UNIVERSITÀ DEGLI STUDI DI PADOVA

Facoltà di Ingegneria
Corso di Laurea in Ingegneria dell'Informazione

Convex Optimization

Relatore:Chiarissimo Prof. **Sandro Zampieri**

Tesi di Laurea di:

Stefano Zorzi

Matr. n. 593820

ANNO ACCADEMICO 2012/2013

Indice

Indice delle figure

1.Introduzione

1.1 Problema dell'ottimizzazione

Nel linguaggio matematico ottimizzare significa trovare il valore delle variabili di una funzione tale per cui questa assuma il suo minimo o il suo massimo.
Il problema dell'ottimizzazione è un astrazione del problema di eseguire la migliore scelta in un insieme di possibili candidati.
Un problema di ottimizzazione può essere:

$$\begin{array}{lll} \text{minimizzo} & f_0(\mathrm{x}) & \\ \text{soggetto a} & f_i(\mathrm{x}) \leq 0 \quad i = 1, \ldots \ldots, m & \\ & h_i(\mathrm{x}) = 0 \quad i = 1, \ldots \ldots, p & \end{array} \tag{1.1}$$

La funzione f_0(x) viene chiamata la funzione Obiettivo $f_0 : \boldsymbol{R}^n \rightarrow \boldsymbol{R}$ e risulta essere la funzione da minimizzare.
Bisogna però rispettare le condizioni determinate da f_i(x)≤ 0 definiti come Vincoli di Disuguaglianza dove $f_i : \boldsymbol{R}^n \rightarrow \boldsymbol{R}$ e le condizioni dei Vincoli di Uguaglianza dove $h_i : \boldsymbol{R}^n \rightarrow \boldsymbol{R}$.
Se $m = p = 0$ il problema (1.1) non presenta vincoli e si studia come un calcolo di massimi e minimi non vincolato.
La soluzione al problema di ottimizzazione va ricercata in tutti quei punti che appartengono al dominio della funzione Obiettivo e devono obbligatoriamente soddisfare anche le altre condizioni.
Il dominio del problema di ottimizzazione e definito come:

$$D = \bigcap_{i=1}^{m} dom\, f_i \quad \cap \quad \bigcap_{i=1}^{p} dom\, h_i$$

Un punto x si dice ammissibile se soddisfa le condizioni sui vincoli cioè se f_i (x)≤ 0 $i = 1, \ldots \ldots, m$ e h_i(x)=0 $i = 1, \ldots \ldots, p$, l'insieme dei punti ammissibili è detto insieme Ammissibile o Regione Ammissibile del Problema ed è l'insieme dove devo trovare la soluzione.

I vincoli possono essere di 2 tipi e posso avere vincoli attivi e passivi:
un vincolo si definisce attivo se $\forall\, x \in S\ e\ f_i(x)= 0$, mentre se f_i(x)< 0 il vincolo si definisce passivo.

Il valore ottimale del problema è definito come :

$$p^* = \inf \{ \ f_0(x) \mid f_i(x) \leq 0 \quad i = 1, \ldots\ldots, m\ , \ h_i(x) = 0 \quad i = 1, \ldots\ldots, p \ \}.$$

Una soluzione al Problema 1.1 è un punto x^* che oltre ad appartenere all'insieme dei punti Ammissibili, minimizza la funzione f_0(x) cioè $f_0(x^*) = p^*$.
Solitamente non è detto che esista una soluzione per il problema e, se esiste, non è detto che questa sia unica.
Nel caso in cui siano presenti più soluzioni si definisce l'insieme di soluzioni ottime :

$$X_{opt} = \{x \mid f_i(x) \leq 0 \quad i = 1, \ldots\ldots, m\ , \ h_i(x) = 0 \quad i = 1, \ldots\ldots, p\ , f_0(x) = p^*\}\ .$$

Se esiste almeno una soluzione il problema è risolvibile, altrimenti se l'insieme X_{opt} è vuoto il problema è impossibile.
L'ottimizzazione viene affrontata valutando un punto ottimo localmente oppure globalmente.
Ottimizzare localmente significa che x minimizza f_0 in un intorno di un punto ammissibile, mentre ottimizzare globalmente significa che x minimizza f_0 su tutto il dominio.

1.2 Ottimizzazione Convessa

L'ottimizzazione convessa è un problema di ottimizzazione che si presenta nella forma:

$$\begin{array}{lll} \text{minimizzo} & f_0(\text{x}) & \\ \text{soggetto a} & f_i(\text{x}) \leq 0 & i = 1, \ldots \ldots, m \\ & h_i(\text{x}) = 0 & i = 1, \ldots \ldots, p \end{array} \qquad (1.1)$$

Inoltre risultano essere presenti 3 proprietà:

1. La funzione obiettivo è una funzione convessa.
2. I vincoli di disuguaglianza sono funzioni convesse.
3. I vincoli di uguaglianza devono essere affini.

Una funzione $f : R^n \rightarrow R$ si dice convessa se $dom\, f$ è un insieme convesso e se per ogni $x, y \in dom\, f$ e θ con $0 \leq \theta \leq 1$ abbiamo

$$f(\theta\, \text{x} + (1 - \theta)\text{y}) \leq \theta\, \text{f(x)} + (1 - \theta)\text{f(y)}$$

Una funzione $f : R^n \rightarrow R$ si dice affine se è la somma di una funzione lineare e di una costante, cioè si presenta nella forma $f(x) = Ax + b$ dove $A \in R^{m \times n}$ e $b \in R^m$.

L'insieme dei punti ammissibili del problema di ottimizzazione convessa è a sua volta un insieme convesso.

Infatti il dominio del problema è definito come $D = \bigcap_{i=1}^{n} dom\, f_i$ è un insieme convesso costituito da m sottoinsiemi convessi intersecati con p iperpiani rappresentati dai vincoli di uguaglianza.

Concludendo il problema dell'ottimizzazione convessa non è altro che la minimizzazione di una funzione convessa in un insieme convesso.

Una proprietà fondamentale dell'ottimizzazione convessa è che un punto localmente ottimo è anche ottimo globalmente; questo significa che se io ho un problema di ottimizzazione convessa e restringo la mia ricerca ad un intorno di un punto x, che è candidato punto ottimo e nell'intorno non c'e alcun punto migliore di x, si può arrivare alla conclusione che il punto è un punto ottimo globalmente.

Un esempio interessante è :

$$
\begin{aligned}
&\text{Minimizzo} && f_0(x) = x_1^2 + x_2^2 \\
&\text{Soggetto a} && f_1(x) = \frac{x_1}{1+x_2^2} \leq 0 \\
& && h_1(x) = (x_1 + x_2)^2 = 0
\end{aligned}
$$

Apparentemente il problema risulta non essere un problema di ottimizzazione convessa in quanto $h_1(x)$ non è affine e $f_1(x)$ non è convessa, ma l'insieme dei valori ammissibili costituito da $\{x|\, x_1 \leq 0, x_1 + x_2 = 0\}$ è convesso, quindi stiamo minimizzando una funzione convessa su un insieme convesso.

Il problema può essere riscritto come :

$$
\begin{aligned}
&\text{minimizzo } f_0(x) = x_1^2 + x_2^2 \\
&\text{soggetto a } \tilde{f}\,(x) = x_1 \leq 0 \\
&\qquad\qquad\quad \tilde{h}(x) = x_1 + x_2 = 0
\end{aligned}
$$

I due problemi non sono identici perché due problemi di ottimizzazione si dicono identici solo se la funzione obiettivo e le funzioni del vincolo sono uguali, in questo caso i due problemi sono equivalenti in quanto portano allo stesso risultato.

2. La dualità

2.1 La funzione Lagrangiana

Considerando un problema di Ottimizzazione:

$$\begin{array}{lll} \text{minimizzo} & f_0(x) & \\ \text{soggetto a} & f_i(x) \leq 0 \quad i = 1, \dots\dots, m & \\ & h_i(x) = 0 \quad i = 1, \dots\dots, p & \end{array} \qquad (2.1)$$

Dove x$\in R^n$, il dominio è definito come $D = \{\bigcap_{i=1}^{m} dom\, f_i \quad \cap \quad \bigcap_{i=1}^{p} dom\, h_i\}$ $\neq \emptyset$ e p^* è il valore ottimale.

L'idea base nella dualità Lagrangiana consiste nell'introdurre una nuova funzione costituita dalla somma della funzione obiettivo con delle somme pesate delle funzioni dei vincoli.

Si definisce funzione Lagrangiana $L(x, \lambda, \mu) : R^n \times R^m \times R^p \rightarrow R$ come:

$$L(x, \lambda, \mu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \mu_i h_i(x) \qquad (2.2)$$

Dove $dom\, L = D \times R^m \times R^p$ mentre λ_i assume il nome di moltiplicatore di Lagrange associato allo i-esimo vincolo di disuguaglianza e μ_i è il moltiplicatore di Lagrange associato allo i-esimo vincolo di uguaglianza.

Considerando che $f_i(x) \leq 0$ la funzione Lagrangiana assumerà valori inferiori rispetto alla funzione obiettivo oppure valori uguali quando $f_i(x) = 0$ così trovando i punti di comuni a $f_0(x)$ e $L(x, \lambda, \mu)$.

Strettamente collegata alla funzione Lagrangiana è la funzione duale di Lagrange definita come il minimo valore di $L(x, \lambda, \mu)$ su x per $\lambda \in R^m$, $\mu \in R^p$

$$g(\lambda, \mu) = inf_{x \in D}\, L(x, \lambda, \mu) = inf_{x \in D}(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \mu_i h_i(x))$$

La funzione duale fornisce un limite inferiore del valore ottimale p^* del problema 2.1. Infatti una proprietà molto importante è che $g(\lambda, \mu)$ è sempre minore o uguale al valore ottimale.

$$g(\lambda, \mu) \leq p^* \qquad (2.3)$$

Supponendo che $\check{x}$ sia un punto ammissibile cioè $f_i(\check{x}) \leq 0$, $h_i(\check{x})=0$ e $\lambda \geq 0$ allora

$$\sum_{i=1}^{m} \lambda_i f_i(\check{x}) + \sum_{i=1}^{p} \mu_i h_i(\check{x}) \leq 0$$

La prima sommatoria è negativa, mentre la seconda sommatoria risulta essere nulla. Quindi:

$$L(\check{x}, \lambda, \mu) = f_0(\check{x}) + \sum_{i=1}^{m} \lambda_i f_i(\check{x}) + \sum_{i=1}^{p} \mu_i h_i(\check{x}) \leq f_0(\check{x}).$$

E:

$$g(\lambda, \mu) = inf_{x \in D}\, L(x, \lambda, \mu) \leq L(\check{x}, \lambda, \mu) \leq f_0(\check{x}).$$

Per cui $g(\lambda, \mu) \leq f_0(\check{x})$ è valida per ogni punto ammissibile e anche per il punto ottimale, da questo deriva la (2.3).

E' importante notare di come il duale Lagrangiano è una funzione concava perché è un estremo inferiore puntuale di una funzione affine.

2.2 Interpretazione Grafica

La funzione Lagrangiana e la funzione duale di Lagrange possono essere interpretate graficamente:

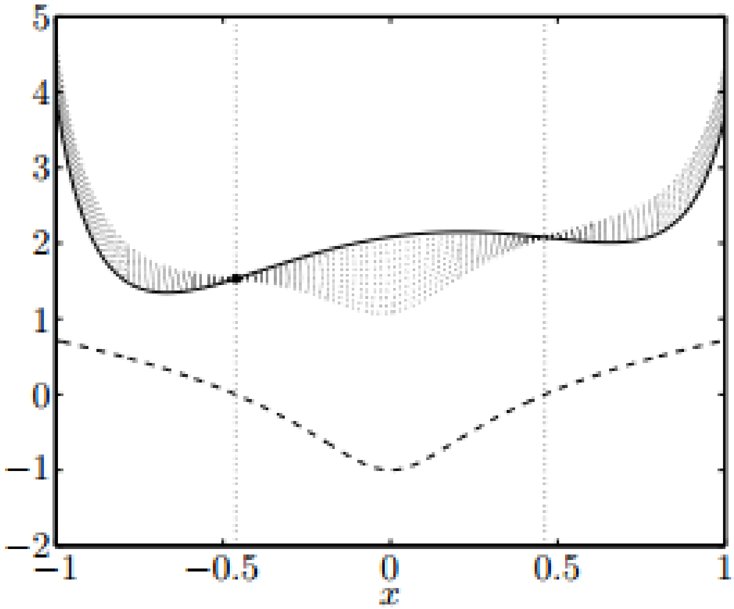

Figura 1

Nel grafico in figura 1 si può vedere la funzione obiettivo f_0(x), rappresentata in linea continua, il vincolo $f_1(x)$, rappresentata da una linea tratteggiata e l'insieme dei punti ammissibili compresi tra [-0.46,0.46].

Le curve tratteggiate rappresentano $L(x,\lambda)$ al variare di λ, si può notare come il punto ottimale è x^*= 0.46 ed il valore ottimale $p^* = 1.54$.

Nel grafico per ogni valore ammissibile vale la disequazione $L(x,\lambda) \leq f_0(x)$ al variare di λ.

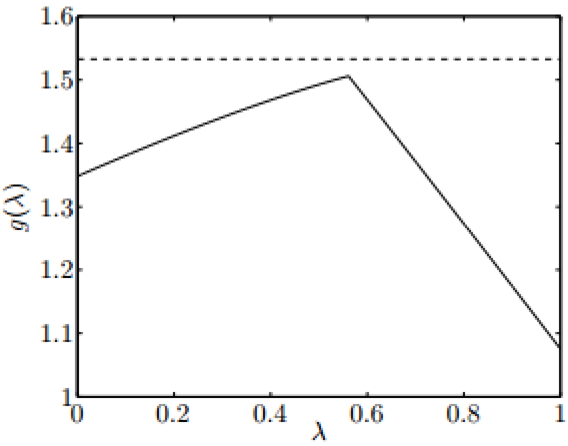

Figura 2

In questo secondo grafico è rappresentato l'andamento della funzione $g(\lambda)$ al variare di λ nell intervallo [0,1].

In figura 2 la linea tratteggiata rappresenta il valore ottimale p^*.

L'estremo superiore della funzione $g(\lambda)$ differisce dal punto ottimale a meno di una differenza che definiamo gap ottimale, inoltre nel grafico è sempre rispettata la proprietà (2.3).

Un esempio del calcolo della funzione duale di Lagrange può essere :

minimizzo	f_0(x)=$c^T x$
soggetto a	$Ax = b$
	$x \geq 0$

Il vincolo di disuguaglianza ha funzioni f_i(x)=x_i $i = 1, \ldots, n$ mentre il vincolo di uguaglianza è $Ax - b = 0$, la funzione Lagrangiana $L(x, \lambda, \mu)$ si ottiene come:

$$L(x, \lambda, \mu) = c^T x - \sum_{i=1}^{n} \lambda_i x_i + \mu^T (Ax - b) = -b^T \mu + (c + A^T \mu - \lambda)^T x$$

Mentre la funzione duale è :

$$g(\lambda,\mu) = inf_x\, L(x,\lambda,\mu) = -b^T\mu + inf_x(c + A^T\mu - \lambda)^T x$$

Si osserva che $g(\lambda,\mu)$ è una funzione affina costituita da una somma di x più una costante, quindi il limite inferiore è illimitato a meno che $c + A^T\mu - \lambda = 0$:

$$g(\lambda,\mu) = \begin{cases} -b^T\mu & se\ c + A^T\mu - \lambda = 0 \\ -\infty & altrimenti \end{cases}$$

Quindi $-b^T\mu$ è il limite inferiore del valore ottimo.

2.3 Problema duale di Lagrange

La funzione duale di Lagrange ci fornisce un limite inferiore del punto ottimale p^*, questo limite dipende dai valori assunti da λ e μ, quindi dobbiamo valutare qual' è il migliore valore della funzione duale di Lagrange.

Definisco il Problema Duale di Lagrange come:

$$\text{massimizzo } g(\lambda,\mu) \qquad (2.4)$$
$$\text{soggetta a } \lambda \geq 0$$

La soluzione a (2.4) sarà una coppia di valori (λ^*,μ^*) che rappresenta la coppia di valori che massimizza la funzione duale di Lagrange e verrà chiamata Moltiplicatori Ottimi di Lagrange.

Il problema duale di Lagrange è un problema di Ottimizzazione Convessa fino a che la funzione obiettivo da massimizzare è concava e i vincoli sono convessi.

Si possono fare delle osservazione inerenti al dominio di $g(\lambda,\mu)$ in quanto la funzione duale di Lagrange fornisce un limite inferiore quando $\lambda \geq 0$ e $(\lambda,\mu) \in dom\ g$ dove il $dom\ g = \{(\lambda,\mu)\ |\ g(\lambda,\mu) \geq -\infty\ \}$.

Si definisce Coppia Ammissibile le Coppie (λ,μ) tale per cui $\lambda \geq 0$ e $(\lambda,\mu) \in dom\ g$.

Riprendendo l'esempio del paragrafo precedente la funzione duale di Lagrange è definita come:

$$g(\lambda,\mu) = \begin{cases} -b^T\mu & se\ c + A^T\mu - \lambda = 0 \\ -\infty & altrimenti \end{cases}$$

Dalla funzione duale di Lagrange si può ricavare il Problema duale di Lagrange utilizzando $g(\lambda,\mu)$, in quanto si va a valutare quale tra i limiti inferiore è il migliore.

Quindi si deve massimizzare $g(\lambda,\mu)$ quando non è infinita:

massimizzo

$$g(\lambda,\mu) = \begin{cases} -b^T\mu & se\ c + A^T\mu - \lambda = 0 \\ -\infty & altrimenti \end{cases}$$

soggetto a $\lambda \geq 0$

Quindi massimizzo

$$g(\lambda,\mu) = \begin{cases} -b^T\mu & se\ \lambda = c + A^T\mu \\ -\infty & se\ \lambda \neq c + A^T\mu \end{cases}$$

soggetto a $\lambda \geq 0$

dovendo essere $\lambda \geq 0$ allora identifico due regioni:

$$M_1 = \{\mu | c + A^T\mu \geq 0\}$$
$$M_2 = \{\mu | c + A^T\mu < 0\}$$

Si va quindi a massimizzare $g(\lambda,\mu)$ nelle due regioni M_1 e M_2 prendendo in considerazione il valore più grande tra i due al variare di μ sulle due regioni;

$$max\{sup_{\mu\in M_1}\, g(\lambda,\mu), sup_{\mu\in M_2}\, g(\lambda,\mu)\} =$$
$$= max\{sup_{\mu\in M_1} - b^T\mu\, , sup_{\mu\in M_2} - \infty\} =$$
$$= sup_{\mu\in M_1} - b^T\mu$$

Il passaggio dal problema primale al problema duale può portare dei vantaggi dal punto di vista dimensionale in quanto si va ad affrontare un problema di dimensioni minori se la matrice A ha dimensioni minori di c.

2.4 La dualità

Nei paragrafi precedenti abbiamo introdotto la funzione duale di Lagrange per risolvere un problema di ottimizzazione.
Si è verificato di come il risultato è lo stesso, sia se si minimizza la funzione obiettivo soggetta ad alcuni vincoli, sia se si massimizza la funzione duale di Lagrange.
Si definisce d^* il valore ottenuto risolvendo il Problema duale di Lagrange, mentre p^* è il limite inferiore della funzione duale di Lagrange.
I punti d^* e p^* sono legati da una dualità che può essere di due tipi:

- Dualità debole
- Dualità forte

La dualità debole si ha quando $d^* \leq p^*$, questa dualità esiste anche se il problema di ottimizzazione non è convesso.
Se la funzione duale di Lagrange non è limitata, quindi il limite inferiore $p^* = -\infty$, si ha che $d^* = -\infty$ e il problema di Lagrange è inammissibile.
Si definisce gap duale ottimale la differenza tra p^* e d^* e fornisce quindi la differenza tra il valore ottimale ottenuto dal problema principale e il miglior limite inferiore della funzione duale di Lagrange.
La dualità forte si ha se $d^* = p^*$ cioè i due valori coincidono e il gap ottimale risulta essere nullo.
La dualità forte solitamente risulta essere presente se il problema di ottimizzazione è convesso.
Le condizioni sotto le quali si verifica la dualità forte vengono chiamate qualificazioni del vincolo.
Una qualificazione del vincolo è la condizione di Slater: $\exists\, x \in int\, D : \; f_i(x) < 0 \quad i = 1, \dots, m$ e x soddisfa il vincolo di uguaglianza, allora è presente una dualità forte.

Inoltre se tra le $f_i(x)$ funzioni del vincolo si ha k funzioni affine, la condizione di strettamente minore di zero per un punto $x \in int\ D$ deve valere per le restanti funzioni del vincolo e non per le funzioni del vincolo affine:

$$f_i(x) \leq 0 \qquad i = 1, \dots, k \qquad\qquad f_i(x) < 0 \qquad i = k+1, \dots, m \, .$$

Considero l'esercizio 5.1 tratto da "Convex Optimization" dove il problema di ottimizzazione è :

minimizzo $f_0(\mathrm{x}) = x^2 + 1$

soggetta a $f_1(\mathrm{x}) = (x-2)(x-4) \leq 0$

con $x \in R$.

L'insieme dei punti ammissibili è [2,4], il punto ottimale $x^* = 2$ e il valore ottimale $p^* = 5$.

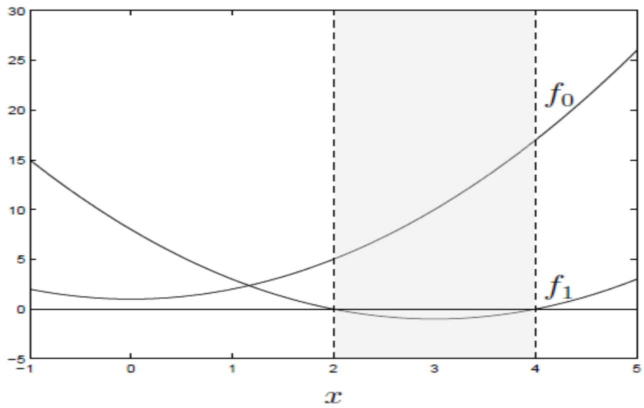

Figura 3

La funzione Lagrangiana è:

$$L(x, \lambda) = (1+\lambda)x^2 - 6\lambda x + (1+8\lambda)$$

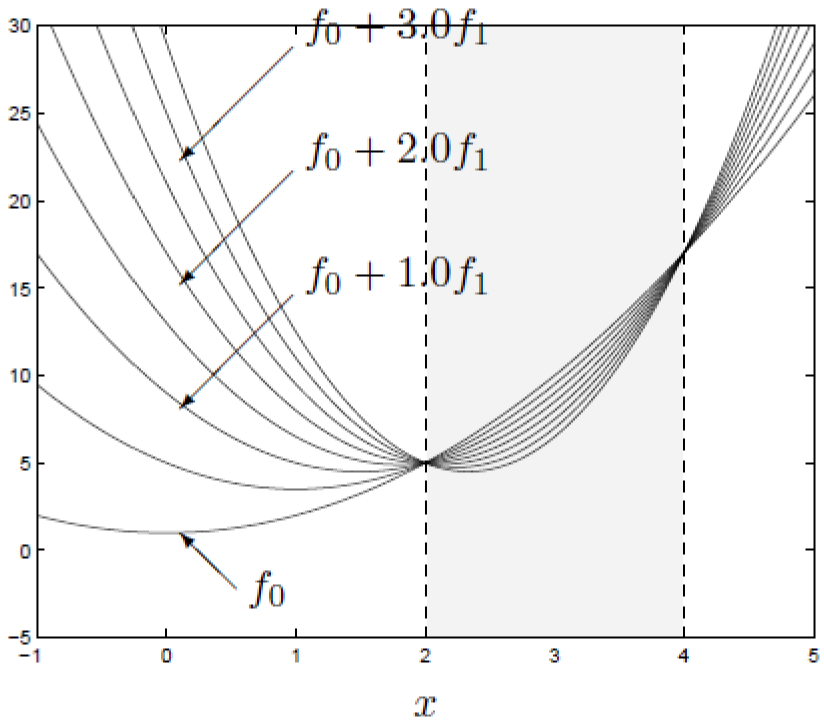

Figura 4

In figura 4 è rappresentata graficamente la funzione Lagrangiana $L(x, \lambda) = f_0 + \lambda\ f_1$ quando λ assume i valori [1,2,3], si nota che il minimo valore di $L(x, \lambda)$ su xè sempre meno di p^*,aumenta quando λ cresce da 0 a 2 mentre decresce quando λ diventa maggiore di 2.

In λ=2 abbiamo che $g(\lambda) = p^*$.

Per $\lambda > -1$ la funzione Lagrangiana ha il minimo in $\tilde{x} = 3\lambda/(1+\lambda)$, per $\lambda \leq -1$ la funzione è illimitata, quindi la funzione $g(\lambda)$ assume i valori:

$$g(\lambda) = \begin{cases} -\dfrac{9\lambda^2}{1+\lambda} + 1 + 8\lambda & \lambda > -1 \\ \infty & altrimenti \end{cases}$$

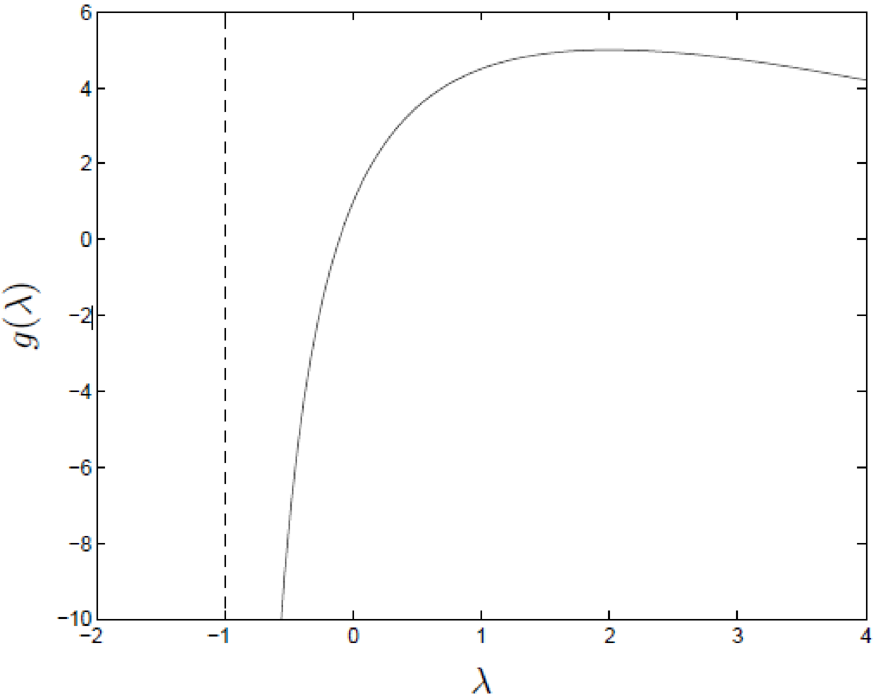

Figura 5

In figura 5 si vede l'andamento della funzione $g(\lambda)$.

La funzione duale di Lagrange è una funzione concava, si vede che p^*= 5 per λ = 2, poi la funzione decresce.

Il problema duale di Lagrange è:

massimizzo $\quad g(\lambda) = -9\lambda^2/(1+\lambda) + 1 + 8\lambda$

soggetta a $\quad \lambda \geq 0$

La soluzione al problema duale di Lagrange si ha per λ=2 e $d^* = 5$, quindi $d^* = 5 = p^*$.

Risultano soddisfatte le condizioni di Slater, perciò è presente una dualità forte.

2.5 Interpretazione Geometrica

I problemi affrontati fino ad ora possono essere reinterpretati da un punto di vista geometrico utilizzando un' analisi basata sugli insiemi.
Si definisce G l'insieme dei valori assunti dai vincoli e dalla funzione obiettivo:

$$G = \left\{ \left(f_1(x), \ldots, f_m(x), h_1(x), \ldots, h_p(x), f_0(x) \right) \in R^n \times R^m \times R \,\middle|\, x \in D \right\}$$

E p^* viene espresso nei termini di G come:

$$p^* = \inf\{t|(u,v,t) \in G\ u \leq 0, v = 0\}$$

La funzione duale di Lagrange è definita come minimizzazione della funzione:

$$(\lambda, \mu, 1)^T(u,v,t) = \sum_{i=1}^{m} \lambda_i u_i + \sum_{i=1}^{p} \mu_i v_i + t$$
$$g(\lambda, \mu) = inf\{ (\lambda, \mu, 1)^T(u,v,t) \mid (u,v,t) \in G\}$$

Da tale minimizzazione si trova che la disequazione $(\lambda, \mu, 1)^T(u,v,t) \geq g(\lambda, \mu)$ definisce un iperpiano di supporto all' insieme G.
Per semplificare la trattazione grafica si ipotizza di avere solo un vincolo $f_1(x) \leq 0$.

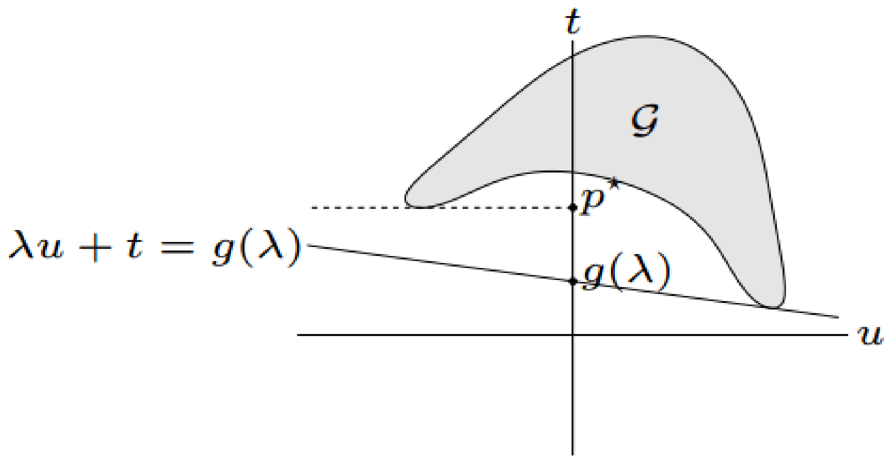

Figura 6

In figura 6 si vede l'interpretazione geometrica del limite inferiore $g(\lambda)$, che è minore del valore ottimale della funzione obiettivo ed identifica un iperpiano di supporto.
L'asse delle ascisse rappresenta i valori della funzione obiettivo, mentre l'asse delle ordinate rappresenta il valore del vincolo.
Tutti i punti appartenenti a G con $u \leq 0$ sono i punti ammissibili, in quanto u rappresenta i valori del vincolo $f_1(x)$ e per soddisfare la disugualianza del vincolo $f_1(x) \leq 0$.
Risolvere il problema duale di Lagrange da un punto di vista grafico significa minimizzare la funzione $f_0(x) + \lambda f_1(x) = g(\lambda) = \lambda\mu + t.$
Si deve fissare una pendenza λ e l'intersezione tra l'asse delle ascisse e l'iperpiano trovato fornisce il valore di $g(\lambda)$.
Quindi più si aumenta la pendenza più $g(\lambda)$ aumenta e il gap tra p^*e d^* diminuisce.

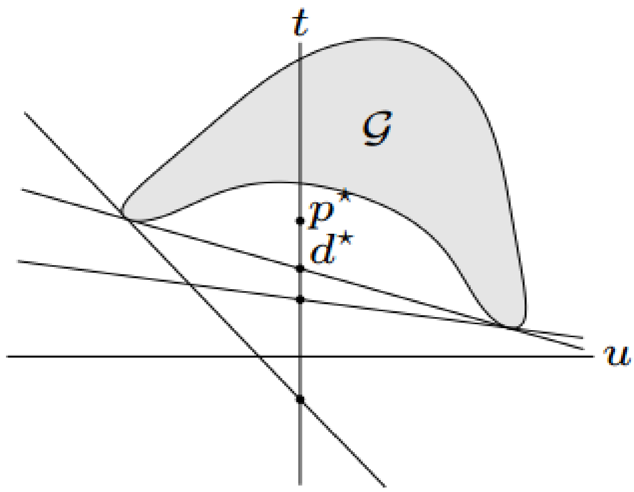

Figura 7

Dalla figura 4 al variare di λ e μ si identificano diversi iperpiani ma, solo una coppia (λ, μ) permette di identificare il valore d^* che rappresenta la soluzione al problema duale di Lagrange.
Dall'interpretazione grafica si riesce ad analizzare meglio il motivo per cui si ottiene una dualità forte nei casi di ottimizzazione convessa.

Infatti se $f_0(x)$ e $f_1(x)$ sono convessi allora anche l'insieme G è convesso quindi il gap di dualità è zero.

Se modifichiamo l'insieme G come in figura 5

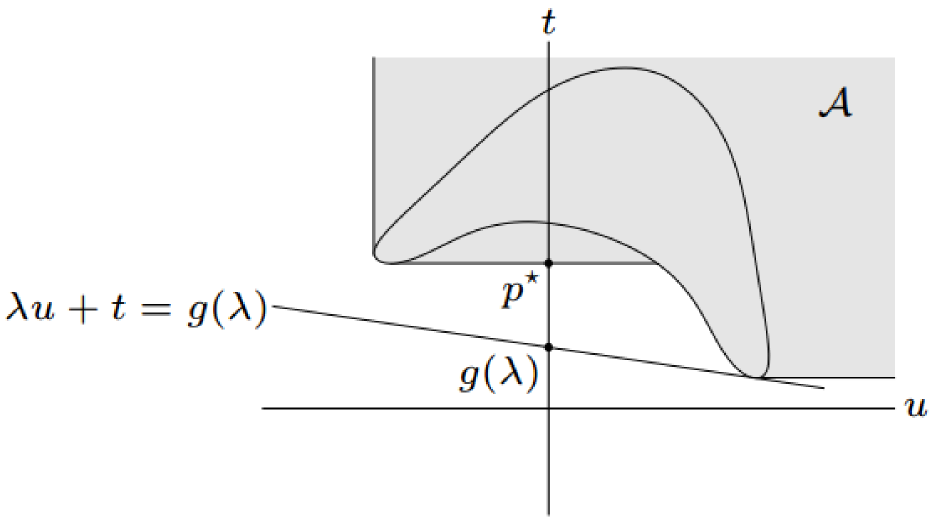

Figura 8

cioè si ridefinisce l'insieme $A = \{(u,t) | \exists\, x \in D,\ f_0(x) \leq t\,, f_1(x) \leq u\}$ per valutare la funzione duale al punto (λ, μ) con $\lambda \geq 0$ devo minimizzare la funzione $(\lambda, \mu, 1)^T(u, v, t)$ su A.

L'estremo inferiore ovviamente definisce un iperpiano di supporto ad A e in particolare se $(0,0,p^*) \in \text{bd}\, A$ si dice che

$$p^* = (\lambda, \mu, 1)^T(u, v, t) \geq g(\lambda, \mu)$$

verificandosi una dualità debole.

La dualità forte si ottiene quando al posto della disuguaglianza si ha una uguaglianza.

La dualità forte è presente se al punto $(0, p^*)$ c'è un iperpiano non verticale di supporto ad A.

Nei problemi di ottimizzazione convessa A è sicuramente convesso e quindi c'è un iperpiano di supporto non verticale in $(0, p^*)$.

Le condizioni di Slater per un problema convesso dicono che è presente una dualità forte se $\exists$ un punto $(\tilde{u}, \tilde{t}) \in A$ con $\tilde{u} < 0$.

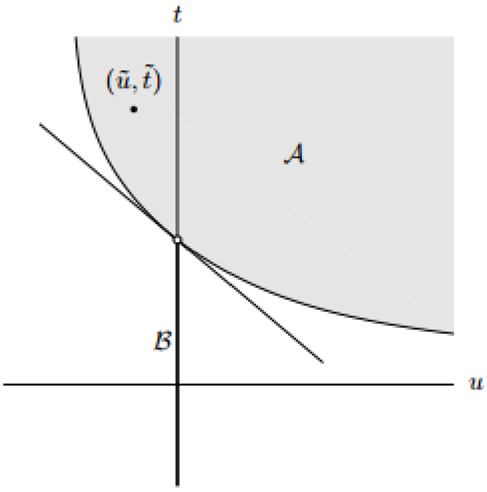

Figura 9

Si può provare come le condizioni di Slater garantiscono una dualità forte per un problema convesso ($f_0(x)$ convessa e f_i(x)≤ 0 $\ i = 1, \dots \dots, m$ con f_i(x) convesse).

Si consideri un problema primario convesso e si assuma che siano verificate le condizioni di Slater : $\exists\ \tilde{x} \in\ int\ D \mid\ f_i(\tilde{x})$<0 e $A\tilde{x} = b$.

Sia $A = \{(u, t) | \exists\ x\ \in D,\ f_0(x) \leq t\ , f_1(x) \leq u\}$ insieme convesso e definisco $B = \{(0,0, s) \in\ R^m\ \times\ R^p\ \times R\ |s < p^*\}$ cioè B è la retta che va da -∞ a p^* escluso.

Sfruttando il teorema dell'iperpiano che dice :

dati C e D due insiemi convessi con intersezione vuota allora $\exists\ a\ \neq 0$ e b tale che $a^T x\ \leq b$ per ogni $x \in C$ e $a^T x\ \geq b$ per ogni $x \in D$.

In altre parole $\exists$ un iperpiano $\{x\ |a^T x = b\}$ chiamato iperpiano di separazione degli insiemi C e D.

Tornando alla dimostrazione precedente si può affermare che $\exists$ un iperpiano di separazione tra A e B che definisce un iperpiano di supporto ad A in (0, p^*).

Le condizioni di Slater, guardando la figura 6, si possono interpretare come se $\exists$ un punto interno $(\tilde{u}, \tilde{t})\ con\ \tilde{u} < 0$ dove $\tilde{u}$ rappresenta il valore di $f_1(x)$.

La dualità forte è presente se esiste un iperpiano di supporto non verticale ad A nel punto (0,0, p^*) ma essendo $\tilde{u} < 0$ e ricordando di come l'insieme A è un insieme convesso, allora l'iperpiano di supporto non può essere verticale.

Considerando l'esempio del paragrafo 2.2 il problema primale è definito come:

$$\begin{array}{ll} \text{minimizzo} & f_0(\text{x})=c^T x \\ \text{soggetto a} & Ax = b \\ & x \geq 0 \end{array}$$

Da questo si ricava la funzione duale espressa come:

$$g(\lambda, \mu) = inf_x\, L(x, \lambda, \mu) = -b^T\mu + inf_x(c + A^T\mu - \lambda)^T x$$

E si definisce il problema duale di Lagrange:

$$\begin{array}{ll} \text{massimizzo} & -b^T\mu \\ \text{soggetta a} & c + A^T\mu - \lambda = 0 \qquad \lambda \geq 0 \end{array}$$

Il problema duale appena scritto non è il vero problema duale ma è un problema equivalente e quindi accettabile come problema duale.

Le condizioni di Slater dicono geometricamente che se il poliedro $Ax = b$ non ha l'insieme interno vuoto, esiste allora una dualità forte.

Nell' esempio è sempre presente la dualità forte a meno che il problema primario e duale non presentano soluzione, cioè l'insieme dei punti ammissibili è vuoto.

Considero l'esercizio 5.21 tratto da "Convex Optimization" dove il problema di ottimizzazione è:

$$\begin{array}{ll} \text{minimizzo} & f_0(x) = e^{-x} \\ \text{soggetta a} & f_1(x, y) = x^2/y \leq 0 \end{array}$$

funzioni definite nel dominio $D = \{(x, y) \mid y > 0 \quad \}$.

Si può verificare di come questo è un problema di ottimizzazione convessa, infatti il dominio è un insieme convesso, la funzione obiettivo $f_0(x)$ è una funzione convessa fino a quando $\ddot{f_0}(x) \geq 0$.

La funzione $f_1(x, y)$ è una funzione convessa in quanto lo si può verificare dalla matrice hessiana.

Quindi il problema considerato è un problema di ottimizzazione convessa.

Il punto di ottimizzazione è $p^* = 1$ quando $x^*= 0$.

Il problema duale di Lagrange può quindi essere formulato come:

massimizzo	$g(\lambda) = inf_x \, L(x, \lambda) = inf_x(e^{-x} + \lambda \, (x^2/y))$
soggetta a	$\lambda \geq 0$

Per ogni scelta di λ, lasciando x $\rightarrow \infty$ e $x^2/y \rightarrow 0$ abbiamo che $g(\lambda) = 0$, quindi $d^* = 0$ e il gap di dualità è pari a 1.

Fino a quando $x^2/y \leq 0$ non è una funzione affine e non c'è nessun punto interno che soddisfa $x^2/y < 0$, le condizioni di Slater non sono verificate.

3.Proprietà della dualità e condizioni ottime

3.1 Caratterizzazione della dualità

È possibile rappresentare in forma simmetrica la relazione che esiste tra il problema primale e il problema duale di Lagrange.

Ipotizzando di non avere vincoli di uguaglianza si può definire :

$$\sup_{\lambda\geq 0} L(x,\lambda) = \sup_{\lambda\geq 0}\left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(\text{x})\right)$$

$$= \begin{cases} f_0(x) & f_i(x) \leq 0 \\ \infty & altrimenti \end{cases}$$

Infatti sommando a $f_0(x)$ valori negativi ovviamente si ottiene $f_0(x)$ come limite superiore, mentre se $f_i(x) \geq 0$ allora il limite superiore è infinito.

Posso esprimere il valore ottimale del problema primale come:

$$p^* = inf_x sup_{\lambda\geq 0}\, L(x,\lambda)$$

Mentre

$$d^* = sup_{\lambda\geq 0}\, inf\ L(x,\lambda)$$

Da queste definizioni si può esprimere la dualità debole come la disequazione:

$$sup_{\lambda\geq 0}\, inf\ L(x,\lambda) \leq inf_x sup_{\lambda\geq 0}\, L(x,\lambda)$$

Mentre la dualità forte ha una relazione di uguaglianza:

$$sup_{\lambda\geq 0}\, inf\ L(x,\lambda) = inf_x sup_{\lambda\geq 0}\, L(x,\lambda)$$

Il risultato ottenuto è una importante proprietà della dualità forte in quanto questa uguaglianza ci dice che posso invertire l'ordine con il quale si affronta il problema primale di ottimizzazione e il problema duale di Lagrange mantenendo inalterato il risultato finale.

Quindi, si può prima minimizzare la funzione obiettivo su x e poi massimizzare la funzione duale su $\lambda \geq 0$ o viceversa senza modificare l'esito finale.

3.2 Teorema dello scarto complementare

Sia x^* soluzione al problema primale e (λ^*, μ^*) soluzione ottimale al problema duale di Lagrange con presenza di una dualità forte allora:

$$\begin{aligned} f_0(x^*) &= g(\lambda^*, \mu^*) \\ &= inf_{x \in D}\left(f_0(x) + \sum_{i=1}^{m} \lambda^*{}_i f_i(\mathrm{x}) + \sum_{i=1}^{p} \mu^*{}_i h_i(\mathrm{x})\right) \\ &\leq f_0(x^*) + \sum_{i=1}^{m} \lambda^*{}_i f_i(x^*) + \sum_{i=1}^{p} \mu^*{}_i h_i(x^*) \\ &\leq f_0(x^*) \end{aligned}$$

Il risultato ottenuto è di particolare interesse perché applicando semplicemente le definizioni di dualità forte e valutando $g(\lambda^*, \mu^*)$ in x^* si arriva ad una identità.

Quindi tutte le disuguaglianze lungo la catena risultano essere uguaglianze sotto le condizioni di dualità forte arrivando così a due conclusioni:

- x^* minimizza la funzione $L(x, \lambda^*, \mu^*)$
- $\lambda^*{}_i f_i(\mathrm{x}) = 0 \;\; i = 1, \ldots\ldots, m$

Se il moltiplicatore di Lagrange è positivo, $f_i(x^*)$ deve essere nulla.

Se $f_i(x^*) < 0$ allora il moltiplicatore di Lagrange associato al vincolo deve essere nullo.

Queste sono conosciute come le condizioni dello scarto complementare e dicono che lo i-esimo moltiplicatore ottimo di Lagrange è zero, a meno che lo i-esimo vincolo sia attivo.

3.3 Condizioni di Karush-Kuhn-Tucker

Le condizioni di ottimalità KKT variano a seconda che si affrontino problemi convessi o non convessi.

3.3.1 Condizioni KKT per problemi non convessi

Sia x^* soluzione per il problema primale, (λ^*, μ^*) soluzione del problema duale con gap di dualità pari a zero, la funzione obiettivo $f_0(x)$ e le funzioni dei vincoli $f_i(x)$ differenziabili, allora x^* minimizza $L(x, \lambda^*, \mu^*)$ solo se è nullo il gradiente di L cioè se

$$\nabla_x L = 0$$

$$\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda^*{}_i \nabla f_i(x^*) + \sum_{i=1}^{p} \mu^*{}_i \nabla h_i(x^*) = 0$$

Quindi:

$$f_i(x^*) \leq 0 \qquad i = 1, \dots \dots, m$$

$$h_i(x^*) = 0 \qquad i = 1, \dots \dots, p$$

$$\lambda^*{}_i \geq 0 \qquad i = 1, \dots \dots, m$$

$$\lambda^*{}_i f_i(x^*) = 0 \qquad i = 1, \dots \dots, m$$

$$\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda^*{}_i \nabla f_i(x^*) + \sum_{i=1}^{p} \mu^*{}_i \nabla h_i(x^*) = 0$$

Queste sono chiamate le condizioni di Karush-Kuhn-Tucker (KKT).
Le condizioni devono essere soddisfatte per ogni coppa di soluzioni del problema primale e duale dove è rispettata la condizione di dualità forte.

3.3.2 Condizioni KKT per problemi convessi

Se il problema primale è convesso le condizioni KKT sono sufficienti per essere punti ottimali.
Se $f_i(x)$ sono convesse mentre le funzioni $h_i(x)$ sono affine e $\tilde{x}, \tilde{\lambda}, \tilde{\mu}$ sono punti che soddisfano le condizioni KKT :

$$f_i(\tilde{x}) \leq 0 \qquad i = 1, \dots \dots, m$$

$$h_i(\tilde{x}) = 0 \qquad i = 1, \dots \dots, p$$

$$\tilde{\lambda}_i \geq 0 \qquad i = 1, \dots \dots, m$$

$$\tilde{\lambda}_i f_i(\tilde{x}) = 0 \qquad i = 1, \dots \dots, m$$

$$\nabla f_0(\tilde{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla f_i(\tilde{x}) + \sum_{i=1}^{p} \tilde{\mu}_i \nabla h_i(\tilde{x}) = 0$$

Allora $\tilde{x}$ è soluzione ottimale del problema primale, mentre la coppia $(\tilde{\lambda}, \tilde{\mu})$ è soluzione ottimale del problema duale,con gap di dualità pari a zero.
Ogni condizione delle KKT ha il suo significato:
La prima e la seconda mi dicono che $\tilde{x}$ è un punto ammissibile per il problema primale, la terza e la quarta mi dicono che finchè $\tilde{\lambda}_i \geq 0$ $L(x, \tilde{\lambda}, \tilde{\mu})$ è convessa, l'ultima

condizione mi dice che il gradiente della funzione Lagrangiana si annulla per $x = \tilde{x}$ quindi $\tilde{x}$ minimizza la funzione in x.

Quindi per ogni problema di ottimizzazione convessa con funzioni obiettivo e funzioni del vincolo differnziabili, qualsiasi punto che soddisfa le condizioni KKT sono punti ottimali con gap di dualità pari a zero.

Se la dualità è forte allora:

$$\begin{aligned} g(\lambda^*, \mu^*) &= L(x, \tilde{\lambda}, \tilde{\mu}) \\ &= f_0(\tilde{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{x}) + \sum_{i=1}^{p} \tilde{\mu}_i h_i(\tilde{x}) \\ &= f_0(\tilde{x}). \end{aligned}$$

Se un problema di ottimizzazione con la funzione obiettivo e le funzioni del vincolo differenziabili soddisfa le condizioni di Slater, allora le condizioni KKT sono condizioni necessarie e sufficienti per l'ottimalità.

Infatti le condizioni di Slater implicano che il gap di dualità è zero quindi x è ottimo se e solo se esiste una coppia $(\tilde{\lambda}, \tilde{\mu})$ che insieme con x soddisfano le condizioni KKT.

Bibliografia

- Stephen Boyd, Lieven Vandenberghe, 2003, Convex Optimization, Cambridge University Press.
- Appunti corso di Analisi Matematica 2 anno 2009/2010 professor Andrea Marson