande, la soluzione

$$\begin{cases} \min f(x) \\ h(x) = 0 \end{cases}$$

risulterà vicina a

$$\begin{cases} \min F(x) \\ x \in \mathbb{R}^n \end{cases}$$

Il modo di procedere consiste allora nel risolvere una successione di problemi del tipo

$$\begin{cases} \min F(x) \\ x \in \mathbb{R}^n \end{cases}$$

per valori crescenti di ρ , ottenendo così una successione di punti che convergono alla soluzione ottima del problema vincolato. Si ha che la successione $\{\rho_k\}$ si suppone monotonicamente crescente e tale che

$$\lim_{k \rightarrow \infty} \rho_k = +\infty$$

Per quanto concerne la funzione di penalità sono possibili molte scelte diverse. Si noti che per poter risolvere il problema

$$\begin{cases} \min F(x) \\ x \in \mathbb{R}^n \end{cases}$$

con i metodi visti per il caso non vincolato è necessario garantire che la funzione complessiva $F(x)$ risulti sufficientemente regolare, in particolare differenziabile nei punti ammissibili per il problema vincolato. Una scelta abbastanza comune è

$$F(x) = f(x) + \rho \cdot \sum_{i=1}^m h_i(x)^2$$

quindi

$$p(h(x)) = \sum_{i=1}^m h_i(x)^2$$

4.8.2 Metodi di barriera

Un altro approccio sequenziale, applicato a problemi con soli vincoli di disuguaglianza.

$$\begin{cases} \min f(x) \\ g_1(x) \geq 0 \\ g_2(x) \geq 0 \\ \vdots \\ g_p(x) \geq 0 \end{cases}$$

Definiamo

$$g(x) = \begin{bmatrix} g_1(x) \\ g_2(x) \\ \vdots \\ g_p(x) \end{bmatrix}$$

L'idea è di trasformare il problema di ottimizzazione vincolata in un problema di ottimizzazione non vincolata in cui minimizzo una nuova funzione definita come

$$\begin{cases} \min F(x) \\ x \in \mathbb{R}^n \end{cases}$$

dove

$$F(x) = f(x) + \varepsilon \cdot v(x) \quad \varepsilon \in \mathbb{R}^+$$

e la funzione $v(x)$ detta funzione di barriera è definita come

$$v(x) \rightarrow +\infty \text{ se } g(x) \rightarrow 0$$

quindi se il punto x si sta avvicinando alla frontiera dell'insieme ammissibile X . Inoltre, come per il metodo delle funzioni di penalità, è possibile dimostrare che la successione delle soluzioni ottime dei problemi non vincolati converge a un minimo locale del problema vincolato. La funzione di barriera più utilizzata è quella logaritmica, definita come

$$v(x) = - \sum_{j=1}^p \log(g_j(x))$$

quindi

$$g_j(x) \rightarrow 0 \quad \implies \quad -\log(g_j(x)) \rightarrow +\infty$$

Capitolo 5

Programmazione Lineare

5.1 Introduzione

Un problema di Programmazione Lineare (PL) è un problema di ottimizzazione vincolata in cui la funzione obiettivo f è lineare e la regione ammissibile X è formata da m equazioni lineari e da n disequazioni lineari. Le funzioni lineari sono sia concave che convesse, siamo nelle ipotesi del Teorema 21, e quindi la regione ammissibile X è convessa. Si ha che

- la PL è un caso particolare di programmazione convessa
- un minimo locale è anche un minimo globale
- le KKT (condizioni di Karush-Kuhn-Tucker) sono condizioni necessarie e sufficienti di minimo globale anche nei punti non regolari

5.1.1 Teorema 25

Segue che $-g_j(x)$ è convessa se e solo se $g_j(x)$ è concava.

5.1.1.1 Dimostrazione

Devo dimostrare che

$$\forall x, y \in X \quad z = \lambda x + (1 - \lambda)y \in X \quad \lambda \in [0, 1]$$

Abbiamo che una funzione $h_i(x) = 0$ è lineare se

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n + b_i = 0$$

dove

$$a_{i1}, a_{i2}, \dots, a_{in}, b_i \in \mathbb{R}^n$$

Possiamo esprimere l'equazione in forma compatta con

$$a_i = \begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{in} \end{bmatrix} \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

e si ha

$$a_i^T x + b_i = 0$$

Dobbiamo dimostrare che l'insieme ammissibile è convesso

$$x, y \in X \iff \begin{cases} h_i(x) = 0 & i = 1, \dots, m \\ h_i(y) = 0 & i = 1, \dots, m \\ g_j(x) \geq 0 & j = m+1, \dots, m+p \\ g_j(y) \geq 0 & j = m+1, \dots, m+p \end{cases}$$

$$z = \lambda x + (1 - \lambda)y \in X \iff \begin{cases} h_i(z) = 0 & i = 1, \dots, m \\ g_j(z) \geq 0 & j = m+1, \dots, m+p \end{cases}$$

$$x, y \in X \iff \begin{cases} h_i(x) = 0 & i = 1, \dots, m \\ h_i(y) = 0 & i = 1, \dots, m \end{cases} \iff \begin{cases} a_i^T x + b_i = 0 \\ a_i^T y + b_i = 0 \end{cases}$$

La tesi è

$$a_i^T z + b_i = 0$$

con

$$z = \lambda x + (1 - \lambda)y \quad \lambda \in [0, 1]$$

Sostituendo otteniamo

$$a_i^T(\lambda x + (1 - \lambda)y) + b_i = \lambda a_i^T x + (1 - \lambda)a_i^T y + b_i = -\lambda b_i - (1 - \lambda)b_i + b_i = 0$$

dato che, per ipotesi, abbiamo che $a_i^T x = -b_i$ e $a_i^T y = -b_i$. Quindi

$$h_i(z) = 0 \quad i = 1, \dots, m$$

Inoltre dobbiamo dimostrare che

$$\begin{cases} g_j(x) \geq 0 \\ g_j(y) \geq 0 \end{cases} \implies g_j(z) \geq 0$$

Per la convessità di $g_j(x)$ si ha che

$$-g_j(z) = -g_j(\lambda x + (1 - \lambda)y) \leq -\lambda g_j(x) - (1 - \lambda)g_j(y) \leq 0$$

dato che per ipotesi

$$\begin{cases} g_j(x) \geq 0 \\ g_j(y) \geq 0 \end{cases}$$

e $\lambda \in [0, 1]$. Quindi

$$-g_j(z) \leq 0 \implies g_j(z) \geq 0$$

5.2 Problema primale in forma standard

Definiamo un problema primale in forma standard come

$$\left\{ \begin{array}{l} \min c_1x_1 + c_2x_2 + \dots + c_nx_n \\ a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m \\ x_1 \geq 0 \\ x_2 \geq 0 \\ \vdots \\ x_n \geq 0 \end{array} \right.$$

dove $x_1, \dots, x_n$ sono le variabili, $c_1, \dots, c_n$ sono i coefficienti di costo delle variabili nella funzione obiettivo, i termini a_{ij} sono chiamati coefficienti delle variabili nei vincoli e i termini $b_1, \dots, b_m$ sono chiamati termini noti dei vincoli. Definiamo

$$c = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \in \mathbb{R}^n \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \in \mathbb{R}^{m \times n} \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \in \mathbb{R}^m$$

dove c è il vettore dei costi delle variabili nella funzione obiettivo, b è il vettore dei termini noti, x è il vettore delle variabili ed A è la matrice dei coefficienti delle variabili nei vincoli. Il problema primale in forma standard può essere scritto in forma compatta come

$$\left\{ \begin{array}{l} \min c^T x \\ Ax = b \\ x \geq 0 \end{array} \right.$$

5.2.1 Funzione lagrangiana

Nello scrivere la funzione lagrangiana di un problema di PL in forma standard, esprimiamo separatamente i moltiplicatori associati ad vincoli di uguaglianza (che indichiamo con u) da quelli associati ai vincoli di non negatività (che indichiamo con s). Definiamo

$$u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{bmatrix} \in \mathbb{R}^m \quad s = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix} \in \mathbb{R}^n$$

La funzione lagrangiana è definita come

$$L(x, u, s) = c^T x - u^T (Ax - b) - s^T x$$

5.2.2 Condizioni KKT del problema primale

Sia x^* un punto di minimo, allora esistono u^*, s^* tali che valgono le condizioni KKT

1. $\nabla_x L(x^*, u^*, s^*) = c^T - u^{*T} A - s^{*T} = 0^T$
2. $Ax^* = b$
3. $x^* \geq 0$
4. $s^* \geq 0$
5. $s_i^* \cdot x_i^* = 0 \quad i = 1, \dots, n$

Da queste condizioni, ricaviamo altre condizioni come

- da (1) e (4)

$$s^{*T} = c^T - u^{*T} A \geq 0$$

- da (3), (4) e (5)

$$s^{*T} \cdot x^* = 0$$

- sostituendo la prima condizione nella seconda otteniamo

$$(c^T - u^{*T} A) \cdot x^* = 0$$

In definitiva, le condizioni KKT per il problema primale sono

1. $Ax^* = b$
2. $x^* \geq 0$
3. $c^T - u^{*T} A \geq 0^T$
4. $(c^T - u^{*T} A) \cdot x^* = 0$

Da queste condizioni, ricaviamo altre condizioni come

- da (4)

$$c^T x^* - u^{*T} Ax^* = 0$$

- da (1)

$$c^T x^* - u^{*T} b = 0$$

quindi se x^*, u^*, s^* soddisfano le KKT, allora vale la relazione

$$c^T x^* = u^{*T} b$$

5.3 Problema duale

Il problema duale è definito come

$$\begin{cases} \max b^T y \\ A^T y \leq c \end{cases}$$

dove

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \in \mathbb{R}^m \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \in \mathbb{R}^m \quad c = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \in \mathbb{R}^n \quad A^T \in \mathbb{R}^{n \times m}$$

Le variabili y sono chiamati duali, la funzione obiettivo del problema duale è ottenuta moltiplicando le variabili duali con il vettore b dei termini noti. Il vettore dei termini noti del problema primale diventa il vettore dei coefficienti di costo delle variabili duali, il vettore c è il vettore dei coefficienti di costo del problema primale. Segue che i vincoli sono formati da m disequazioni. Il problema duale è un problema di massimo e lo possiamo trasformare in un problema di minimo, si ha

$$\begin{cases} \max b^T y \\ A^T y \leq c \end{cases} \equiv \begin{cases} -\min -b^T y \\ A^T y \leq c \end{cases}$$

Poi associamo ad ogni vincolo un moltiplicatore di Lagrange, esso è rappresentato da

$$v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \geq 0$$

Inoltre trasformiamo i vincoli di $\leq$ in vincoli di $\geq$ e otteniamo

$$\begin{cases} -\min -b^T y \\ c - A^T y \geq 0 \end{cases}$$

5.3.1 Funzione lagrangiana

La funzione lagrangiana del problema duale è definita come

$$L(y, v) = -b^T y - v^T (c - A^T y)$$

5.3.2 Condizioni KKT del problema duale

Sia y^* un punto di minimo, allora esiste v^* tale che valgono le KKT

$$1. \nabla_y L(y^*, v^*) = -b^T + v^{*T} A^T = 0^T$$

2. $c - A^T y^* \geq 0$
3. $v^* \geq 0$
4. $v_i^* \cdot (c - A^T y^*)_i = 0 \quad i = 1, \dots, n$

Da queste condizioni, ricaviamo altre condizioni come

- da (1)

$$-b + Av^* = 0 \quad \implies \quad Av^* = b$$
- da (2), (3) e (4)

$$(c^T - y^{*T} A) \cdot v^* = 0$$

In definitiva, le condizioni KKT per il problema primale sono

1. $Av^* = b$
2. $v^* \geq 0$
3. $c^T - y^{*T} A \geq 0^T$
4. $(c^T - y^{*T} A) \cdot v^* = 0$

5.4 Condizioni KKT del problema primale e duale

- Le KKT del problema primale sono

1. $Ax^* = b$
2. $x^* \geq 0$
3. $c^T - u^{*T} A \geq 0^T$
4. $(c^T - u^{*T} A) \cdot x^* = 0$

- Le KKT del problema duale sono

1. $Av^* = b$
2. $v^* \geq 0$
3. $c^T - y^{*T} A \geq 0^T$
4. $(c^T - y^{*T} A) \cdot v^* = 0$

Se poniamo $v^* = x^*$ e $y^* = u^*$ allora le KKT dei problemi primale e duale coincidono ovvero i moltiplicatori di Lagrange all'ottimo del problema duale corrispondono con la soluzione ottima del problema primale e i moltiplicatori di Lagrange all'ottimo del problema primale corrispondono con la soluzione ottima del problema duale. Segue che

- le condizioni di ammissibilità per il problema primale sono

$$\begin{cases} Av^* = b \\ v^* \geq 0 \end{cases} \quad \begin{cases} Ax^* = b \\ x^* \geq 0 \end{cases}$$

- le condizioni di ammissibilità per il problema duale sono

$$c^T - y^{*T} A \geq 0^T \quad c^T - u^{*T} A \geq 0^T$$

- le condizioni di complementarità o di ortogonalità sono

$$(c^T - y^{*T} A) \cdot v^* = 0 \quad (c^T - u^{*T} A) \cdot v^* = 0$$

Quindi le KKT per il problema primale le posso scrivere con $y^* = u^*$ come

1. $Ax^* = b$
2. $x^* \geq 0$
3. $c^T - y^{*T} A \geq 0^T$
4. $(c^T - y^{*T} A) \cdot x^* = 0$

dove y^* è la soluzione ottima del problema primale. La condizione (4) la possiamo riscrivere componente per componente come

$$(c^T - y^{*T} A) \cdot x^* = 0 \quad \Longleftrightarrow \quad (c_j^T - y^{*T} A_j) \cdot x_j^* = 0 \quad j = 1, \dots, n$$

dove

$$A = \begin{bmatrix} A_1 & A_2 & \dots & A_j & \dots & A_n \end{bmatrix}$$

e $(c_j^T - y^{*T} A_j)$ è il j-esimo vincolo del problema duale.

5.5 Dualità debole

La proprietà di dualità debole è

$$b^T \bar{y} \leq c^T \bar{x} \quad \forall \bar{x}, \bar{y} \text{ ammissibili per il primale e per il duale}$$

Questa relazione afferma che le funzioni obiettivo del problema primale e del problema duale sono legate da una relazione di disuguaglianza. Infatti $c^T \bar{x}$ è il valore della funzione obiettivo del primale calcolato nel punto ammissibile $\bar{x}$ mentre $b^T \bar{y}$ è il valore della funzione obiettivo del duale calcolato nel punto ammissibile $\bar{y}$. Quindi il valore della funzione obiettivo del problema duale è sempre minore o uguale della funzione obiettivo del problema primale in un qualunque punto $\bar{x}$ e $\bar{y}$ ammissibili per il primale e per il duale.

5.5.1 Dimostrazione

Siano $\bar{x}$ e $\bar{y}$ soluzioni ammissibili (non necessariamente minimizzando la funzione obiettivo ma per essere ammissibile devono appartenere all'insieme ammissibile X) per i problemi primali e duali, rispettivamente. Appliciamo l'operatore di trasposizione ai vincoli che definiscono la regione ammissibile del problema duale

$$A^T \bar{y} \leq c \quad \Longleftrightarrow \quad (A^T \bar{y})^T \leq c^T \quad \Longleftrightarrow \quad \bar{y}^T A \leq c^T$$

Poi moltiplichiamo per $\bar{x}$ a destra e otteniamo

$$\bar{y}^T A \bar{x} \leq c^T \bar{x} \quad \Longleftrightarrow \quad \bar{y}^T b \leq c^T \bar{x} \quad \Longleftrightarrow \quad b^T \bar{y} \leq c^T \bar{x}$$

5.6 Dualità forte

Due soluzioni x^* e y^* ammissibili per il primale ed il duale, rispettivamente, sono soluzioni ottime del primale e del duale se e solo se

$$c^T x^* = b^T y^*$$

5.6.1 Dimostrazione 1

5.6.1.1 Minimo globale per il problema primale

Consideriamo la proprietà di dualità debole

$$b^T \bar{y} \leq c^T \bar{x} \quad \forall \bar{x}, \bar{y} \text{ ammissibili per il primale e per il duale}$$

Se y^* è soluzione ottima del problema duale, possiamo scrivere

$$c^T \bar{x} \geq b^T y^* = \max b^T \bar{y} \quad \forall \bar{x} \text{ ammissibile per il primale}$$

Se x^* è soluzione ammissibile del problema primale, tale che

$$c^T x^* = \max b^T y = b^T y^*$$

allora x^* è un minimo globale per il problema primale.

5.6.1.2 Minimo globale per il problema duale

Consideriamo la proprietà di dualità debole

$$b^T \bar{y} \leq c^T \bar{x} \quad \forall \bar{x}, \bar{y} \text{ ammissibili per il primale e per il duale}$$

Se x^* è soluzione ottima del problema primale, possiamo scrivere

$$b^T \bar{y} \leq c^T x^* = \min c^T x \quad \forall \bar{y} \text{ ammissibile per il duale}$$

Se y^* è soluzione ammissibile del problema duale, tale che

$$b^T y^* = \min c^T x = c^T x^*$$

allora y^* è un minimo globale per il problema duale.

5.6.2 Dimostrazione 2

La dimostrazione deriva dalle condizioni KKT, che verificano l'ottimalità. In particolare da

$$(c^T - y^{*T} A) \cdot x^* = 0$$

che risulta

$$c^T x^* = A^T x^{*T} y^*$$

e dato che $Ax = b$ segue che

$$c^T x^* = b^T y^*$$

5.7 Soluzioni ottime per il primale e per il duale

Le condizioni KKT sono condizioni necessarie e sufficienti di minimo globale anche per i punti non regolari. Segue che x^* e y^* sono soluzioni ottime del problema primale e duale, rispettivamente, se e solo se valgono le seguenti condizioni

1. $Ax^* = b$
2. $x^* \geq 0$
3. $A^T y^* \leq c$
4. $x^{*T} \cdot (c - A^T y^*) = (c^T - y^{*T} A) \cdot x^* = 0$

Le condizioni (1) e (2) sono le condizioni di ammissibilità per il primale, la (3) è la condizione di ammissibilità per il duale e la (4) è la condizione di complementarità o ortogonalità. La condizione (4) può essere scritta come

$$(c^T - y^{*T} A) \cdot x^* = 0 \quad \Longleftrightarrow \quad (c_j^T - y^{*T} A_j) \cdot x_j^* = 0$$

Se x^* e y^* sono soluzioni ottime, allora $x_j^* = 0$ e/o $c_j = y^{*T} A_j$ che equivale a $A_j^T y^* = c_j$

5.7.1 Teorema 26

Supponiamo di avere un problema primale e il corrispondente problema duale.

$$\begin{cases} \min c^T x \\ Ax = b \\ x \geq 0 \end{cases} \quad \begin{cases} \max b^T y \\ A^T y \leq c \end{cases}$$

Segue che, se uno dei due problemi è illimitato allora l'altro problema è impossibile cioè non ha soluzioni ammissibili.

5.7.1.1 Dimostrazione

Un problema è illimitato quando la funzione obiettivo può decrescere in modo indefinito rimanendo sempre nell'insieme ammissibile, dalla dualità debole si ha un assurdo e quindi il problema duale non ha una soluzione ammissibile. Si ha

$$\min c^T x \rightarrow -\infty \geq b^T y$$

Analogamente, assumiamo che un problema duale sia illimitato

$$\max b^T y \rightarrow +\infty \leq c^T x$$

ed è un assurdo per la dualità debole, il problema primale non ha soluzioni ammissibili.

5.8 Problema primale in forma canonica

Un problema di PL in forma canonica è espresso come

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

e il corrispondente problema duale è espresso come

$$\begin{cases} \max b^T y \\ A^T y \leq c \\ y \geq 0 \end{cases}$$

La matrice $A \in \mathbb{R}^{m \times n}$ può essere espressa come

$$A = \begin{bmatrix} A_1 & A_2 & \cdots & A_n \end{bmatrix} \quad \text{dove } A_i = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$$

5.8.1 Condizioni KKT del problema primale

Ripetendo i ragionamenti precedenti, le KKT per il primale sono

1. $Ax^* \geq b$
2. $x^* \geq 0$
3. $(a_i^T x^* - b_i) \cdot y_i^* = 0 \quad i = 1, \dots, n$

Le condizioni (1) e (2) sono di ammissibilità e la condizione (3) è di complementarità, all'ottimo $y_i^* = 0$ e/o l' i -esimo vincolo del primale deve essere soddisfatto all'uguaglianza).

5.8.2 Condizioni KKT del problema duale

Ripetendo i ragionamenti precedenti, le KKT per il duale sono

1. $A^T y^* \leq c$
2. $y^* \geq 0$
3. $(c_j - A_j^T y^*) \cdot x_j^* = 0 \quad j = 1, \dots, n$

Le condizioni (1) e (2) sono di ammissibilità e la condizione (3) è di complementarità, all'ottimo $x_j^* = 0$ e/o l' j -esimo vincolo del duale deve essere soddisfatto all'uguaglianza).

5.9 Problema duale associato a un problema primale

Problema primale	Problema duale
min	max

Segno dei vincoli	Segno delle variabili
$a_i^T x \geq b_i$	$y_i \geq 0$
$a_i^T x \leq b_i$	$y_i \leq 0$
$a_i^T x = b_i$	y_i libera

Segno delle variabili	Segno dei vincoli
$x_j \geq 0$	$A_j^T y \leq c_j$
$x_j \leq 0$	$A_j^T y \geq c_j$
x_j libera	$A_j^T y = c_j$

Funzione obiettivo	Funzione obiettivo
$c^T x$	$b^T y$

Possiamo ricavare il problema duale associato a un problema primale tramite le regole di trasformazione, la tabella si legge da sinistra verso destra se si ha un problema primale in forma di minimo e da destra verso sinistra se si ha un problema primale in forma di massimo. Per verificare se la trasformazione è corretta, è necessario calcolare il duale del duale il quale deve essere uguale al primale di partenza.

Ad esempio, consideriamo un problema di PL in forma canonica. Se il primale è

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

allora il duale è

$$\begin{cases} \max b^T y \\ A^T y \leq c \\ y \geq 0 \end{cases}$$

Quindi i coefficienti di costo delle variabili della funzione obiettivo del problema duale è il vettore b che è il vettore dei termini noti che definiscono i vincoli del problema primale, i coefficienti di costo delle variabili della funzione obiettivo del problema primale è il vettore c che è il vettore dei termini noti che definiscono i vincoli del problema duale. Questo ragionamento vale qualsiasi sia il problema primale, anche non in forma canonica.

5.9.1 Esempio

Supponiamo di avere un problema di PL. Determinare il corrispondente problema duale.

$$\left\{ \begin{array}{l} \min 10x_1 + 20x_2 + 30x_3 \\ 2x_1 - x_2 \geq 1 \\ x_2 + x_3 \leq 2 \\ x_1 - x_3 = 3 \\ x_1 \geq 0 \\ x_2 \leq 0 \\ x_3 \text{ libera} \end{array} \right.$$

Studiamo il segno dei vincoli e delle variabili

$$\left\{ \begin{array}{ll} \min 10x_1 + 20x_2 + 30x_3 & \\ 2x_1 - x_2 \geq 1 & \implies y_1 \geq 0 \\ x_2 + x_3 \leq 2 & \implies y_2 \leq 0 \\ x_1 - x_3 = 3 & \implies y_3 \text{ libera} \\ x_1 \geq 0 & \implies A_1^T y \leq c_1 \\ x_2 \leq 0 & \implies A_2^T y \geq c_2 \\ x_3 \text{ libera} & \implies A_3^T y = c_3 \end{array} \right.$$

Quindi

$$A = \begin{bmatrix} A_1 & A_2 & A_3 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & -1 \end{bmatrix} \quad b = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad c = \begin{bmatrix} 10 \\ 20 \\ 30 \end{bmatrix} \quad y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

Segue che il problema duale è

$$\left\{ \begin{array}{l} \max y_1 + 2y_2 + 3y_3 \\ 2y_1 + y_3 \leq 10 \\ -y_1 + y_2 \geq 20 \\ y_2 - y_3 = 30 \\ y_1 \geq 0 \\ y_2 \leq 0 \\ y_3 \text{ libera} \end{array} \right.$$

5.10 Trasformazione dei vincoli

Siano $x_1, \dots, x_n$ le variabili, $a_{i1}, \dots, a_{in}$ i coefficienti delle variabili e b_i il termine noto. I vincoli che si possono trasformare sono

- Vincolo $\geq$

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

Dobbiamo trasformare il vincolo di $\geq$ in un vincolo di uguaglianza, introducendo una variabile di surplus $s_i \geq 0$. Si ha

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n - s_i = b_i$$

- Vincolo $\leq$

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

Dobbiamo trasformare il vincolo di $\leq$ in un vincolo di uguaglianza, introducendo una variabile di slack $s_i \geq 0$. Si ha

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n + s_i = b_i$$

- Vincolo $=$

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

è equivalente al sistema

$$\begin{cases} a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n \geq b_i \\ a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n \leq b_i \end{cases}$$

Se abbiamo un problema di PL e la variabile x_i è libera, allora possiamo trasformare la variabile x_i in un sistema in cui tutte le variabili sono non negative. Introducendo le variabili s_i^+ e s_i^- segue che

$$\begin{cases} x_i = s_i^+ - s_i^- \\ s_i^+ \geq 0 \\ s_i^- \geq 0 \end{cases}$$

5.11 Problema di allocazione di risorse

Consideriamo

- n prodotti: $j = 1, \dots, n$
- m risorse: $i = 1, \dots, m$
- a_{ij} : quantità di risorsa i necessaria a produrre un'unità di prodotto j
- b_i : disponibilità di risorsa i
- c_j : profitto derivante dalla produzione di un'unità di prodotto j

Il problema è decidere la produzione di ciascun prodotto j in modo tale che

- non si eccede il consumo di ciascuna risorsa i
- il profitto totale sia massimo

Quando dobbiamo formulare il problema come un problema di PL devo implicitamente fare delle assunzioni che sono

1. proporzionalità: ad esempio se $c_1 = 3$ e produciamo 2 unità, il profitto è $2 \cdot c_1 = 3$
2. continuità: le variabili del modello assumono tutti i valori reali in un dato intervallo
3. additività

Prodotto	a_{ij}
1	3
2	4

5.11.1 Esempio

Consideriamo un'azienda che produce 2 coloranti c_1, c_2 con 3 sostanze s_1, s_2, s_3

- Matrice di produzione (coefficienti a_{ij})

Litri	c_1	c_2
s_1	1	1
s_2	1	2
s_3	0	1

- Disponibilità delle sostanze

Litri	s_1	s_2	s_3
b_i	750	1000	400

- Profitto dalla vendita dei coloranti

Euro/Litro	c_1	c_2
c_j	7	10

Quanti coloranti di ogni tipo bisogna produrre, in modo tale da non eccedere la disponibilità delle risorse e da massimizzare il profitto?

- Variabili
 - litri di c_1 da produrre: x_1
 - litri di c_2 da produrre: x_2
- Ipotesi di proporzionalità
 - profitto di x_1 : $7x_1$

- profitto di x_2 : $10x_2$
- Ipotesi di additività
 - profitto totale: $7x_1 + 10x_2$
- Funzione obiettivo
 - $\max 7x_1 + 10x_2$
- Consumi
 - s_1 : $x_1 + x_2 \leq 750$
 - s_2 : $x_1 + 2x_2 \leq 1000$
 - s_3 : $x_2 \leq 400$
- Ipotesi di continuità
 - condizione $x_1 \geq 0$
 - condizione $x_2 \geq 0$

5.12 Problema di miscelazione

Consideriamo

- n sostanze contenenti elementi "utili": $j = 1, \dots, n$
- m elementi "utili": $i = 1, \dots, m$
- a_{ij} : quantità "utile" i presente in un'unità di sostanza j
- b_i : quantità minima di elemento utile i presente nella combinazione (miscela) di sostanze
- c_j : costo unitario della sostanza j

Il problema è determinare la quantità di ciascuna sostanza da inserire nella miscela in modo tale che

- la miscela contenga almeno una quantità di elemento utile i pari a b_i
- il costo complessivo sia minimo

5.12.1 Esempio

Le sostanze sono gli alimenti e gli elementi utili sono le caratteristiche degli elementi (calorie, proteine, grassi, ...). Si ha

	Pane	Latte	Uova	Carne	Dolce
Calorie (cal)	110	160	180	260	420
Proteine (gr)	4	8	13	14	4
Calcio (mg)	2	285	54	80	22
Costo (euro)	2	3	4	19	20
# massimo di porzioni	4	8	3	2	2

I requisiti minimi degli elementi utili sono

$$\begin{cases} \text{calorie} \geq 2000 \text{ cal} \\ \text{proteine} \geq 50 \text{ gr} \\ \text{calcio} \geq 700 \text{ mg} \end{cases}$$

Determinare il numero di porzioni di ciascun elemento in modo da rispettare i requisiti minimi e minimizzare i costi. Si ha

- x_j : numero di porzioni di alimento j nella dieta

- Ipotesi di continuità

$$x_j \geq 0 \quad j = 1, \dots, 5$$

- funzione obiettivo

$$\min 2x_1 + 3x_2 + 4x_3 + 19x_4 + 20x_5$$

- quantità di calorie

$$110x_1 + 160x_2 + 180x_3 + 260x_4 + 420x_5 \geq 2000 \text{ cal}$$

- quantità di proteine

$$4x_1 + 8x_2 + 13x_3 + 14x_4 + 4x_5 \geq 50 \text{ gr}$$

- quantità di calcio

$$2x_1 + 285x_2 + 54x_3 + 80x_4 + 22x_5 \geq 700 \text{ mg}$$

- condizioni

$$\begin{cases} x_1 \leq 4, & x_1 \geq 0 \\ x_2 \leq 8, & x_2 \geq 0 \\ x_3 \leq 3, & x_3 \geq 0 \\ x_4 \leq 5, & x_4 \geq 0 \\ x_5 \leq 2, & x_5 \geq 0 \end{cases}$$

5.13 Problema di trasposto

Consideriamo

- m origini: $i = 1, \dots, m$
- n destinazioni: $j = 1, \dots, n$
- a_i : bene disponibile nell'origine i
- b_j : quantità di bene richiesta da j
- c_{ij} : costo per trasportare un'unità di bene da i a j

Il problema è trasportare il bene dalle origini alle destinazioni in modo tale che

- ad ogni destinazione arrivi una quantità almeno pari a b_j
- da ogni origine porto al massimo una quantità a_i
- il costo complessivo di trasporto sia minimo

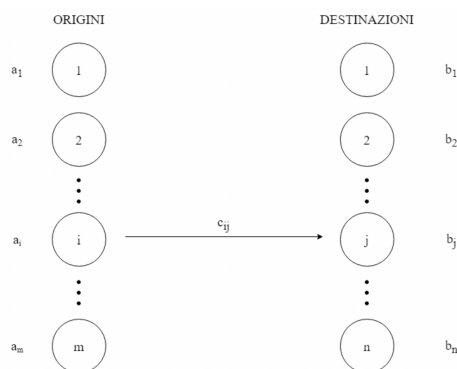
Definiamo x_{ij} come la quantità di bene trasportato da i a j . La formulazione è

$$\min \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij}$$

I vincoli sono

$$\begin{cases} \sum_{i=1}^m x_{ij} \geq b_j & j = 1, \dots, n \\ \sum_{j=1}^n x_{ij} \leq a_i & i = 1, \dots, m \\ x_{ij} \geq 0 & i = 1, \dots, m \text{ e } j = 1, \dots, n \end{cases}$$

Il problema può essere rappresentato graficamente come



Capitolo 6

Programmazione Lineare con Vertici

6.1 Introduzione

6.1.1 Poliedro

Consideriamo un problema di PL in forma standard

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

un vettore α e uno scalare α_0 definiti come

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \in \mathbb{R}^n \quad \alpha_0 \in \mathbb{R}$$

Segue che

- $\alpha^T x \leq \alpha_0$ è un semispazio affine
- $\alpha^T x = \alpha_0$ è un semipiano indotto da α e α_0

Il poliedro è dato dall'intersezione di un certo numero di semispazi e semipiani, consideriamo per esempio

$$X \equiv \begin{cases} Ax = b & m \text{ semipiani} \\ -x \leq 0 & n \text{ semispazi} \end{cases}$$

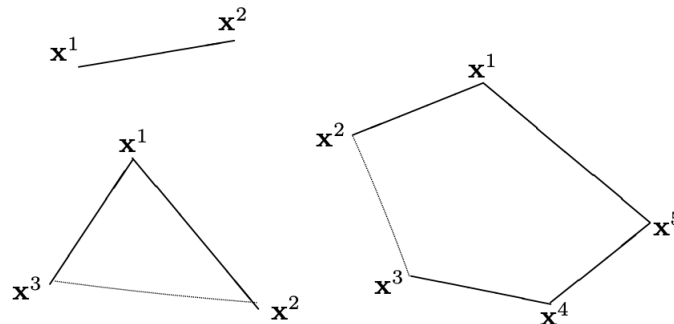
Un poliedro è un insieme convesso, esso è formato da un insieme di equazioni lineari e di disequazioni lineari (definite da funzioni lineari che sono sia concave che convesse).

$$X \equiv \begin{cases} Ax = b & \text{equazioni lineari} \\ x \geq 0 & \text{disequazioni lineari} \end{cases} \implies X \text{ è un insieme convesso}$$

6.1.2 Politopo

Un politopo è un poliedro limitato P cioè

$$\exists M : \|x\| \leq M \quad \forall x \in P$$



6.1.3 Vertice o punto estremo di un poliedro

Un vertice o un punto estremo di un poliedro P è un punto $x \in P$ se non si può esprimere come combinazione convessa stretta di altri due punti del poliedro P cioè se

$$\nexists y, z \in P : x = \lambda y + (1 - \lambda)z \quad \lambda \in (0, 1)$$

6.2 Teorema di Minkowski-Weyl

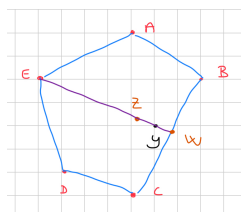
Ogni punto di un politopo P può essere espresso come combinazione convessa dei suoi vertici. Supponiamo $x^1, x^2, \dots, x^k \in \mathbb{R}^n$ vertici di P e allora

$$y = \lambda_1 x^1 + \lambda_2 x^2 + \dots + \lambda_k x^k \quad \forall y \in P$$

con $\lambda_1, \lambda_2, \dots, \lambda_k \geq 0$ e $\sum_{i=1}^k \lambda_i = 1$

6.2.1 Dimostrazione

Ad esempio, consideriamo il seguente politopo e i suoi vertici.



Segue che il punto W lo possiamo esprimere come combinazione convessa dei vertici B e C perché giace sul segmento che unisce B e C . Il segmento EW lo possiamo esprimere come combinazione convessa del vertice E e del punto W , il punto y lo possiamo esprimere come combinazione convessa del vertice E e del punto W , il quale può essere a sua volta espresso dai vertici B e C .

6.3 Teorema 27

Consideriamo il seguente problema di PL

$$\begin{cases} \min c^T x \\ x \in P \end{cases}$$

dove P è un politopo. Se un problema di PL ha soluzione ottima allora esiste un vertice del politopo che è soluzione ottima del problema di PL.

6.3.1 Dimostrazione

Supponiamo che $x^1, x^2, \dots, x^k \in \mathbb{R}^k$ siano i vertici di P . Sia

$$z^* = \min\{c^T x^i : i = 1, \dots, k\}$$

La tesi è

$$\forall y \in P \quad c^T y \geq z^*$$

La dimostrazione parte da $y \in P$ quindi per il Teorema di Minkowski-Weyl si ha

$$y = \sum_{i=1}^k \lambda_i x^i \quad \text{con } \lambda_1, \lambda_2, \dots, \lambda_k \geq 0 \quad \text{e} \quad \sum_{i=1}^k \lambda_i = 1$$

Segue che

$$c^T y = c^T \sum_{i=1}^k \lambda_i x^i = \sum_{i=1}^k (\lambda_i c^T x^i)$$

e dato che $c^T x^i \geq z^*$ si ha

$$\sum_{i=1}^k (\lambda_i c^T x^i) \geq \sum_{i=1}^k (\lambda_i z^*) = z^* \cdot \sum_{i=1}^k \lambda_i = z^* \cdot 1 = z^*$$

6.4 Sistema ammette un'unica soluzione

Consideriamo un problema di PL in forma standard e sia P un poliedro.

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

quindi abbiamo m equazioni in n variabili. L'ipotesi è

$$m \leq n \quad \text{e} \quad \text{rango}(A) = m \quad \Longleftrightarrow \quad A \text{ ha rango pieno}$$

Se annulliamo $n-m$ variabili, allora il sistema $Ax = b$ diventa un sistema di m equazioni in m variabili. Quindi se le colonne (o le righe) dei coefficienti delle variabili che definiscono il sistema sono linearmente indipendenti, allora il sistema ammette un'unica soluzione.

6.5 Base di una matrice

La base di una matrice è l'insieme di m colonne della matrice A linearmente indipendenti. Consideriamo il sistema $Ax = b$

$$A = \begin{bmatrix} A_1 & A_2 & \cdots & A_n \end{bmatrix} \in \mathbb{R}^{m \times n}$$

$$Ax = b \implies A_1x_1 + A_2x_2 + \dots + A_nx_n = b$$

Quando annulliamo $n - m$ variabili della matrice A , il sistema $Ax = b$ ha un'unica soluzione se le colonne associate alle variabili della matrice A che non annulliamo sono corrispondenti a m colonne di A linearmente indipendenti ovvero le colonne di A formano una base di A . Se

$$A_{\beta(1)}, A_{\beta(2)}, \dots, A_{\beta(m)}$$

sono linearmente indipendenti, allora

$$B = (A_{\beta(1)}, A_{\beta(2)}, \dots, A_{\beta(m)}) \in \mathbb{R}^{m \times m}$$

è una base di A dove

$$x_{\beta(1)}, x_{\beta(2)}, \dots, x_{\beta(m)}$$

sono m variabili in base e

$$\{x_1, x_2, \dots, x_n\} \setminus \{x_{\beta(1)}, x_{\beta(2)}, \dots, x_{\beta(m)}\}$$

sono $n - m$ variabili fuori base. Inoltre

$$\{A_1, A_2, \dots, A_n\} \setminus \{A_{\beta(1)}, A_{\beta(2)}, \dots, A_{\beta(m)}\}$$

sono le colonne di A fuori base. Se annulliamo le variabili fuori base allora possiamo determinare in modo univoco, dato che $\det(B) \neq 0$, le variabili in base.

6.6 Soluzione di base e soluzione di base ammissibile

Consideriamo il sistema $Ax = b$ e supponiamo $B \in \mathbb{R}^{m \times m}$ base di A . La matrice A è formata da due parti

$$A = \begin{bmatrix} B & F \end{bmatrix}$$

dove $B \in \mathbb{R}^{m \times m}$ sono le colonne in base e $F \in \mathbb{R}^{m \times (n-m)}$ sono le colonne fuori base. Inoltre anche il vettore delle variabili è formato da due parti

$$x = \begin{bmatrix} x_B \\ x_F \end{bmatrix}$$

dove x_B sono le variabili associate alle colonne in base e x_F sono le variabili associate alle colonne fuori base. Quindi l'equazione la possiamo scrivere come

$$Ax = b \iff \begin{bmatrix} B & F \end{bmatrix} \begin{bmatrix} x_B \\ x_F \end{bmatrix} = b \iff Bx_B + Fx_F = b$$

Assumiamo che $\det(B) \neq 0$, allora esiste l'inversa della matrice B cioè B^{-1} e moltiplichiamo il sistema a sinistra per B^{-1} ottenendo

$$B^{-1}(Bx_B + Fx_F) = B^{-1}b$$

$$x_B + B^{-1}Fx_F = B^{-1}b \quad \implies \quad x_B = B^{-1}b - B^{-1}Fx_F$$

La soluzione in cui $x_B = B^{-1}b$ e $x_F = 0$ è chiamata soluzione di base. Se

$$x_B = B^{-1}b \geq 0$$

allora la soluzione di base è ammissibile per il sistema

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

ed è chiamata soluzione di base ammissibile, essa corrisponde a

$$x = \begin{bmatrix} x_B \\ x_F \end{bmatrix} = \begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix}$$

che è un punto di P e la base B si dice ammissibile.

6.7 Teorema 28

Il punto $x \in P$ è un vertice di P se e solo se x è soluzione ammissibile di base.

6.7.1 Dimostrazione $\Leftarrow$

- Ipotesi: x è soluzione di base ammissibile
- Tesi: x è un vertice di P

Per assurdo, supponiamo che x non sia un vertice. Consideriamo

$$x = \begin{bmatrix} x_B \\ x_F \end{bmatrix} = \begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix} = \begin{bmatrix} x_1 \\ \vdots \\ x_k \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

dove $B^{-1}b \in \mathbb{R}^m$ e $x_1, \dots, x_k$ sono le componenti positive $k \geq m$ perché così otteniamo un punto x come combinazione di altri due punti. Se x non è un vertice

$$\exists y, z \in P, y \neq z : x = \lambda y + (1 - \lambda)z \quad \lambda \in (0, 1)$$

cioè x è all'interno del segmento che unisce i punti y e z . Poiché

$$y, z \in P \equiv \begin{cases} Ax = b \\ x \geq 0 \end{cases} \implies y \geq 0 \quad \text{e} \quad z \geq 0$$

Inoltre x è ottenuto come combinazione convessa dei punti y e z infatti

$$x = \lambda y + (1 - \lambda)z$$

e dato che

$$\lambda > 0 \implies (1 - \lambda) > 0 \quad \text{e} \quad y \geq 0 \quad \text{e} \quad z \geq 0$$

allora

$$\lambda y \geq 0 \quad \text{e} \quad (1 - \lambda)z \geq 0$$

Segue che

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_k \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_k \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad z = \begin{bmatrix} z_1 \\ \vdots \\ z_k \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Poiché

$$y, z \in P \equiv \begin{cases} Ax = b \\ x \geq 0 \end{cases} \implies Ay = b \quad \text{e} \quad Az = b$$

dove

$$A = \begin{bmatrix} A_1 & \dots & A_k & 0 & \dots & 0 \end{bmatrix}$$

Si ha

$$Ay = b \implies A_1 y_1 + A_2 y_2 + \dots + A_k y_k = b$$

$$Az = b \implies A_1 z_1 + A_2 z_2 + \dots + A_k z_k = b$$

$$\begin{array}{r} A_1 y_1 + A_2 y_2 + \dots + A_k y_k = b \\ - \quad A_1 z_1 + A_2 z_2 + \dots + A_k z_k = b \\ \hline A_1 \underbrace{(y_1 - z_1)}_{\alpha_1} + A_2 \underbrace{(y_2 - z_2)}_{\alpha_2} + \dots + A_k \underbrace{(y_k - z_k)}_{\alpha_k} = 0 \end{array}$$

Poiché $y \neq z$ allora $\alpha_1, \alpha_2, \dots, \alpha_k$ non sono tutti nulli,

$$A_1 \alpha_1 + A_2 \alpha_2 + \dots + A_k \alpha_k = 0$$

e implica che $A_1, A_2, \dots, A_k$ sono linearmente dipendenti. Questo è un assurdo dato che abbiamo supposto che $A_1, A_2, \dots, A_k$ erano le colonne associate ad una base di A .

6.7.2 Dimostrazione $\Rightarrow$

- Ipotesi: x è un vertice di P
- Tesi: x è soluzione di base ammissibile

Per assurdo, supponiamo che x non sia soluzione di base ammissibile. Se x è un vertice allora x ha almeno $n - m$ componenti nulle. Abbiamo

$$x = \begin{bmatrix} x_1 & \cdots & x_k & 0 & \cdots & 0 \end{bmatrix}^T$$

dove $x_1, \dots, x_k$ sono le componenti positive con $k \geq m$. Se x non è SBA (soluzione di base ammissibile) allora le colonne $A_1, \dots, A_k$ sono linearmente dipendenti. Se $A_1, \dots, A_k$ fossero colonne linearmente indipendenti potremmo aggiungere $n - k$ colonne linearmente indipendenti della matrice A e ottenere così una base, quindi x sarebbe una SBA. Segue che esistono dei coefficienti $\alpha_1, \dots, \alpha_k$ non tutti nulli tale che

$$A_1\alpha_1 + A_2\alpha_2 + \dots + A_k\alpha_k = 0$$

per definizione di vettore linearmente indipendenti. Poiché

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

si ha

$$Ax = b \quad \Rightarrow \quad A_1x_1 + A_2x_2 + \dots + A_kx_k = b$$

e moltiplicando per ε

$$A_1\varepsilon x_1 + A_2\varepsilon x_2 + \dots + A_k\varepsilon x_k = 0$$

$$A_1x_1 + A_2x_2 + \dots + A_kx_k = b$$

$$+ \quad A_1\varepsilon x_1 + A_2\varepsilon x_2 + \dots + A_k\varepsilon x_k = 0$$

$$A_1(x_1 + \varepsilon\alpha_1) + A_2(x_2 + \varepsilon\alpha_2) + \dots + A_k(x_k + \varepsilon\alpha_k) = b$$

$$A_1x_1 + A_2x_2 + \dots + A_kx_k = b$$

$$- \quad A_1\varepsilon x_1 + A_2\varepsilon x_2 + \dots + A_k\varepsilon x_k = 0$$

$$A_1(x_1 - \varepsilon\alpha_1) + A_2(x_2 - \varepsilon\alpha_2) + \dots + A_k(x_k - \varepsilon\alpha_k) = b$$

Si ottiene

$$y = \begin{bmatrix} x_1 + \varepsilon\alpha_1 \\ x_2 + \varepsilon\alpha_2 \\ \vdots \\ x_k + \varepsilon\alpha_k \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad z = \begin{bmatrix} x_1 - \varepsilon\alpha_1 \\ x_2 - \varepsilon\alpha_2 \\ \vdots \\ x_k - \varepsilon\alpha_k \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \Rightarrow \quad \begin{cases} Ay = b \\ Az = b \end{cases}$$

$$\exists \varepsilon \text{ sufficientemente piccolo} : \begin{cases} y \geq 0 \\ z \geq 0 \end{cases} \implies y, z \in P$$

Inoltre possiamo scrivere x come

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

ovvero x è esprimibile come combinazione convessa (con $\lambda = 1/2$) di y e z . Questo è assurdo perché x è un vertice e avevamo assunto che x non fosse SBA.

6.8 Riassumendo

Consideriamo un problema di PL in forma standard

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

Allora

- se esiste soluzione ottima del problema (cioè se l'insieme ammissibile non è vuoto oppure se il problema non è illimitato), allora esiste un vertice di P ottimo
- si ha che $\bar{x} \in P$ è vertice se e solo se $\bar{x}$ è soluzione di base ammissibile (SBA) cioè esiste una base B di A tale che

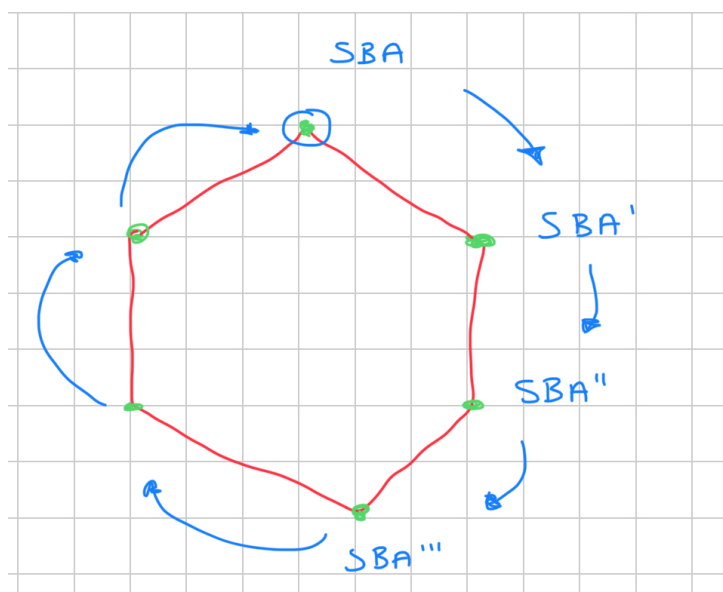
$$\bar{x} = \begin{bmatrix} x_B \\ x_F \end{bmatrix} = \begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix}$$

Capitolo 7

Metodo del Simpleso

7.1 Introduzione

L'idea del metodo del simpleso è quella di visitare i vertici di un poliedro ovvero visitare le soluzioni di base ammissibili associate al poliedro finché non troviamo la soluzione ottima. Per fare ciò dobbiamo enumerare tutti i vertici del poliedro (SBA) ed individuare il vertice ottimo ovvero quello con costo minore.



Vedremo che questo metodo, così come descritto, non è molto efficiente perché i vertici di un poliedro possono essere tantissimi. Sappiamo che $A \in \mathbb{R}^{m \times m}$ e una base $B \in \mathbb{R}^{m \times m}$ quindi il numero massimo di basi ammissibili è

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

Possiamo migliorare questa idea con

- un test per controllare se una SBA (vertice) è ottima oppure no
- un modo per passare da una SBA (vertice) non ottima ad una SBA "vicina" senza peggiorare il valore della funzione obiettivo.

7.2 Test di ottimalità per una SBA

Sia B una base ammissibile della matrice A . La SBA

$$x = \begin{bmatrix} x_B \\ x_F \end{bmatrix} = \begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix}$$

è ottima? Devo trasformare il problema di PL

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

rispetto alla base B dove

$$\begin{aligned} A &= \begin{bmatrix} B & F \end{bmatrix} & B &\in \mathbb{R}^{m \times m}, F \in \mathbb{R}^{m \times (n-m)} \\ x &= \begin{bmatrix} x_B \\ x_F \end{bmatrix} & Ax = b &\implies Bx_B + Fx_F = b \\ c^T &= \begin{bmatrix} c_B & c_F \end{bmatrix} & c^T x &\implies c_B^T x_B + c_F^T x_F \end{aligned}$$

Data la base B ammissibile, possiamo riscrivere il problema come

$$\begin{cases} \min c_B^T x_B + c_F^T x_F \\ Bx_B + Fx_F = b \\ x_B \geq 0 \\ x_F \geq 0 \end{cases}$$

Dato che

$$\det(B) \neq 0 \implies \exists B^{-1} \implies x_B = B^{-1}b - B^{-1}Fx_F$$

possiamo riscrivere la funzione obiettivo

$$c_B^T x_B + c_F^T x_F = c_B^T \cdot (B^{-1}b - B^{-1}Fx_F) + c_F^T x_F = c_B^T B^{-1}b + (c_F^T - c_B^T B^{-1}F) \cdot x_F$$

dove $c_B^T B^{-1}b$ è un termine costante e rappresenta il valore della funzione obiettivo in corrispondenza della SBA

$$x = \begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix}$$

Definiamo il vettore dei costi ridotti delle variabili fuori base come

$$\bar{c}_F^T = (c_F^T - c_B^T B^{-1}F)$$

7.2.1 Teorema 29

Sia B una base ammissibile. Se $\bar{c}_F^T \geq 0$ allora la SBA

$$x = \begin{bmatrix} x_B \\ x_F \end{bmatrix} = \begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix}$$

è ottima.

7.2.1.1 Dimostrazione

La funzione obiettivo la possiamo esprimere come

$$c^T x = c_B^T B^{-1} b + (c_F^T - c_B^T B^{-1} F) \cdot x_F = c_B^T B^{-1} b + \bar{c}_F^T \cdot x_F$$

Consideriamo un punto $\bar{x} \in P$ e calcoliamo

$$c^T \tilde{x} = c_B^T B^{-1} b + \bar{c}_F^T \cdot \tilde{x}_F$$

Per ipotesi ho che $\bar{c}_F^T \geq 0$ e poiché

$$\tilde{x} \in P \implies \tilde{x} \geq 0 \implies \tilde{x}_F \tilde{x} \geq 0 \implies \bar{c}_F^T \cdot x_F \geq 0$$

ed eliminando questo termine dall'uguaglianza

$$c^T \tilde{x} \geq c_B^T B^{-1} b$$

dove $c^T \tilde{x}$ è il valore della funzione obiettivo in $\tilde{x}$ e

$$c_B^T B^{-1} b$$

è la funzione obiettivo calcolata nella SBA

$$x = \begin{bmatrix} B^{-1} b & 0 \end{bmatrix}$$

Quindi ho dimostrato che la SBA è soluzione ottima per il problema di PL cioè non esiste un altro punto $\tilde{x} \in P$ che ha valore della funzione obiettivo che è inferiore al valore della funzione obiettivo calcolata nella SBA perché il vettore dei costi ridotti delle variabili fuori base, quando esprimiamo la funzione obiettivo in funzione della base B e delle variabili fuori base, è non negativo. Data una base B il vettore dei costi ridotti è definito come

$$\begin{aligned} \bar{c}^T &= c^T - c_B^T B^{-1} A = \begin{bmatrix} c_B^T & c_F^T \end{bmatrix} - c_B^T B^{-1} \begin{bmatrix} B & F \end{bmatrix} = \\ &= \begin{bmatrix} c_B^T & c_F^T \end{bmatrix} - \begin{bmatrix} c_B^T & c_B^T B^{-1} F \end{bmatrix} = \begin{bmatrix} 0 & c_F^T - c_B^T B^{-1} F \end{bmatrix} \end{aligned}$$

dove

$$\bar{c}_F^T = c_F^T - c_B^T B^{-1} F$$

è il vettore dei costi ridotti delle variabili fuori base e

$$\bar{c}_B^T = 0$$

è il vettore dei costi ridotti delle variabili in base.

7.3 Operazione di Pivot

Consideriamo il seguente problema di PL in forma standard e supponiamo di avere una base ammissibile B non ottima. Trasformiamo il problema di PL in forma canonica rispetto alla base B , la quale esprime le variabili x_B e la funzione obiettivo in funzione delle variabili fuori base

$$\begin{cases} x_B = B^{-1} b - B^{-1} F x_F \\ c^T x = c_B^T B^{-1} b + (c_F^T - c_B^T B^{-1} F) \cdot x_F \end{cases}$$

dove

$$\bar{c}_F^T = (c_F^T - c_B^T B^{-1} F)$$

è il vettore dei costi ridotti di x_F . Definiamo $\bar{c}_h < 0$ il coefficiente di costo ridotto della variabile fuori base h quindi

$$x_F = \begin{bmatrix} 0 & \cdots & 0 & x_h & 0 & \cdots & 0 \end{bmatrix}^T$$

e se aumentiamo la variabile x_h . Allora

$$c^T x = c_B^T B^{-1} b + \bar{c}_F^T x_F = c_B^T B^{-1} b + \bar{c}_h x_h$$

e dato che $\bar{c}_h < 0$ allora devo ottenere $x_h > 0$ così che

$$\bar{c}_h x_h < 0$$

e quindi aumentando x_h la funzione obiettivo decresce. Quale è il valore massimo di x_h ?

$$x_B = B^{-1} b - B^{-1} F x_F$$

ma se

$$x_F = \begin{bmatrix} 0 & \cdots & 0 & x_h & 0 & \cdots & 0 \end{bmatrix}^T$$

allora

$$F x_F = A_h x_h \implies x_B = B^{-1} b - B^{-1} A_h x_h$$

e l'aumento di x_h deve soddisfare la condizione

$$x_B = B^{-1} b - B^{-1} A_h x_h \geq 0$$

Definisco

$$\bar{b} := B^{-1} b$$

e

$$\bar{A}_h := B^{-1} A_h \implies x_B = \bar{b} - \bar{A}_h x_h \geq 0$$

Supponiamo

$$B = (A_{\beta(1)}, A_{\beta(2)}, \dots, A_{\beta(m)})$$

e

$$x_B = \begin{bmatrix} x_{\beta(1)} & x_{\beta(2)} & \cdots & x_{\beta(m)} \end{bmatrix}^T$$

Si ha

$$B^{-1} b = \bar{b} = \begin{bmatrix} \bar{b}_1 \\ \bar{b}_2 \\ \vdots \\ \bar{b}_m \end{bmatrix} \quad \bar{A}_h = \begin{bmatrix} a_{1h} \\ a_{2h} \\ \vdots \\ a_{mh} \end{bmatrix}$$

Possiamo scrivere

$$x_B = \bar{b} - \bar{A}_h x_h \geq 0$$

in forma estesa come

$$\begin{cases} x_{\beta_1} = \bar{b}_1 - a_{1h} x_h \geq 0 \\ x_{\beta_2} = \bar{b}_2 - a_{2h} x_h \geq 0 \\ \vdots \\ x_{\beta_m} = \bar{b}_m - a_{mh} x_h \geq 0 \end{cases} \implies x_{\beta(i)} = \bar{b}_i - a_{ih} x_h \geq 0$$

7.3.1 Caso 1

Consideriamo come caso 1: $\bar{a}_{ih} \leq 0$ e dividiamo in

- se $\bar{a}_{ih} = 0$ allora $\bar{b}_i \geq 0$ è sempre verificata perché $B^{-1}b \geq 0$
- se $\bar{a}_{ih} < 0$ allora

$$x_h = \frac{\bar{b}_i}{\bar{a}_{ih}} > 0$$

e abbiamo una limitazione per x_h

7.3.1.1 Caso particolare

$$\begin{cases} x_{\beta_1} = \bar{b}_1 - \bar{a}_{1h}x_h \geq 0 \\ x_{\beta_2} = \bar{b}_2 - \bar{a}_{2h}x_h \geq 0 \\ \vdots \\ x_{\beta_m} = \bar{b}_m - \bar{a}_{mh}x_h \geq 0 \end{cases}$$

Se

$$\begin{cases} \bar{a}_{1h} \leq 0 \\ \bar{a}_{2h} \leq 0 \\ \vdots \\ \bar{a}_{mh} \leq 0 \end{cases}$$

allora il sistema è soddisfatto $\forall x_h > 0$ quindi soddisfa che x_h può crescere in modo indefinito, mantenendo l'ammissibilità, e si dice che il problema è illimitato cioè non esiste una soluzione ottima finita. Il valore massimo per x_h è definito da

$$\vartheta = \min \left\{ \frac{\bar{b}_i}{\bar{a}_{ih}} : \bar{a}_{ih} > 0 \quad i \in \{\beta(1), \dots, \beta(m)\} \right\}$$

Se ϑ lo troviamo in corrispondenza al valore relativo alla t -esima disequazione

$$x_h = \vartheta = \frac{\bar{b}_t}{\bar{a}_{th}} \geq 0$$

succede che una riga di pivot x_{β_t} è

$$x_{\beta_t} = \bar{b}_t - \bar{a}_{th}x_h = \bar{b}_t - \bar{a}_{th}\frac{\bar{b}_t}{\bar{a}_{th}} = 0$$

Allora x_{β_t} diventa una variabile fuori base, si ha che $A_{\beta(t)}$ esce dalla base, x_h diventa variabile in base e A_h entra in base.

7.3.1.2 Caso particolare: valore massimo ϑ

Consideriamo

$$x_h = \vartheta = \frac{\bar{b}_t}{\bar{a}_{th}} \geq 0$$

Se $\bar{b}_t = 0$ allora esiste una variabile in base uguale a 0, implica

$$\vartheta = 0 \implies x_h = 0$$

Quindi x_h diventa una variabile di base con valore 0 ma anche $x_{\beta(t)}$ era in base con valore 0. Questo si chiama "operazione di pivot degenerare" e si ha quando $B^{-1}b$ ha qualche componente nulla, segue che

$$x = \begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix}$$

si dice SBA degenerare e B si dice base degenerare.

7.3.2 Caso 2

Consideriamo come caso 2: $\bar{a}_{ih} > 0$ allora

$$x_h \leq \frac{\bar{b}_i}{\bar{a}_{ih}} > 0$$

e abbiamo una limitazione per x_h

7.3.3 Teorema 30

La nuova base

$$\tilde{B} = (A_{\beta(1)}, A_{\beta(2)}, \dots, A_h, \dots, A_{\beta(m)})$$

è ancora una base di A ovvero $\tilde{B}$ ha colonne tutte linearmente indipendenti. In particolare si dimostra che

$$\det(\tilde{B}) = (-1)^{t+h} \cdot \bar{a}_{th} \neq 0$$

7.4 Base ammissibile per il metodo del simplesso

7.4.1 Metodo del Big-M

L'idea è quella di trasformare il problema in forma standard in un altro problema che utilizza un nuovo vettore come $y \in \mathbb{R}^m$ chiamato vettore delle variabili artificiali. Si ha

$$\begin{cases} \min c^T x + M \cdot \sum_{i=1}^m y_i \\ Ax + Iy = b \\ x \geq 0 \\ y \geq 0 \end{cases}$$

dove M è una costante molto grande, $I \in \mathbb{R}^{m \times m}$ è la matrice identità. Poi

$$Ax + Iy = b$$

lo possiamo scrivere come

$$\begin{bmatrix} A & I \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = b$$

dove

$$A \in \mathbb{R}^{m \times n} \implies \begin{bmatrix} A & I \end{bmatrix} \in \mathbb{R}^{m \times (n+m)}$$

e una base di

$$\begin{bmatrix} A & I \end{bmatrix}$$

è una matrice $B \in \mathbb{R}^{m \times m} : \det(B) \neq 0$. In questo caso, segue che $B = I$ è una base di

$$\begin{bmatrix} A & I \end{bmatrix}$$

dato che $\det(I) \neq 0$ e la soluzione in cui base ci sono le variabili y e fuori base le variabili x è una SBA per il sistema

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

quindi pongo

$$\begin{cases} x_B = y = b \\ x_F = x = 0 \end{cases}$$

Inoltre supponiamo che $b \geq 0$ ed eventualmente se b ha qualche componente negativa allora possiamo moltiplicare per -1 l'equazione. Quindi $B = I$ è una base ammissibile dato che $B^{-1}b = b \geq 0$ per ipotesi. Se il problema

$$\begin{cases} \min c^T x \\ x \in P \equiv \begin{cases} Ax \geq 0 \\ x \geq 0 \end{cases} \end{cases}$$

ha una soluzione ammissibile cioè $P \neq \emptyset$ allora la soluzione ottima del problema artificiale avrà tutte le variabili

$$y_i = 0 \quad i = 0, \dots, m$$

Lo svantaggio di usare questo metodo è che il coefficiente M porta nell'algoritmo del metodo del simplesso ad avere problemi di instabilità numerica.

7.4.2 Metodo delle 2 fasi

- Fase 1: si applica il metodo del simplesso per trovare una base ammissibile
- Fase 2: si applica il metodo del simplesso a partire dalla base trovata nella fase 1 per trovare la soluzione ottima

7.4.2.1 Fase 1

A partire dal problema di PL in forma standard

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

Consideriamo il problema artificiale, con ipotesi $b \geq 0$, si ha

$$\begin{cases} \min y_1 + y_2 + \dots + y_m \\ Ax + Iy = b \\ x \geq 0 \\ y \geq 0 \end{cases}$$

e quindi $y \in \mathbb{R}^m$ ed è il vettore delle variabili artificiali. In questo caso $B = I$ è una base ammissibile per

$$\begin{bmatrix} A & I \end{bmatrix}$$

perché

$$x_B = B^{-1}b = b \geq 0$$

da ipotesi. Poiché $y \geq 0$ allora

$$y_1 + y_2 + \dots + y_m \geq 0$$

quindi il valore minimo della funzione obiettivo cioè l'ottimo del problema artificiale lo otteniamo quando

$$y_1 + y_2 + \dots + y_m = 0$$

per qualche soluzione ammissibile del problema iniziale. Sia $\bar{x}$ una soluzione ammissibile per il problema iniziale cioè

$$\begin{cases} A\bar{x} = b \\ \bar{x} \geq 0 \end{cases} \implies \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \bar{x} \\ 0 \end{bmatrix}$$

è la soluzione ottima del problema artificiale.

7.4.2.2 Fase 2

Determinare una base ammissibile per il problema iniziale risolvendo il problema artificiale. Sia x^*, y^* una soluzione ottima del problema artificiale. Abbiamo 2 casi

- Caso 1

$$\exists y_i^* > 0 \implies \begin{cases} Ax = b \\ x \geq 0 \end{cases}$$

non ha soluzione ammissibile, il poliedro P è vuoto e il problema iniziale non ha soluzione quindi non esiste una soluzione ottima.

- Caso 2

$$y_i^* = 0 \quad i = 1, \dots, m$$

implica che x^* è soluzione ammissibile per il sistema

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

Il problema iniziale ha una soluzione ammissibile x^* e abbiamo due sottocasi

- Nella soluzione ottima x^*, y^* del problema artificiale, tutte le variabili y sono fuori base (ottima). Questo implica che in base ho soltanto variabili x che forniscono una SBA iniziale per la fase 2.

– Si ha

$$\exists y_h^* = 0$$

ed è in base nella soluzione ottima del problema artificiale e ciò implica che

$$x_B = \begin{bmatrix} x_{\beta(1)} & x_{\beta(2)} & \cdots & y_h & \cdots \end{bmatrix}^T$$

e se la SBA ha qualche variabili in base che è nulla, allora SBA si chiama soluzione degenera. In questo caso, si può applicare un'iterazione del metodo del simplesso per far uscire la variabile y_h dalla base e far entrare al suo posto una variabile x_j , la quale entra con valore 0 quindi i valori delle due variabili sono sempre 0. Possiamo far uscire dalla base tutte le variabili y ed ottenere una base con soltanto variabili x ovvero una SBA per il problema iniziale.

7.4.3 Basi e SBA degeneri

Sia B una base ammissibile di A . Consideriamo il problema di PL

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

Allora B si dice degenera se $B^{-1}b$ ha qualche componente nulla. Se ho una base degenera,

$$\begin{cases} x_B = B^{-1}b \\ x_F = 0 \end{cases}$$

si dice SBA degenera. Dal punto di vista geometrico, ho un vertice $\tilde{x}$ che è un punto del poliedro, segue che a questo punto $\tilde{x}$ possono corrispondere più basi e più SBA. Quindi in caso di degenerazione diverse basi ammissibili corrispondono allo stesso vertice, questo fatto può essere un problema per il metodo del simplesso perché esso si potrebbe muovere da una SBA ad un'altra SBA rimanendo sempre sullo stesso vertice quindi il punto geometrico non cambia ma varia le basi e le SBA. Il metodo del simplesso entra in un ciclo infinito, questi si chiamano fenomeni di ciclaggio.

7.4.4 Esercizio

Consideriamo

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

$$A = \begin{bmatrix} A_1 & A_2 & A_3 & A_4 & A_5 \end{bmatrix} = \begin{bmatrix} 4 & 2 & 2 & 1 & 1 \\ 2 & 0 & 1 & 3 & 2 \end{bmatrix} \quad b = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}$$

Una base di A sarà $B \in \mathbb{R}^{2 \times 2}$ e implica che

$$x_B \in \mathbb{R}^2 \quad x_F \in \mathbb{R}^3$$

Consideriamo

$$B = \begin{bmatrix} A_1 \\ A_5 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}$$

sono linearmente indipendenti allora

$$B = \begin{bmatrix} A_1 \\ A_5 \end{bmatrix}$$

è una base di A . Si ha che

$$Bx_B + Fx_F = b$$

dato che

$$x_F = 0 \quad \Rightarrow \quad Bx_B = b \quad \Rightarrow \quad \begin{cases} 4x_1 + 2x_5 = 2 \\ 2x_1 + 2x_5 = 2 \end{cases}$$

$$\begin{cases} x_1 = 0 \\ x_5 = 1 \end{cases} \quad \Rightarrow \quad x_B = \begin{bmatrix} x_1 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \geq 0$$

allora B è base ammissibile. Il vertice associato a questa base B è

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

La SBA è degenera e B è una base ammissibile degenera. A meno di situazioni singolari, le coppie

$$\begin{bmatrix} A_2 & A_5 \end{bmatrix} \quad \begin{bmatrix} A_3 & A_5 \end{bmatrix} \quad \begin{bmatrix} A_4 & A_5 \end{bmatrix}$$

danno luogo allo stesso vertice. Consideriamo

$$B = \begin{bmatrix} A_2 & A_5 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 0 & 2 \end{bmatrix}$$

allora B è una base di A e

$$Bx_B = b \quad \Rightarrow \quad \begin{cases} 2x_2 + 2x_5 = 2 \\ 2x_5 = 2 \end{cases}$$

$$\begin{cases} x_5 = 1 \\ x_2 = 0 \end{cases} \quad \Rightarrow \quad x_B = \begin{bmatrix} x_2 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \geq 0$$

allora B è base ammissibile e il vertice associato a questa base B è

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

come prima ma è cambiata la SBA. Consideriamo

$$B = \begin{bmatrix} A_3 & A_5 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 1 & 2 \end{bmatrix}$$

allora B è una base di A e

$$Bx_B = b \quad \Longrightarrow \quad \begin{cases} 2x_3 + 2x_5 = 2 \\ x_3 + 2x_5 = 2 \end{cases}$$

$$\begin{cases} x_3 = 0 \\ x_5 = 1 \end{cases} \quad \Longrightarrow \quad x_B = \begin{bmatrix} x_3 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \geq 0$$

allora B è base ammissibile e il vertice associato a questa base B è

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

come prima ma è cambiata la SBA. Consideriamo

$$B = \begin{bmatrix} A_4 & A_5 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}$$

allora B è una base di A e

$$Bx_B = b \quad \Longrightarrow \quad \begin{cases} x_4 + 2x_5 = 2 \\ 3x_4 + 2x_5 = 2 \end{cases}$$

$$\begin{cases} x_4 = 0 \\ x_5 = 1 \end{cases} \quad \Longrightarrow \quad x_B = \begin{bmatrix} x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \geq 0$$

allora B è base ammissibile e il vertice associato a questa base B è

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

come prima ma è cambiata la SBA.

7.4.5 Regola di Bland (regola anticiclo)

Durante il metodo del simplesso, ogni volta che si può scegliere, bisogna scegliere la variabile entrante o uscente dalla base con indice minimo. Possiamo scegliere di

1. far entrare in base la variabile fuori base x_h con $c_h < 0$ ed indice h minimo
2. far entrare in base le variabili in base $x_{\beta(t)}$ tale che $\frac{\bar{b}_t}{\bar{a}_{th}} = \vartheta$ ed indice $\beta(t)$ minimo

Utilizzando la regola di Bland, il metodo del simplesso termina (converge) dopo al più

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

iterazioni.

7.5 Metodo del simplesso e teoria della dualità

Consideriamo il problema primale

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

e il problema duale

$$\begin{cases} \max b^T u \\ A^T u \leq c \end{cases}$$

Nel metodo del simplesso, x^* è una soluzione ottima se

1. x^* è SBA cioè

$$x^* = \begin{bmatrix} x_B \\ x_F \end{bmatrix} = \begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix} \geq 0$$

2. $\bar{c}_F^T = c_F^T - c_B^T B^{-1} F \geq 0$

Il vettore dei costi ridotti lo possiamo definire come

$$\bar{c}^T = c^T - c_B^T B^{-1} A$$

e considerando

$$\begin{aligned} c^T &= \begin{bmatrix} c_B^T & c_F^T \end{bmatrix} & A &= \begin{bmatrix} B & F \end{bmatrix} \\ \bar{c}^T &= \begin{bmatrix} c_B^T & c_F^T \end{bmatrix} - c_B^T B^{-1} \begin{bmatrix} B & F \end{bmatrix} = \begin{bmatrix} c_B^T & c_F^T \end{bmatrix} - \begin{bmatrix} c_B^T & c_B^T B^{-1} F \end{bmatrix} = \\ &= \begin{bmatrix} 0^T & c_F^T - c_B^T B^{-1} F \end{bmatrix} = \begin{bmatrix} 0^T & \bar{c}_F^T \end{bmatrix} \end{aligned}$$

La condizione di ottimalità (2) possiamo esprimerla come

$$\bar{c}^T = c_F^T - c_B^T B^{-1} A \geq 0$$

Dalla dualità forte

$$c^T x^* = b^T u^*$$

se e solo se x^* e u^* sono soluzioni ottime, implica che

$$c_B^T B^{-1} b = b^T u^* = u^{*T} b \iff u^{*T} = c_B^T B^{-1}$$

è la soluzione ottima del duale. La condizione di ottimalità (2) è

$$\bar{c}^T = c^T - c_B^T B^{-1} A \geq 0$$

e la possiamo riscrivere come

$$\bar{c}^T = c^T - u^{*T} A \geq 0$$

che implica

$$u^{*T} A \leq c^T \implies (u^{*T})^T A \leq (c^T)^T \implies A^T u^* \leq c$$

Segue che, la condizione di ottimalità per il metodo del simplesso corrisponde alla condizione di ammissibilità del problema duale. Quindi il metodo del simplesso, iterazione dopo iterazione, si sposta da una soluzione ammissibile primale ad un'altra soluzione ammissibile primale mantenendo l'ottimalità della soluzione del duale e si ferma quando è verificata la condizione di ottimalità ovvero quando la soluzione u^* diventa ammissibile per il problema duale.

7.6 Analisi di sensibilità della soluzione ottima

7.6.1 alla variazione del vettore dei termini noti

Consideriamo un problema di PL in forma standard e sia B una base di A ottima. Si ha

$$x^* = \begin{bmatrix} x_B \\ x_F \end{bmatrix} = \begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix}$$

Studiamo la variazione del vettore dei termini noti

$$b \rightarrow b + \Delta b$$

Segue che la condizione di ottimalità rimane invariata dato che non è presente il vettore b , essa è

$$\bar{c}^T = c^T - c_B^T B^{-1} A \geq 0$$

mentre la condizione di ammissibilità è influenzata dal vettore b , si ha

$$B^{-1}b \geq 0 \quad \implies \quad B^{-1}(b + \Delta b) \geq 0 \quad \implies \quad B^{-1}b \geq -B^{-1}\Delta b$$

cioè un sistema di m disequazioni in n variabili. Inoltre

$$x^* = \begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix} \rightarrow \begin{bmatrix} B^{-1}(b + \Delta b) \\ 0 \end{bmatrix}$$

Segue che

$$c^T x^* = c_B^T B^{-1}b \rightarrow c_B^T B^{-1}(b + \Delta b) = c_B^T B^{-1}b + c_B^T B^{-1}\Delta b$$

dove

$$c_B^T B^{-1} = u^{*T}$$

è la soluzione ottima del problema duale. Questo implica che

$$c_B^T B^{-1}\Delta b = u^{*T}\Delta b = \sum_{i=1}^m u_i^* \Delta b_i$$

dove u_i^* è il valore ottimo della variabile duale u_i che "misura" la sensibilità della funzione obiettivo del primale alla variazione del termine noto b_i del vincolo dell' i -esimo del primale.

7.6.2 alla variazione del vettore dei costi

Consideriamo un problema di PL in forma standard e sia B una base di A ottima. Si ha

$$x^* = \begin{bmatrix} x_B \\ x_F \end{bmatrix} = \begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix}$$

Studiamo la variazione del vettore dei costi c

1. variazione c_F

$$c_F + \Delta c_F \quad c_F \in \mathbb{R}^m, \Delta c_F \in \mathbb{R}^{n-m}$$

Studiamo se x^* è ancora ammissibile e ottima? Abbiamo che

$$B^{-1}b \geq 0$$

è la condizione di ammissibilità che rimane invariata, la condizione di ottimalità è

$$\bar{c}_F^T = c_F^T - c_B^T B^{-1}F \geq 0$$

$$\bar{c}_F^T = c_F^T + \Delta c_F^T - c_B^T B^{-1}F \geq 0 \quad \implies \quad \Delta c_F^T \geq -\bar{c}_F^T$$

e otteniamo $n - m$ disequazioni indipendenti cioè

$$\Delta c_j \geq -\bar{c}_j \quad \forall x_j \text{ fuori base}$$

Si ha che il costo ridotto $\bar{c}_j \geq 0$ è il massimo decremento di c_j per cui la base B rimane ottima.

2. variazione c_B

$$c_B + \Delta c_B \quad c_B \in \mathbb{R}^m, \Delta c_B \in \mathbb{R}^{n-m}$$

Studiamo se x^* è ancora ammissibile e ottima? Abbiamo che

$$B^{-1}b \geq 0$$

è la condizione di ammissibilità che rimane invariata, la condizione di ottimalità è

$$\bar{c}_F^T = c_F^T - c_B^T B^{-1}F \geq 0 \quad \implies \quad c_F^T - c_B^T B^{-1}F - \Delta c_B^T B^{-1}F \geq 0$$

$$\bar{c}_F^T - \Delta c_B^T B^{-1}F \geq 0 \quad \implies \quad \Delta c_B^T B^{-1}F \leq \bar{c}_F^T$$