Capitolo 5

Dualità

Si è visto nel capitolo precedente come sia possibile rappresentare un insieme convesso internamente tramite punti oppure esternamente tramite piani. Vedremo in questo capitolo che questa duplice possibilità permette di caratterizzare gli ottimi sia direttamente tramite punti interni (problema primale), sia indirettamente tramite piani esterni (problema duale). Le due rappresentazioni sono equivalenti in presenza di convessità. Se invece non c'è convessità e quindi le rappresentazioni non sono equivalenti, la comprensione del problema che si ricava dallo studio del problema duale è comunque fondamentale per risolverlo. Per questi motivi la teoria della dualità riveste un ruolo centrale nell'ottimizzazione.

Storicamente fu Lagrange [1797] che per primo ebbe l'idea di caratterizzare i minimi di una funzione soggetti a vincoli di uguaglianza tramite una nuova funzione, ottenuta aggiungendo alla funzione da minimizzare una combinazione lineare dei vincoli. Questa nuova funzione prese poi il nome di Lagrangiana. Il caso affrontato da Lagrange, in connessione con problemi di meccanica, fu limitato a condizioni di tipo differenziale e quindi locale. L'estensione a vincoli di diseguaglianza fu poi condotta da Fourier [1798] e Cournot [1826]. Il lavoro successivo più importante anche se limitato al caso di problemi lineari fu il celebre lemma di Farkas [1895]. Fu solo con Karush [1939] e Kuhn e Tucker [1951] che fu affrontato anche il caso di vincoli non lineari di diseguaglianza. Una panoramica molto esauriente degli studi che nel passato hanno poi condotto alla teoria della dualità si può trovare in Schrijver [1986].

5.1. Problemi primale e duale

La teoria della dualità si applica a problemi che possono essere posti nella seguente forma:

5.1 Definizione. *Si definisce* problema primale *il seguente problema di minimizzazione:*

$$\begin{aligned} v := \inf f(x) & \\ g(x) \le 0 & \\ x \in X & \end{aligned} \tag{5.1}$$

dove $f : E \to \mathbb{R}$, $g : E \to \mathbb{R}^m$, $X \subset E$. ∎

Si suppone quindi che una parte dei vincoli caratterizzanti l'insieme ammissibile possa essere rappresentata come un insieme di m diseguaglianze. Il vincolo $g(x) \le 0$ prende il nome di vincolo *esplicito* mentre il vincolo $x \in X$ viene detto *implicito*. Si sottolinea che la suddivisione dei vincoli in impliciti ed espliciti è arbitraria, nel senso che si possono introdurre nel vincolo implicito un numero qualsiasi di diseguaglianze, purché il vincolo esplicito contenga almeno una diseguaglianza.

La caratterizzazione dei punti ottimi è fatta tramite un altro problema, ricavato da quello primale, che viene detto *duale*. Per introdurre il problema duale si definiscono le seguenti funzioni:

5.2 Definizione. *Si dice* funzione Lagrangiana *la funzione*

$$L : E \times \mathbb{R}^m \to \mathbb{R} \qquad L(x,u) := f(x) + u\, g(x)$$

e si dice funzione duale *la funzione :*

$$L : \mathbb{R}^m \to \mathbb{R} \qquad L(u) := \inf_{x \in X} L(x,u)$$

∎

La variabile u viene indicata come *variabile duale*. Sia inoltre

$$X(u) := \{x \in X : L(x,u) = L(u)\}$$

l'insieme degli x che minimizzano la Lagrangiana per un fissato valore della variabile duale.

Siccome $L(x,u)$ è funzione affine di u, $L(u)$ è concava. Pur non essendo strettamente necessario, è tuttavia conveniente indicare in modo esplicito il dominio effettivo di $L(u)$. Sia quindi $U := \operatorname{dom} L(u)$. Possiamo a questo punto definire il seguente problema:

5.3 Definizione. *Si definisce* problema duale *il seguente problema di massimizzazione:*

$$\begin{aligned} d \;:=\; & \sup L(u) \\ & u \in U \\ & u \geq 0 \end{aligned} \tag{5.2}$$

∎

Nel problema duale viene definito come vincolo esplicito il vincolo $u \in U$ ed implicito il vincolo $u \geq 0$. Il perché di questa definizione risulterà chiaro più avanti. La relazione fondamentale che lega il problema primale e quello duale è data dal seguente risultato, che prende il nome di *dualità debole*:

5.4 Teorema. (dualità debole) $d \leq v$

Dimostrazione.

$$L(u) = \inf_{x \in X} f(x) + ug(x) \leq f(x) + ug(x) \leq f(x)$$

Le due diseguaglianze sono valide per qualsiasi x e u ammissibili rispettivamente per il primale e il duale, da cui la tesi. ∎

In generale vale quindi $d \leq v$. L'interesse maggiore tuttavia ricade su quei problemi per i quali $v = d$, o quanto meno la quantità $(v - d)$ sia piccola. La differenza $(v - d)$ prende il nome di *scarto di dualità* (*duality gap*).

5.5 Definizione. (dualità forte) *Si dice che vale la* dualità forte *se lo scarto di dualità è nullo e i problemi primale e duale ammettono ottimi.* ∎

È importante far notare che lo scarto di dualità non è una proprietà intrinseca del problema primale, ma dipende anche da come l'insieme ammissibile F del problema primale viene rappresentato tramite i vincoli impliciti ed espliciti. Sia ad esempio $F \subset X' \subset X$. Ovviamente il problema primale è equivalente al seguente problema:

$$\begin{aligned} v \,=\, \inf\, & f(x) \\ & g(x) \leq 0 \\ & x \in X' \end{aligned}$$

Il duale tuttavia è un problema differente e vale $d' \geq d$, come è facile verificare. Se addirittura $X' = F$ e il primale ammette ottimo allora vale la dualità forte (infatti $L'(0) = d' = v$). È chiaro che operare una restrizione del tipo $F \subset X' \subset X$ non è sempre facile e può addirittura rappresentare un problema più difficile di quello che si deve risolvere.

Si noti che se il problema primale è illimitato ($v = -\infty$) allora necessariamente $L(u) = -\infty$, $\forall u \geq 0$, e quindi il duale non è ammissibile. Viceversa il primale non è ammissibile se il duale è illimitato. Si tenga ancora presente che primale e duale possono essere ambedue non ammissibili come dimostra il seguente semplice esempio: $f(x_1, x_2) := x_1$, $\quad g(x_1, x_2) := x_2 + 1$, $\quad X := \{x \in \mathbb{R}^2 : x_2 \geq 0\}$.

Nel problema primale possono anche essere presenti dei vincoli del tipo $g_i(x) = 0$ (con $g_i : \mathbb{R} \to \mathbb{R}$). In teoria un tale vincolo può essere facilmente riformulato tramite la coppia di vincoli $g_i(x) \leq 0$ e $-g_i(x) \leq 0$, in modo da renderlo compatibile con la forma data di problema primale. In tal modo si introducono due variabli duali $u_i^+ \geq 0$ e $u_i^- \geq 0$ che però compaiono sempre nella Lagrangiana come differenza $u_i^+ - u_i^-$. Pertanto indicando $u_i := u_i^+ - u_i^-$ si ha direttamente un problema duale senza il vincolo di non negatività sulla variabile u_i. Indicando con J l'insieme di indici corrispondenti alle diseguaglianze del vincolo esplicito, il problema duale è allora esprimibile come

$$\begin{aligned} d \,=\, \sup\, & L(u) \\ & u \in U \\ & u_i \geq 0 \quad i \in J \end{aligned}$$

Per comprendere meglio la natura del problema duale è utile costruirne a sua volta il duale. Per semplicità di trattazione consideriamo soltanto il caso di un insieme X finito i cui elementi vengano enumerati come $x^1, x^2, \ldots, x^p$. Allora il problema duale (5.2) può essere riscritto come:

$$\begin{aligned} d = \sup\, & L \\ & L \leq f(x^k) + u\, g(x^k) \qquad k = 1, \ldots, p \\ & u \geq 0 \end{aligned}$$

o, equivalentemente,

$$\begin{aligned} -d = \inf\, & -L \\ & L \leq f(x^k) + u\, g(x^k) \qquad k = 1, \ldots, p \\ & u \geq 0 \end{aligned} \tag{5.3}$$

Prendendo $u \geq 0$ come vincolo implicito e gli altri p vincoli come espliciti, ai quali associamo le variabili duali α_k, $k := 1, \ldots, p$, e indicando con $\tilde{L}(\alpha)$ la funzione duale di (5.3), abbiamo:

$$\tilde{L}(\alpha) = \inf_{\substack{L \\ u \geq 0}} -L + \sum_{k=1}^{p} \alpha_k \left(L - f(x^k) - u\, g(x^k)\right) =$$

$$-\sum_{k=1}^{p} \alpha_k\, f(x^k) + \inf_{L}(-1 + \sum_{k=1}^{p} \alpha_k)\, L + \inf_{u \geq 0} -u \sum_{k=1}^{p} \alpha_k\, g(x^k)$$

da cui il dominio effettivo delle variabili duali α_k risulta essere

$$A = \left\{ \alpha : \sum_{k=1}^{p} \alpha_k = 1,\ \sum_{k=1}^{p} \alpha_k\, g(x^k) \leq 0 \right\}$$

e, per $\alpha \in A$, si ha

$$\tilde{L}(\alpha) = -\sum_{k=1}^{p} \alpha_k\, f(x^k)$$

Quindi il duale di (5.3) risulta

$$\begin{aligned} -dd := \sup\ & -\sum_{k=1}^{p} \alpha_k\, f(x^k) \qquad (\text{ovvero } dd = \inf \sum_{k=1}^{p} \alpha_k\, f(x^k)) \\ & \sum_{k=1}^{p} \alpha_k = 1 \\ & \sum_{k=1}^{p} \alpha_k\, g(x^k) \leq 0 \\ & \alpha_k \geq 0, \qquad \forall k \end{aligned} \tag{5.4}$$

Come si vede si tratta di una versione rilassata del problema primale, nel senso che ogni soluzione di (5.1) (nel caso considerato di X finito) corrisponde ad una soluzione di (5.4), e cioè a quella con un particolare $\alpha_j = 1$ e tutti gli altri uguali a 0, mentre in (5.4) vi sono infinite soluzioni non presenti in (5.1), e cioè tutte quelle in cui vi sono valori di α_j strettamente compresi fra 0 e 1. Quindi $dd \leq v$. Inoltre per la dualità debole fra problema duale e duale del duale si ha $-dd \leq -d$ e quindi otteniamo $d \leq dd \leq v$. Si vede allora che se il rilassamento (5.4) è effettivo, nel senso che $dd < v$, non vi può essere dualità forte. Si può dimostrare in base ai risultati della sezione successiva che vale la dualità forte fra duale e duale del duale e quindi $d = dd$.

Si noti che le variabili α sono coefficienti di una combinazione convessa e quindi il rilassamento (5.4), e di conseguenza anche il duale a causa della dualità forte fra duale e duale del duale, corrisponde ad una versione 'convessificata' del problema (5.1). Allora se il valore della funzione obiettivo si abbassa convessificando il problema, non vi può essere dualità forte. Da queste considerazioni risulta evidente il ruolo importante che riveste la convessità. Preciseremo meglio questo fatto nella sezione 3. È interessante notare che, se il vincolo implicito del primale, anziché da un insieme finito X, è costituito da $\operatorname{conv} X$, e cioè da un poliedro con vertici X e le funzioni f e g sono lineari, allora si ottiene lo stesso problema duale (5.3) ed è quindi facile vedere che il duale del duale è esattamente il primale. Preciseremo meglio queste proprietà nel capitolo sulla Programmazione Lineare.

5.2. Condizioni di ottimalità

Daremo prima delle condizioni di ottimalità per il problema duale. Poi vedremo che sotto particolari ipotesi queste condizioni possono diventare condizioni globali di ottimalità per la coppia di problemi primale–duale.

Essendo la funzione duale concava, è relativamente facile dare delle condizioni di ottimalità per il problema duale. Se il problema primale ha nel vincolo esplicito delle diseguaglianze, il problema duale contiene dei vincoli $u_i \geq 0$ che restringono il dominio effettivo di $L(u)$. Pertanto la condizione di ottimalità $0 \in \partial L_u$ va modificata tenendo conto del vincolo. Sia quindi $J(u) := \{j \in J : u_j = 0\}$ dove J è l'insieme precedentemente definito. Il cono delle direzioni ammissibili da u è $D(u) = \{h \in \mathbb{R}^m : h_i \geq 0,\ i \in J(u)\}$ e il suo polare è dato da

$$D^*(u) = \{w \in \mathbb{R}^m : w_i \leq 0 \text{ se } i \in J(u),\ w_i = 0 \text{ se } i \notin J(u)\}$$

Quindi la condizione d'ottimalità $\partial L_u \cap D^*(u) \neq \emptyset$ (teorema 4.127 nella versione per le funzioni concave) si può esprimere come:

5.6 Teorema. *$\hat{u}$ è ottimo duale se e solo se esiste $\hat{\gamma} \in \partial L_u$ tale che:*

$$\begin{array}{lll} 1) & \hat{u}\,\hat{\gamma} = 0 & \\ 2) & \hat{\gamma}_i \leq 0 & i \in J \\ & \hat{\gamma}_i = 0 & i \notin J \end{array}$$

Dimostrazione. La condizione d'ottimalità $\partial L_{\hat{u}} \cap D^*(\hat{u}) \neq \emptyset$ implica l'esistenza di un subgradiente $\hat{\gamma}$ tale che $\hat{\gamma}_i \leq 0$ se $i \in J(\hat{u})$ e $\hat{\gamma}_i = 0$ se $i \notin J(\hat{u})$. Quindi la terza condizione è verificata. Dalla definizione di $J(\hat{u})$ risulta $\hat{u}_i\,\hat{\gamma}_i = 0$, $\forall i$, cioè $\hat{u}\,\hat{\gamma} = 0$.

Per dimostrare la sufficienza è immediato notare che $\hat{u}_i\,\hat{\gamma}_i = 0$ per ogni $i \notin J(\hat{u})$ e $\hat{u}_i\,\hat{\gamma}_i \leq 0$ per ogni $i \in J(\hat{u})$. Però la prima condizione impone che $\hat{u}_i\,\hat{\gamma}_i = 0$ anche per $i \in J(\hat{u})$. Quindi o $\hat{u}_i$ o $\hat{\gamma}_i$ deve essere uguale a zero in modo coerente con la definizione di $D^*(\hat{u})$ e la condizione d'ottimalità. ∎

Ora si noti che vale

$$\partial L_u \supset \overline{\text{conv}}\{\partial L(x,\cdot)_u : x \in X(u)\} = \overline{\text{conv}}\{g(x) : x \in X(u)\}$$

Se il vettore $\hat{\gamma}$ appartiene a $\overline{\text{conv}}\{g(x) : x \in X(u)\}$ allora $\hat{\gamma}$ può essere caratterizzato come combinazione convessa dei punti $\{g(x) : x \in X(u)\}$. Se tale combinazione convessa è stretta, non si giunge a nessuna conclusione interessante per il problema primale. Se però avviene $\hat{\gamma} = g(x)$ per qualche $x \in X(\hat{u})$ allora si può concludere, come vedremo fra poco, che $\hat{x}$ è ottimo primale e vale la dualità forte. Si consideri dapprima la seguente definizione:

5.7 Definizione. *Un punto $(\hat{x}, \hat{u}) \in X \times \mathbb{R}^m_+$ soddisfa le* condizioni globali di ottimalità (CGO) *se:*

$$\begin{array}{llll} 1) & \hat{x} \in X(\hat{u}) & & \\ 2) & \hat{u}\,g(\hat{x}) & = 0 & \\ 3) & g_i(\hat{x}) & \leq 0 & i \in J \\ & g_i(\hat{x}) & = 0 & i \notin J \end{array}$$

∎

Le CGO costituiscono quindi una riformulazione delle precedenti condizioni di ottimalità duale imponendo l'ulteriore condizione che il subgradiente $\hat{\gamma}$ sia direttamente esprimibile come $g(\hat{x})$ e non già come combinazione convessa di tali punti.

Riassumendo, si vede che $\hat{x}$ deve minimizzare la Lagrangiana per il valore $\hat{u}$ della variabile duale (dalla condizione 1), che $\hat{x}$ deve essere ammissibile (dalla condizione 3) e che per ogni $i = 1, \ldots, m$, deve valere $\hat{u}_i\, g_i(\hat{x}) = 0$ (dalla condizione 2). In altre parole o $\hat{u}_i$ o $g_i(\hat{x})$ deve essere nullo e quindi se una variabile duale è strettamente positiva, il vincolo corrispondente deve essere soddisfatto come eguaglianza, mentre se il vincolo non è attivo in $\hat{x}$ (definiamo un vincolo g_i come *attivo* in x se $g_i(x) = 0$) la corrispondente variabile duale deve essere nulla. Per questo motivo la condizione 2) delle CGO prende il nome di *condizione di complementarità*. La condizione di complementarità diventa superflua, perché banalmente verificata, se il vincolo esplicito è tutto composto da equazioni.

La seguente definizione rivela inoltre il ruolo della funzione Lagrangiana ai fini dell'ottimalità della coppia di problemi primale–duale.

5.8 Definizione. *Un punto* $(\hat{x}, \hat{u}) \in X \times \mathbb{R}^m_+$ *si dice* punto di sella *per la Lagrangiana* $L(x, u)$ *se*

$$L(\hat{x}, u) \le L(\hat{x}, \hat{u}) \le L(x, \hat{u}) \qquad \forall x \in X \quad \forall u \ge 0$$

∎

Si noti che nella definizione non compaiono i vincoli espliciti dei due problemi e che, rispetto ai vincoli impliciti, un punto di sella si comporta come minimo in x e come massimo in u.

Ricordiamo ancora che vale la dualità forte se esistono $\hat{x}$ e $\hat{u}$ ammissibili nei rispettivi problemi tali che $f(\hat{x}) = L(\hat{u})$ e quindi necessariamente ottimi. Faremo ora vedere che le CGO, la condizione di punto di sella e la dualità forte sono condizioni equivalenti fra di loro.

5.9 Teorema. $(\hat{x}, \hat{u}) \in X \times \mathbb{R}^m_+$ *è un punto di sella se e solo se soddisfa le CGO.*

Dimostrazione. ($\Rightarrow$) Siccome $L(\hat{x}, \hat{u}) \le L(x, \hat{u})$, $\forall x \in X$, la CGO 1) vale banalmente. Inoltre $L(\hat{x}, u) \le L(\hat{x}, \hat{u})$ implica $(u - \hat{u})\, g(\hat{x}) \le 0$, $\forall u \ge 0$. Si scelga in particolare $u = \hat{u} + w$, $w \ge 0$, quindi $w\, g(\hat{x}) \le 0$, $\forall w \ge 0$, che implica $g(\hat{x}) \le 0$ e quindi CGO 3) è verificata. Quindi $\hat{u}\, g(\hat{x}) \le 0$. D'altra parte $u\, g(\hat{x}) \le \hat{u}\, g(\hat{x})$ implica $\hat{u}\, g(\hat{x}) \ge \sup_{u \ge 0} u\, g(\hat{x}) \ge 0$, che, combinato con il risultato precedente fornisce la CGO 2).

($\Leftarrow$) La CGO 1) è equivalente a $L(\hat{x}, \hat{u}) \le L(x, \hat{u}) \quad \forall x \in X$. Inoltre si ha

$$\begin{aligned} f(\hat{x}) &= f(\hat{x}) + \hat{u}\, g(\hat{x}) = L(\hat{x}, \hat{u}) && \text{in base a CGO 2) e} \\ f(\hat{x}) &\ge f(\hat{x}) + u\, g(\hat{x}) = L(\hat{x}, u) \quad \forall u \ge 0 && \text{in base a CGO 3)} \end{aligned}$$

∎

5.10 Teorema. $(\hat{x}, \hat{u}) \in X \times \mathbb{R}^m_+$ *è un punto di sella se e solo se vale la dualità forte per* $(\hat{x}, \hat{u})$.

Dimostrazione. ($\Rightarrow$) Da $L(\hat{x}, \hat{u}) \le L(x, \hat{u})$ discende $L(\hat{x}, \hat{u}) = L(\hat{u})$ (e quindi $\hat{u} \in U$). Inoltre $L(\hat{x}, \hat{u}) = f(\hat{x})$ in quanto $\hat{u}\, g(\hat{x}) = 0$ per il teorema precedente. Sempre dal teorema precedente $\hat{x}$ è ammissibile e quindi vale la dualità forte.

($\Leftarrow$) $f(\hat{x}) = L(\hat{u}) \le L(x, \hat{u})$, $\forall x \in X$ implica $f(\hat{x}) \le f(x) + \hat{u}\, g(x)$ cioè $0 \le \hat{u}\, g(\hat{x})$. Ma siccome $\hat{u} \ge 0$ e $g(\hat{x}) \le 0$, $\hat{u}\, g(\hat{u}) \le 0$, cioè $\hat{u}\, g(\hat{x}) = 0$, e allora $f(\hat{x}) = f(\hat{x}) + \hat{u}\, g(\hat{x}) = L(\hat{x}, \hat{u})$. Quindi $L(\hat{x}, \hat{u}) \le L(x, \hat{u})$. Ovviamente $f(\hat{x}) + u\, g(\hat{x}) \le f(\hat{x})$, $\forall u \ge 0$, e quindi $L(\hat{x}, u) \le L(\hat{x}, \hat{u})$. ∎

Le condizioni enunciate sono condizioni di tipo sufficiente di ottimalità. In altri termini può succedere che per un ottimo primale x non esista nessun u ammissibile tale che (x, u) soddisfi le CGO. Se ciò si verifica è perché lo scarto di dualità è strettamente positivo.

Si supponga che per un dato problema valga la dualità forte. Sappiamo che questa proprietà dipende da come si sono scelti i vincoli impliciti ed espliciti. È interessante vedere tuttavia che la proprietà viene mantenuta se parte dei vincoli espliciti viene considerata vincolo implicito. Infatti sia dato

$$\begin{aligned} \inf\ & f(x) \\ & g_1(x) \le 0 \\ & g_2(x) \le 0 \\ & x \in X \end{aligned}$$

e si considerino le due funzioni duali

$$L(u_1, u_2) := \inf_{x \in X} f(x) + u_1\, g_1(x) + u_2\, g_2(x) \qquad \tilde{L}(u_1) := \inf_{\substack{x \in X \\ g_2(x) \le 0}} f(x) + u_1\, g_1(x)$$

5.11 Teorema. *Se $(\hat{x}, \hat{u}_1, \hat{u}_2)$ punto di sella per la funzione duale $L(u_1, u_2)$ allora $(\hat{x}, \hat{u}_1)$ è punto di sella per la funzione duale $\tilde{L}(u_1)$.*

Dimostrazione.

$$\begin{aligned} \tilde{L}(\hat{u}_1) \le v = L(\hat{u}_1, \hat{u}_2) & \le f(x) + \hat{u}_1 g_1(x) + \hat{u}_2 g_2(x) \le && \forall x \in X \\ & \le f(x) + \hat{u}_1 g_1(x) && \forall x \in X : g_2(x) \le 0 \end{aligned}$$

da cui $\tilde{L}(\hat{u}_1) \le v \le \tilde{L}(\hat{u}_1)$ e cioè $v = \tilde{L}(\hat{u}_1) = f(\hat{x})$. ∎

L'implicazione inversa non è vera in generale, cioè se $(\hat{x}, \hat{u}_1)$ è un punto di sella, potrebbe non esistere nessun $\hat{u}_2$ tale che $(\hat{x}, \hat{u}_1, \hat{u}_2)$ sia punto di sella. Sia ad esempio dato un problema per cui non valga la dualità forte. Si indichi con g_2 il vincolo esplicito di questo problema. Si aggiunga una diseguaglianza, che indichiamo con g_1, che non abbia effetto sull'istanza, sia cioè $g_1(x) < 0$ per ogni x ammissibile. Se ora g_2 viene considerato vincolo implicito la dualità forte vale banalmente con valore duale ottimo $\hat{u}_1 = 0$, mentre non vale se g_2 è vincolo esplicito. Vedremo più avanti che l'implicazione inversa vale in condizioni di convessità.

5.12 Esercizio. Riformulare la funzione Lagrangiana, il problema duale e le condizioni di ottimalità per un problema formulato come max $f(x)$ con il vincolo $g(x) \ge 0$, $x \in X$. ∎

5.3. Esempi

L'approccio lagrangiano alla risoluzione di un problema è uno dei più diffusi e più estesamente applicabili. Si tratta di costruire un opportuno problema duale e di risolverlo, eventualmente anche in modo approssimato, e di ricavare da questa soluzione il massimo d'informazione per ottenere la soluzione del problema primale. Si ritiene utile presentare un'ampia rassegna di esempi di applicazione in modo da far comprendere il più possibile la portata e l'utilità delle tecniche di tipo lagrangiano.

5.13 Esempio. Consideriamo il problema di knapsack continuo dell'esempio 1.74:

$$\begin{aligned} \max \quad & 24\,x_1 + 5\,x_2 + 26\,x_3 + 28\,x_3 \\ & 8\,x_1 + 5\,x_2 + 13\,x_3 + \;\; 7\,x_4 \le 26 \\ & 0 \le x_i \le 1 \quad \forall i \end{aligned}$$

Per applicare i risultati precedenti trasformiamo il problema in una minimizzazione, cioè la funzione obiettiva sia $\min - \sum_i c_i\,x_i$. La scelta naturale è di porre il vincolo $0 \le x_i \le 1$ come vincolo implicito; pertanto

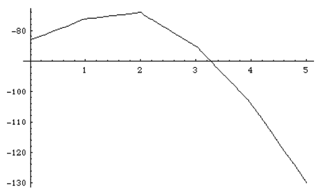
$$L(u) = \min_{0 \le x_i \le 1, \forall i} -24\,x_1 - 5\,x_2 - 26\,x_3 - 28\,x_4 + u\,(8\,x_1 + 5\,x_2 + 13\,x_3 + 7\,x_4 - 26)$$

Si vede che l'operazione di minimo può essere eseguita indipendentemente in ogni variabile. Quindi

$$\begin{aligned} L(u) = -26\,u + \min_{0 \le x_1 \le 1} (-24 + 8\,u)\,x_1 + \min_{0 \le x_2 \le 1} (-5 + 5\,u)\,x_2 + \\ + \min_{0 \le x_3 \le 1} (-26 + 13\,u)\,x_3 + \min_{0 \le x_4 \le 1} (-28 + 7\,u)\,x_4 \end{aligned}$$

I minimi sono raggiunti per $x_i = 0$ se il coefficiente di x_i è positivo, e per $x_i = 1$ se il coefficiente è negativo. Si noti che, se il coefficiente di x_i è nullo, il minimo è raggiunto da un qualsiasi valore di x_i compreso fra 0 e 1. La funzione duale è lineare a tratti con punti di rottura dati da quei valori di u che annullano i coefficienti delle variabili, e quindi in questo esempio, $u = 1$, $u = 2$, $u = 3$ e $u = 4$. Inoltre si ha $u \ge 0$ perché il problema presenta un vincolo di diseguaglianza e $U = \mathbb{R}$ perché X è un insieme limitato. Si ottiene quindi facilmente

$$L(u) = \begin{cases} 7\,u - 83 & \text{se } 0 \le u \le 1 \\ 2\,u - 78 & \text{se } 1 \le u \le 2 \\ -11\,u - 52 & \text{se } 2 \le u \le 3 \\ -19\,u - 28 & \text{se } 3 \le u \le 4 \\ -26\,u & \text{se } 4 \le u \end{cases}$$



Si vede subito che $\hat{u} = 2$ è l'ottimo duale; infatti $\partial L(2) = [-11, 2] \ni 0$. Dobbiamo adesso cercare elementi ammissibili del problema primale che soddisfino le CGO. Innanzitutto dovremo cercare tali candidati nell'insieme $X(\hat{u})$ a causa della prima delle CGO. Quindi $\hat{x}_1 = 1$, $\hat{x}_2 = 0$, $\hat{x}_4 = 1$ mentre $\hat{x}_3$ può assumere qualsiasi valore fra 0 e 1. Siccome $\hat{u} > 0$ la condizione di complementarità impone $g(\hat{x}) = 0$ cioè $8 \cdot 1 + 5 \cdot 0 + 13\,\hat{x}_3 + 7 \cdot 1 = 26$, da cui $\hat{x}_3 = 11/13$, che, essendo compreso fra 0 e 1, garantisce l'esistenza di un punto di sella. Si noti che $L(2) = -74 = -(24 + 26 \cdot 11/13 + 28)$, come verifica della dualità forte, appena dimostrata. ∎

5.14 Esercizio. (prima parte, facile) Basandosi sulla struttura del problema di knapsack continuo e della sua soluzione, si derivi un algoritmo di complessità $O(n \log n)$ che calcoli l'ottimo. Si riveda anche l'esercizio 1.75. (seconda parte, difficile) Questo risultato di complessità si può migliorare sfruttando il fatto che il valore mediano di un insieme di numeri si può calcolare in tempo lineare. Questo risultato va applicato all'insieme $\{a_i/b_i\}$. Poi bisogna trovare l'ottimo duale con una ricerca binaria. Dimostrare che il knapsack continuo si può risolvere con complessità lineare. ∎

5.15 Esempio. Consideriamo nuovamente il precedente problema di knapsack nella sua versione discreta originaria:

$$\begin{aligned} \max \quad & 24\,x_1 + 5\,x_2 + 26\,x_3 + 28\,x_3 \\ & \;8\,x_1 + 5\,x_2 + 13\,x_3 + \;\;7\,x_4 \le 26 \\ & x_i \in \{0,1\} \quad \forall i \end{aligned}$$

La funzione duale è pertanto

$$L(u) = \min_{x_i \in \{0,1\}, \forall i} -24\,x_1 - 5\,x_2 - 26\,x_3 - 28\,x_4 + u\,(8\,x_1 + 5\,x_2 + 13\,x_3 + 7\,x_4 - 26)$$

che è separabile anche in questo caso e quindi

$$\begin{aligned} L(u) = -26\,u + & \min_{x_1 \in \{0,1\}} (-24 + 8\,u)\,x_1 + \min_{x_2 \in \{0,1\}} (-5 + 5\,u)\,x_2 + \\ & + \min_{x_3 \in \{0,1\}} (-26 + 13\,u)\,x_3 + \min_{x_4 \in \{0,1\}} (-28 + 7\,u)\,x_4 \end{aligned}$$

Si noti che la funzione duale è la medesima dell'esempio precedente, quindi ha lo stesso ottimo. Tuttavia in questo caso non è possibile verificare le CGO perché non esiste $\hat{x}_3 \in \{0,1\}$ per cui $g(\hat{x}) = 0$. Quindi c'è scarto di dualità.

Il fatto che ci sia scarto di dualità è dovuto agli elementi discreti del problema, come vedremo nella sezione successiva. Vedremo più avanti che un modo per aggirare questa difficoltà consiste nell'adottare delle tecniche di enumerazione implicita, che prevedono il calcolo del problema duale su sottoinsiemi di X che ne costituiscano una partizione. Senza entrare ora nel merito di tali tecniche, limitiamoci a titolo esemplificativo a calcolare la funzione duale su alcuni particolari sottoinsiemi di X. Sia pertanto

$$\begin{aligned} X_1 &:= \{x \in \{0,1\}^4 : x_1 = 0\} \\ X_2 &:= \{x \in \{0,1\}^4 : x_1 = 1,\ x_3 = 0\} \\ X_3 &:= \{x \in \{0,1\}^4 : x_1 = 1,\ x_3 = 1,\ x_4 = 0\} \\ X_4 &:= \{x \in \{0,1\}^4 : x_1 = 1,\ x_3 = 1,\ x_4 = 1\} \end{aligned}$$

Si noti che X_1, X_2, X_3, X_4 costituiscono una partizione di X. Si indichi inoltre $F_i := X_i \cap \{x : g(x) \le 0\}$. Ovviamente i sottoinsiemi F_i costituiscono una partizione dell'insieme ammissibile F. Si ha quindi

$$L_1(u) = -26\,u + \min_{x_2 \in \{0,1\}} (-5 + 5\,u)\,x_2 + \min_{x_3 \in \{0,1\}} (-26 + 13\,u)\,x_3 + \min_{x_4 \in \{0,1\}} (-28 + 7\,u)\,x_4$$

per cui $d_1 = L_1(0) = -59$ e $X_1(0) = \hat{x}^1 = (0\,1\,1\,1)$. Si verifica che $g(\hat{x}^1) < 0$. Ma $\hat{u}^1 = 0$ e quindi le CGO sono verificate. Questo significa che $\hat{x}^1$ è ottimo relativamente al sottoinsieme F_1 con valore ottimo $v_1 = -59$.

$$L_2(u) = -24 - 18\,u + \min_{x_2 \in \{0,1\}} (-5 + 5\,u)\,x_2 + \min_{x_4 \in \{0,1\}} (-28 + 7\,u)\,x_4$$

Anche in questo caso l'ottimo duale si ottiene per $\hat{u}^1 = 0$, con $d_2 = L_2(0) = -57$ e $X_2(0) = \hat{x}^2 = (1\,1\,0\,1)$. Le CGO sono verificate, quindi $\hat{x}^2$ è ottimo relativamente a F_2 con valore ottimo $v_2 = -57 > v_1$.

$$L_3(u) = -50 - 5\,u + \min_{x_2 \in \{0,1\}} (-5 + 5\,u)\,x_2$$

In questo caso l'ottimo duale non è unico; infatti tutti i valori appartenenti all'intervallo $[0,1]$ sono ottimi duali con valore ottimo $d_3 = -55$. Si noti che $d_3 > v_1$, e siccome $v_3 \geq d_3$ per la dualità debole, si ha che $v_3 > v_1$. Quindi possiamo concludere che è inutile cercare l'ottimo all'interno di F_3 perché il suo valore sarà in ogni caso peggiore della migliore soluzione finora trovata.

$$L_4(u) = -78 + 2\,u + \min_{x_2 \in \{0,1\}} (-5 + 5\,u)\,x_2$$

È immediato verificare che $L_4(u)$ è illimitata e quindi $F_4 = \emptyset$. Possiamo pertanto concludere che $\hat{x}^1$ è ottimo globale del problema. ∎

5.16 Esempio. Sia dato il problema

$$\begin{aligned} \min \quad & x_1 + \ \ x_2 + \ \ x_3 + \ \ x_4 \\ & x_1 + 3\,x_2 - 2\,x_3 - 3\,x_4 = 50 \\ & 2\,x_1 - \ \ x_2 + \ \ x_3 - 2\,x_4 = 50 \\ & x_i \geq 0\,, \text{ intero } \forall i \end{aligned}$$

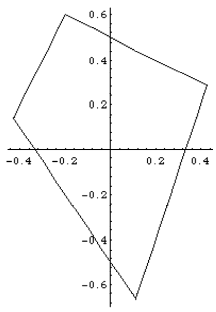
La funzione duale risulta essere:

$$\begin{aligned} L(u_1,u_2) = 50\,u_1 + 50\,u_2 &+ \inf_{x_1 \in Z_+} (1 - u_1 - 2\,u_2)\,x_1 + \inf_{x_2 \in Z_+} (1 - 3\,u_1 + u_2)\,x_2 \\ &+ \inf_{x_3 \in Z_+} (1 + 2\,u_1 - u_2)\,x_3 + \inf_{x_4 \in Z_+} (1 + 3\,u_1 + 2\,u_2)\,x_4 \end{aligned}$$

In questo esempio X è illimitato; come conseguenza $L(u) \to -\infty$ se solo uno dei coefficienti di x_i è negativo. Mentre se tale coefficiente è positivo l'infimo viene realizzato per $x_i = 0$. Allora il problema duale è

$$\begin{aligned} \max \quad & 50\,u_1 + 50\,u_2 \\ & u_1 + \ 2\,u_2 \leq 1 \\ & 3\,u_1 - \quad u_2 \leq 1 \\ & -2\,u_1 + \quad u_2 \leq 1 \\ & -3\,u_1 - \ 2\,u_2 \leq 1 \end{aligned} \tag{5.5}$$

Si noti che non è presente il vincolo di non negatività su u perchè il problema primale ha vincoli d'uguaglianza. L'insieme ammissibile di (5.5) è indicato in figura.



Graficamente si vede che il punto $\hat{u} = (\frac{3}{7}, \frac{2}{7})$ è ottimo del problema duale. Si può inoltre verificare l'ottimalità di $\hat{u}$ da

$$\partial L(\hat{u}) = \begin{pmatrix} 50 \\ 50 \end{pmatrix} - \text{cone}\left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ -1 \end{pmatrix} \right\} \ni 0$$

Si ha $L(\hat{u}) = 250/7$ che non è intero, quindi possiamo già concludere che non c'è dualità forte. Tuttavia cerchiamo ugualmente di verificare le CGO perché otterremo informazione utile per restringere l'insieme del vincolo implicito. Si ottiene

$$X(\hat{u}) = \{x \in Z_+^4 : x_3 = 0,\, x_4 = 0\}$$

Quindi dobbiamo verificare se esistono soluzioni intere positive del sistema

$$\begin{aligned} x_1 + 3\,x_2 &= 50 \\ 2\,x_1 - x_2 &= 50 \end{aligned} \tag{5.6}$$

Se moltiplichiamo la prima riga per due e sottraiamo la seconda, otteniamo $7\,x_2 = 50$. Quindi non esistono soluzioni intere di (5.6) e le CGO non possono essere soddisfatte.

L'approccio che verrà ora presentato per trovare una soluzione intera al problema si è dimostrato alla prova dei fatti meno efficace di altri metodi. Tuttavia ha una sua eleganza teorica che ne giustifica la presentazione, anche se solo a titolo di breve esempio. Il metodo consiste nell'operare una restrizione X' di X tale però che $F \subset X' \subset X$. A tal fine sfruttiamo il procedimento stesso che ha dimostrato la non interezza delle soluzioni di (5.6) e consideriamo il seguente omomorfismo

$$\varphi : Z^2 \to Z_7 \quad \text{definito da} \quad \begin{pmatrix} a \\ b \end{pmatrix} \mapsto 2\,a - b \bmod 7$$

e ricaviamo dal vincolo esplicito del problema la seguente equazione in Z_7:

$$\varphi\begin{pmatrix} 1 \\ 2 \end{pmatrix} x_1 + \varphi\begin{pmatrix} 3 \\ -1 \end{pmatrix} x_2 + \varphi\begin{pmatrix} -2 \\ 1 \end{pmatrix} x_3 + \varphi\begin{pmatrix} -3 \\ -2 \end{pmatrix} x_4 \equiv \varphi\begin{pmatrix} 50 \\ 50 \end{pmatrix} \bmod 7$$

che dà

$$2\,x_3 + 3\,x_4 \equiv 1 \bmod 7 \tag{5.7}$$

Come si vede l'ultima equazione gruppale è certamente soddisfatta dalle soluzioni ammissibili del problema, ma non dagli elementi del precedente insieme $X(\hat{u})$. Quindi, se aggiungiamo come vincolo implicito l'equazione gruppale (5.7) operiamo una restrizione di X, nel senso indicato prima, evitando di ottenere come candidati alla verifica delle CGO i precedenti elementi. La nuova funzione duale è pertanto

$$\begin{aligned} L(u_1, u_2) = 50\,u_1 + 50\,u_2 &+ \inf_{x_1 \in Z_+} (1 - u_1 - 2\,u_2)\,x_1 + \inf_{x_2 \in Z_+} (1 - 3\,u_1 + u_2)\,x_2 \\ &+ \inf_{\substack{x_3 \in Z_+,\, x_4 \in Z_+ \\ 2\,x_3 + 3\,x_4 \equiv 1 \bmod 7}} (1 + 2\,u_1 - u_2)\,x_3 + (1 + 3\,u_1 + 2\,u_2)\,x_4 \end{aligned}$$

Si noti che anche in questo caso si deve avere $1 + 2\,u_1 - u_2 \geq 0$ e $1 + 3\,u_1 + 2\,u_2 \geq 0$ altrimenti $L(u) \to -\infty$. Per il calcolo del terzo infimo possiamo considerare tutti i possibili valori di x_3 da 0 a 6 (di più non servono dato che dobbiamo minimizzare un'espressione con coefficienti non negativi) e calcolare il più piccolo valore non negativo di x_4 che soddisfa (5.7). Le soluzioni che si ottengono sono

x_3	0	1	2	3	4	5	6
x_4	5	2	6	3	0	4	1

Di queste soluzioni alcune sono dominate da altre, nel senso che se $x'_3 \le x''_3$ e $x'_4 \le x''_4$ allora x'' può non essere considerato nella valutazione del minimo. Rimangono da valutare quindi

x_3	0	1	4
x_4	5	2	0

Allora si ha

$$L(u_1, u_2) = 50\,u_1 + 50\,u_2 + \min\big\{(1 + 3\,u_1 + 2\,u_2)\,5\,; \\ (1 + 2\,u_1 - u_2)\,1 + (1 + 3\,u_1 + 2\,u_2)\,2\,; (1 + 2\,u_1 - u_2)\,4\big\}$$

cioè

$$L(u_1, u_2) = 50\,u_1 + 50\,u_2 + \min\{5 + 15\,u_1 + 10\,u_2\,; 3 + 8\,u_1 + 3\,u_2\,; 4 + 8\,u_1 - 4\,u_2\}$$

con u vincolata come nel caso precedente.

Si riconsideri il punto $\hat{u} = (\frac{3}{7}\,,\,\frac{2}{7})$. Si ha $L(\hat{u}) = 4 + 58\,\hat{u}_1 + 46\,\hat{u}_2$, con il minimo ottenuto solo in corrispondenza della terza funzione e quindi abbiamo

$$\partial L(\hat{u}) = \begin{pmatrix} 58 \\ 46 \end{pmatrix} - \text{cone}\left\{\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ -1 \end{pmatrix}\right\} \ni 0$$

e quindi $\hat{u}$ è ottimo duale. Adesso si ha $L(\hat{u}) = 42$ e $X(\hat{u}) = \{x \in Z^4_+ : x_3 = 4,\, x_4 = 0\}$ e quindi l'equazione (5.6) viene trasformata in

$$\begin{aligned} x_1 + 3\,x_2 &= 58 \\ 2\,x_1 - x_2 &= 46 \end{aligned}$$

che dà come soluzioni $\hat{x}_1 = 28$ e $\hat{x}_2 = 10$. Quindi le CGO sono soddisfatte e $\hat{x} = (28, 10, 4, 0)$ è ottimo primale. Si noti incidentalmente che $v = 28 + 10 + 4 = 42 = d = L(\hat{u})$. ∎

5.17 Esercizio. Si risolva l'esempio precedente togliendo il vincolo d'interezza (sempre tramite il problema duale). ∎

5.18 Esercizio. Si risolva il seguente problema

$$\begin{aligned} \min \quad & x_1 + x_2 + x_3 \\ & x_1 \qquad - x_3 \ge 1 \\ & \qquad x_2 + x_3 \ge 2 \\ & x_i \ge 0 \text{ intero} \end{aligned}$$

ponendo come vincolo esplicito le due diseguaglianze, oppure una sola delle due (vale sempre la dualità forte). ∎

5.19 ESERCIZIO. Si risolva il duale del seguente problema

$$\begin{array}{lll}
\max & x_1 \qquad\quad + 2\,x_3 + 3\,x_4 & \\
& 2\,x_1 + \;\; x_2 \qquad\;\; + \;\; x_4 \le 5 & (g_1) \\
& \qquad\; 2\,x_2 + \;\; x_3 + 2\,x_4 \le 2 & (g_2) \\
& 0 \le x_1 \le 2,\ 1 \le x_2 \le 3,\ 0 \le x_3 \le 2,\ 0 \le x_4 \le 1,\ x_i \text{ intero} & X
\end{array}$$

ponendo come vincolo esplicito g_1 e g_2, oppure solo g_1, oppure solo g_2. Si studino le relazioni fra i tre duali. ∎

5.20 ESEMPIO. Il problema del commesso viaggiatore (TSP) (esempio 1.39) può essere formulato nel seguente modo:

$$\begin{aligned}
\min \quad & c\,x \\
& A\,x = \mathbf{2} \\
& x \in X
\end{aligned}$$

dove x è il vettore d'incidenza degli archi di un generico sottografo di supporto del grafo non orientato completo K_n, A è la matrice d'incidenza nodi–archi di K_n (vedi sezione 2.2) e X corrisponde ad un insieme di sottografi che siano formati da un albero di supporto sui nodi $2, \ldots, n$, più due archi qualsiasi uscenti dal nodo 1. Il vincolo d'eguaglianza rappresenta la condizione che il grado in ogni nodo sia esattamente due. Quindi gli unici sottografi che soddisfano sia il vincolo esplicito che quello implicito sono tutti i possibili circuiti hamiltoniani di K_n. L'approccio qui presentato fu proposto da Held e Karp [1970, 1971] e si è dimostrato particolarmente efficace per grafi con diverse centinaia di nodi (per grafi più grandi si sono dimostrati migliori metodi di geometria poliedrale che esamineremo più avanti)

Sia quindi la funzione duale:

$$\begin{aligned}
L(u) &= \min_{x\in X} c\,x + u\,(\mathbf{2} - A\,x) = \mathbf{2}\,u + \min_{x\in X}(c - u\,A)\,x = \\
&= 2\sum_i u_i + \min_{x\in X}\sum_{i,j}(c_{ij} - u_i - u_j)\,x_{ij}
\end{aligned}$$

dove gli indici i e j si riferiscono ai nodi e quindi l'indice composto (ij) si riferisce all'arco incidente nei nodi i e j.

Il calcolo della funzione duale corrisponde al calcolo del minimo albero di supporto sui nodi $2, \ldots, n$, e dei due archi meno costosi uscenti dal nodo 1, rispetto ai costi modificati $(c_{ij} - u_i - u_j)$. Tale calcolo si può effettuare in modo efficiente con diversi algoritmi alternativi, come vedremo più avanti. Per ora basti sapere che il minimo albero di supporto può essere calcolato ordinando gli archi secondo costi non decrescenti, prendendo gli archi nell'ordine e tralasciando quelli che formerebbero cicli fino a formare l'albero.

Consideriamo l'istanza data da K_6 con la seguente tabellina di costi:

$$\begin{bmatrix}
0 & 68 & 36 & 99 & 5 & 21 \\
68 & 0 & 94 & 35 & 21 & 35 \\
36 & 94 & 0 & 73 & 69 & 81 \\
99 & 35 & 73 & 0 & 48 & 70 \\
5 & 21 & 69 & 48 & 0 & 76 \\
21 & 35 & 81 & 70 & 76 & 0
\end{bmatrix}$$

Il calcolo di $L(u^0)$ fornisce il quasi albero x^0 in figura 5.1. Si ha $L(u^0) = 186$. Siccome x^0 è l'unico elemento di $X(u^0)$, $\gamma := g(x^0)$ è il gradiente di $L(u)|_{u=u^0}$. In questo caso si ha $\gamma^0 = \mathbf{2} - A\,x^0 = (0, -1, 1, 1, -1, 0)$. L'origine non è quindi un ottimo del problema duale.

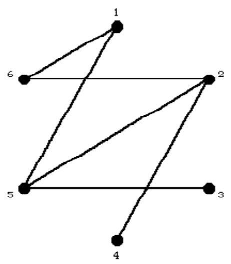


FIGURA 5.1

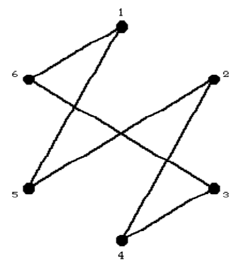


FIGURA 5.2

Per massimizzare $L(u)$ facciamo ricorso ai metodi di quasi ascesa delineati nella sezione 4.11, pag. 121.

È necessario avere una stima del valore ottimo del problema primale. Troviamo allora con un metodo euristico una soluzione che sia abbastanza buona. Ad esempio potremmo usare la soluzione $\tilde{x}$ in figura 5.2, il cui valore è $\tilde{v} := 236$. Si noti il notevole scarto fra il valore della funzione duale e $\tilde{v}$. Il valore ottimo è compreso fra questi due valori. Più si riesce a massimizzare la funzione duale, più si restringe l'intervallo entro cui si trova il valore ottimo primale. Il nuovo valore u^1 è allora dato da

$$u^1 := u^0 + \frac{\tilde{v} - L(u^0)}{\gamma^0\,\gamma^0}\,\gamma^0 = \frac{236 - 186}{4}\,(0,-1,1,1,-1,0) = (0,-12.5,12.5,12.5,-12.5,0)$$

Si ottiene $L(u^1) = 215$, con $x^1 \in X(u^1)$ in figura 5.3 e $\gamma^1 = (0,-1,1,0,0,0)$. Iterando si ottiene un nuovo valore u^2, con $L(u^2) = 219$ e $x^2 \in X(u^2)$ in figura 5.4. In una successiva iterazione si ottiene $L(u^3) = 228$ e $x^3 \in X(u^3)$ in figura 5.5. A questo punto sappiamo che $228 \le \hat{d} \le \hat{v} \le 236$ e quindi il margine d'incertezza nella valutazione dell'ottimo è diminuito di molto.

Proseguendo l'iterazione si nota che i valori cominciano ad oscillare intorno al valore 232 e quindi abbiamo ragione di credere che 236 non sia l'ottimo valore duale. Infatti si può trovare con altri metodi che l'ottimo valore duale è $\hat{d} = 234$ e che l'ottimo duale è

$$\hat{u} = \{0, 0.8, 44.8, 27.8, 13.8, 29.8\}$$

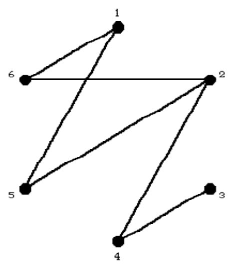


FIGURA 5.3

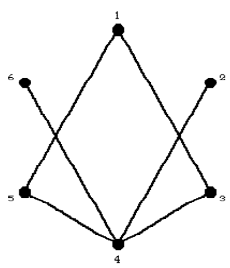


FIGURA 5.4

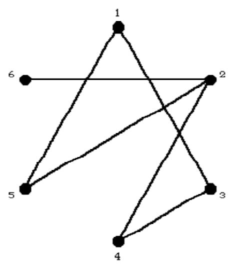


FIGURA 5.5

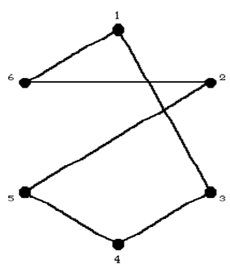


FIGURA 5.6

In corrispondenza di questo valore si ottengono 15 quasi alberi che appartengono a $X(\hat{u})$. Fra questi quasi alberi esiste anche il quasi albero in figura 5.6 che è un circuito, necessariamente ottimo. In questo caso vale allora la dualità forte. Non si creda da questo esempio che sia frequente verificare la dualità forte per il TSP. Anzi, normalmente questo non avviene. Cioè nell'insieme $X(\hat{u})$ non è presente un circuito e invece sono presenti alcuni quasi alberi la cui "combinazione convessa" è un circuito. Più esattamente la combinazione convessa dei subgradienti corrispondenti è uguale a 0, il che garantisce l'ottimalità duale in base alle proprietà dei massimi di una funzione concava. Ad esempio per l'istanza dell'esempio sono presenti in $X(\hat{u})$ anche i quattro quasi alberi in figura 5.7, la cui combinazione convessa è proprio un circuito, come si può facilmente verificare. ∎

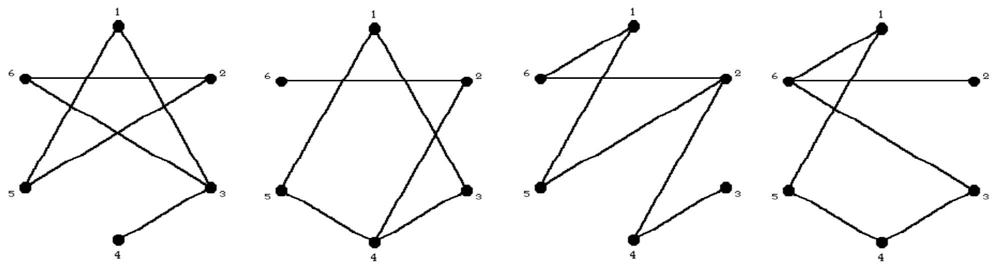


FIGURA 5.7

5.21 ESERCIZIO. Si risolva come nell'esempio precedente la seguente istanza di TSP (basta un'iterazione e vale la dualità forte):

$$\begin{bmatrix} 0 & 64 & 20 & 36 \\ 64 & 0 & 47 & 49 \\ 20 & 47 & 0 & 27 \\ 36 & 49 & 27 & 0 \end{bmatrix}$$

∎

5.22 ESERCIZIO. Come costruire un'istanza (non banale) del TSP per cui valga la dualità forte? ∎

5.23 ESERCIZIO. Come si modifica il problema duale del TSP se l'insieme X viene definito come l'insieme dei quasi alberi di supporto sul grafo dato? ∎

5.24 ESERCIZIO. Dato il grafo di figura 5.8, si determini tramite il problema duale il minimo albero di supporto con grado minore o uguale a 4 nel nodo centrale. Porre come unico vincolo esplicito il vincolo sul grado. ∎

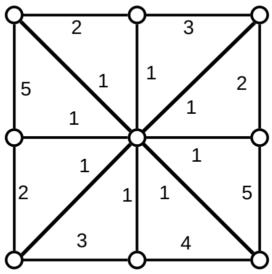


FIGURA 5.8

5.25 ESEMPIO. Consideriamo il seguente problema: dato un grafo G e m coppie di nodi, esistono m cammini disgiunti fra le coppie di nodi? Problemi di questo genere intervengono nella progettazione di circuiti VLSI, dove bisogna realizzare delle connessioni fra definite coppie di punti di un chip. Questo esempio è tratto da un'applicazione reale studiata da Feo e Hochbaum [1986].

Tecnologicamente le connessioni sul chip possono essere realizzate nel seguente modo: esistono due strati sovrapposti a breve distanza. Sul primo strato il canale conduttore può essere realizzato solo lungo un prefissato insieme di rette parallele ed equidistanti. Anche sul secondo strato i conduttori possono essere realizzati lungo un prefissato insieme di rette parallele, che però è ortogonale al primo. Inoltre si possono stabilire delle connessioni (i cosiddetti 'vias') fra un conduttore del primo strato con uno del secondo nei punti di sovrapposizione dei due insiemi di rette. Il grafo G è indotto da questa particolare struttura.

Si tratta pertanto di sapere se le connessioni richieste si possono effettivamente realizzare con la data tecnologia. Una risposta esatta a questo quesito richiede un notevole sforzo computazionale. Tuttavia in molti casi è possibile ottenere velocemente una risposta negativa facendo ricorso al problema duale. Formalizziamo il problema nel seguente modo: sia W l'insieme delle coppie di nodi da connettere, sia S l'insieme dei nodi origine di ogni coppia e sia T l'insieme ordinato dei nodi terminali di ogni coppia. Sia $\mathcal{P}_s$ l'insieme dei cammini da $s \in S$ all'elemento corrispondente della coppia e sia $\mathcal{P}_{si}$ il sottoinsieme dei cammini in $\mathcal{P}_s$ che passano per il nodo i. Per ogni cammino $P \in \mathcal{P}_k$ sia x_P una variabile che vale 1 se viene attivato il cammino P, altrimenti vale 0. Consideriamo allora il seguente problema primale:

$$\begin{aligned} v = \min \quad & -\sum_{s\in S}\sum_{P\in\mathcal{P}_s} x_P \\ & \sum_{P\in\mathcal{P}_s} x_P \le 1 && \forall s \in S \\ & \sum_{s\in S}\sum_{P\in\mathcal{P}_{si}} x_P \le 1 && \forall i \in N \\ & x_P \in \{0,1\} && \forall P \in \mathcal{P}_s \quad \forall s \in S \end{aligned}$$

Il primo gruppo di vincoli impone che venga attivato al più un cammino per ogni coppia. Il secondo gruppo di vincoli impone che i circuiti siano disgiunti. L'obiettivo quindi cerca di massimizzare il numero di circuiti disgiunti (è stato formalizzato come un minimo per omogeneità con la trattazione precedente). In realtà ciò che si vuole non è tanto tale massimo, quanto sapere se $v = -m$ (e in tal caso conoscere la soluzione ottima), e a questo quesito si può rispondere in modo negativo se avviene $d > -m$, in quanto si ricava subito dalla dualità debole $v \ge d > -m$. Quindi se si riesce a risolvere il problema duale in modo efficiente si ha a disposizione un modo rapido per decidere se certe proposte di connessione non sono realizzabili.

Costruiamo pertanto la funzione duale nel seguente modo:

$$L(u) = \min_{\substack{\sum_{P\in\mathcal{P}_s} x_P \le 1 \\ x_P\in\{0,1\}}} -\sum_{s\in S}\sum_{P\in\mathcal{P}_s} x_P + \sum_{i\in N} u_i \Big(\sum_{s\in S}\sum_{P\in\mathcal{P}_{si}} x_P - 1\Big) =$$

$$= -\sum_{i\in N} u_i + \min_{\substack{\sum_{P\in\mathcal{P}_s} x_P \le 1 \\ x_P\in\{0,1\}}} \Big(-\sum_{s\in S}\sum_{P\in\mathcal{P}_s} x_P + \sum_{i\in N} u_i \sum_{s\in S}\sum_{P\in\mathcal{P}_{si}} x_P \Big) =$$

$$= -\sum_{i\in N} u_i + \sum_{s\in S} \min_{\substack{\sum_{P\in\mathcal{P}_s} x_P \le 1 \\ x_P\in\{0,1\}}} \sum_{P\in\mathcal{P}_s} \Big(-1 + \sum_{i\in P} u_i \Big) x_P$$

Ora si noti che $\sum_{i\in P} u_i$ non è altro che la lunghezza del cammino P valutata sommando i costi u_i dei nodi appartenenti al cammino. Si indichi con $V_s(u)$ il valore minimo del cammino con origine in s. Se $V_s(u) < 1$ allora il minimo all'interno della sommatoria viene raggiunto ponendo $x_P := 1$ per il cammino $P \in \mathcal{P}_s$ di minimo costo e $x_P := 0$ per ogni altro cammino in $\mathcal{P}_s$ di costo peggiore. Se invece $V_s(u) > 1$ il minimo viene raggiunto ponendo $x_P := 0$ per ogni cammino in $\mathcal{P}_s$. Quindi la funzione duale diventa

$$L(u) = -\sum_{i\in N} u_i + \sum_{s\in S} \min\{\, -1 + V_s(u)\ ;\ 0\,\}$$

(si noti ad esempio che $L(0) = -|S|$). Come vedremo più avanti (esercizio 8.23) il calcolo di un cammino minimo con lunghezze di questo tipo si può effettuare con complessità computazionale $O(n \log n)$ dove n è il numero di nodi del grafo G.

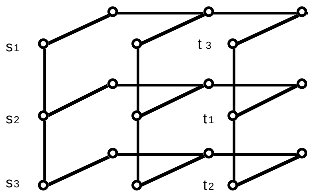


FIGURA 5.9(A)

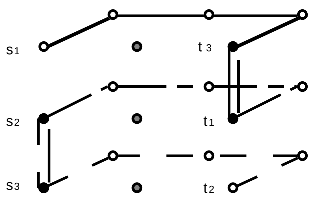


FIGURA 5.9(B)

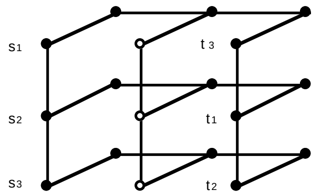


FIGURA 5.9(C)

Consideriamo ad esempio l'istanza in figura 5.9(a). Per $u := 0$ si ha $L(u) = -3$. Si prenda l'insieme di cammini in figura 5.9(b), che corrisponde ad un particolare $x \in X(0)$. Le componenti del relativo subgradiente $\gamma \in \partial L(0)$ valgono 0 sui nodi del grafo attraversati da un solo percorso (indicati in bianco), valgono 1 sui nodi attraversati da due percorsi (indicati in nero) e -1 su quelli non attraversati da alcun percorso (indicati in grigio).

Si noti che $V_s(\alpha\, u) = \alpha\, V_s(u)$ per ogni $\alpha \ge 0$. Quindi se indichiamo con $\alpha(u)$ il massimo valore per cui $V_s(\alpha\, u) \le 1$, $\forall s$, si ha

$$L(\alpha\, u) = -\alpha \sum_{i\in S} u_i + \alpha \sum_{s\in S} V_s(u) - |S| \qquad\qquad 0 \le \alpha \le \alpha(u)$$

e una direzione di ascesa h (dall'origine) si ottiene se $V_s(h) > \sum_{i\in N} h_i$. Ad esempio una tale direzione è quella illustrata in figura 5.9(c) dove le componenti di h valgono 1 sui nodi neri e 0 su quelli bianchi. Come si vede $V_{s_1}(h) = V_{s_2}(h) = V_{s_3}(h) = 6$, da cui $\sum_s V_s(h) = 18 > 15 = \sum_i h_i$. Pertanto $\alpha(h) = 1/6$ e per $0 < \alpha \le 1/6$ si ha $L(\alpha\, h) > -3$. Quindi è impossibile realizzare le connessioni richieste. ∎

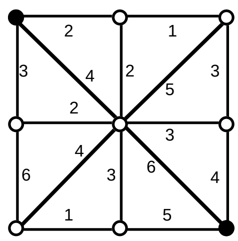


Figura 5.10

5.26 Esercizio. Sia dato il grafo in figura 5.10, dove i numeri rappresentano le lunghezze degli archi. Si voglia trovare il cammino più corto fra i due nodi neri con almeno k archi. Il problema di trovare il cammino più corto (senza ulteriori vincoli) con lunghezze non negative è polinomiale (vedi sezioni 8.6 e 9.3). Ammettendo lunghezze negative, ma non circuiti di lunghezza negativa, il problema rimane polinomiale. Per ora si trovino i cammini minimi in modo manuale, dichiarando illimitata un'istanza in cui sia presente un circuito di lunghezza negativa (un cammino minimo vi si può avvolgere infinite volte diminuendo continuamente la lunghezza). Si affronti il problema dato in modo duale introducendo come vincolo esplicito il vincolo sul numero di nodi. Si noti che per un valore u fissato il calcolo di $L(u)$ diventa il calcolo del cammino minimo con lunghezze diminuite della quantità costante u. Quali valori di u definiscono il dominio effettivo di $L(u)$? Si osservi che la funzione duale può essere ridefinita (ma non calcolata effettivamente) come

$$L(u) = \inf_{h \geq 1} \{V_h + (k - h)\, u\}$$

dove V_h è il valore del cammino minimo con esattamente h archi. Sotto quali ipotesi vale la dualità forte? Si risolva l'istanza data per vari valori di k. ∎

5.27 Esempio. Il *problema dell'assegnamento generalizzato* viene definito come:

$$\begin{aligned} v := \min \quad & \sum_{ij} c_{ij}\, x_{ij} & \\ & \sum_j x_{ij} = 1 & \forall i \\ & \sum_i a_{ij}\, x_{ij} \leq b_j & \forall j \\ & x_{ij} \in \{0, 1\} & \forall i \quad \forall j \end{aligned} \tag{5.8}$$

Questo problema modella una situazione in cui si devono assegnare delle attività A_i a delle macchine M_j, ciascuna delle quali è disponibile per un tempo totale b_j. L'attività A_i, se eseguita sulla macchina B_j, costa c_{ij} ed ha una durata a_{ij}. Si tratta di trovare un assegnamento ammissibile di costo minimo.

Questo problema può essere affrontato tramite il problema duale nei seguenti tre modi alternativi:

a) il vincolo esplicito sia costituito dal vincolo di capacità $\sum_i a_{ij}\, x_{ij} \leq b_j$;
b) il vincolo esplicito sia costituito dal vincolo d'assegnamento $\sum_j x_{ij} = 1$;
c) i vincoli di capacità e di assegnamento siano riformulati come

$$\begin{aligned} & \sum_j x_{ij} = 1 & \forall i \\ & \sum_i a_{ij}\, y_{ij} \leq b_j & \forall j \\ & x_{ij} - y_{ij} = 0 & \forall i \quad \forall j \end{aligned}$$

e il vincolo esplicito sia costituito dal solo vincolo $x_{ij} = y_{ij}$, $\forall i$, $\forall j$.

Definiamo allora:

$$L_a(u) := \min_{\substack{\Sigma_j x_{ij}=1 \\ x_{ij}\in\{0,1\}}} \sum_{ij} c_{ij}\, x_{ij} + \sum_j u_j \,\big(\sum_i a_{ij}\, x_{ij} - b_j\big)$$

$$L_b(v) := \min_{\substack{\Sigma_i a_{ij}\, x_{ij}\le b_j \\ x_{ij}\in\{0,1\}}} \sum_{ij} c_{ij}\, x_{ij} + \sum_i v_i \,\big(\sum_j x_{ij} - 1\big)$$

$$L_c(w) := \min_{\substack{\Sigma_i a_{ij}\, x_{ij}\le b_j \\ x_{ij}\in\{0,1\} \\ \Sigma_j y_{ij}=1 \\ x_{ij}\in\{0,1\}}} \sum_{ij} c_{ij}\, x_{ij} + \sum_{ij} w_{ij}\,(x_{ij} - y_{ij})$$

da cui

$$L_a(u) = -\sum_j u_j\, b_j + \sum_i \min_j \{c_{ij} + u_j\, a_{ij}\}$$

$$L_b(v) = -\sum_i v_i + \sum_j \min_{\substack{\Sigma_i a_{ij}\, x_{ij}\le b_j \\ x_{ij}\in\{0,1\}}} \sum_i (c_{ij} + v_i)\, x_{ij} = -\sum_i v_i + \sum_j K_j(c+v)$$

dove si è indicato con $K_j(c)$ il valore ottimo del problema di knapsack 0-1 con costi c, pesi a_j e limitazione destra b_j,

$$L_c(w) = \sum_j \min_{\substack{\Sigma_i a_{ij}\, x_{ij}\le b_j \\ x_{ij}\in\{0,1\}}} \sum_i (c_{ij} + w_{ij})\, x_{ij} - \sum_i \max_j\; w_{ij} = \sum_j K_j(c+w) - \sum_i \max_j\; w_{ij}$$

e, analogamente

$$d_a := \sup_{u\ge 0} L_a(u) \qquad\qquad d_b := \sup_v L_b(v) \qquad\qquad d_c := \sup_w L_c(w)$$

Cerchiamo ora di stabilire delle relazioni d'ordine per i tre ottimi duali. Introduciamo anche il problema rilassato del problema primale sostituendo in (5.8) il vincolo $x_{ij} \in \{0,1\}$ con il vincolo $0 \le x_{ij} \le 1$. Sia $\bar{v}$ l'ottimo di questo problema rilassato. Ovviamente $\bar{v} \le v$. Si noti che $L_a(u)$ è anche la funzione duale del problema primale rilassato (con la scelta (a) del vincolo esplicito), in virtù del fatto che nell'operazione di infimo che definisce L_a è irrilevante l'interezza di x_{ij}. Dimostreremo nel prossimo capitolo che il problema rilassato ammette dualità forte e quindi $d_a = \bar{v}$.

Sia ora $\bar{L}_b(v)$ il duale del problema rilassato con la scelta (b) del vincolo esplicito. In questo caso l'operazione di infimo prevede il calcolo di problemi di knapsack, per i quali invece fa differenza se le variabili sono intere o meno. Quindi in generale $\bar{L}_b(v) \le L_b(v)$, da cui $\bar{v} \le d_b$, essendo, per l'osservazione precedente, $\bar{v} = \sup_v \bar{L}_b(v)$.

Finora abbiamo quindi stabilito che $\bar{v} = d_a \le d_b \le v$. Per quel che riguarda $L_c(w)$ si definisca questa particolare scelta dei moltiplicatori w, $w_{ij}(v) := v_i$. Si ottiene

$$L_c(w(v)) = \sum_j K_j(c + w(v)) - \sum_i v_i = L_b(v)$$

e quindi, in generale, $d_c = \sup_w L_c(w) \ge \sup_v L_c(w(v)) = \sup_v L_c(v) = d_b$, e allora

$$\bar{v} = d_a \le d_b \le d_c \le v$$

Ad esempio sia data un'istanza con 5 attività e 3 macchine, costi e durate definiti da

$$c = \begin{pmatrix} 4 & 1 & 2 \\ 2 & 2 & 2 \\ 1 & 2 & 1 \\ 2 & 4 & 1 \\ 4 & 3 & 2 \end{pmatrix} \qquad\qquad a = \begin{pmatrix} 5 & 4 & 5 \\ 8 & 6 & 7 \\ 2 & 3 & 2 \\ 3 & 3 & 4 \\ 4 & 2 & 3 \end{pmatrix}$$

e disponibilità delle macchine $b = (10,\ 5,\ 6)$.

L'assegnamento ottimo è $\hat{x}_{13} = \hat{x}_{21} = \hat{x}_{33} = \hat{x}_{42} = \hat{x}_{52} = 1$ (e le altre variabili ovviamente nulle) con valore $\hat{z} = 12$. La soluzione ottima del problema rilassato è $\bar{x}_{12} = 1$, $\bar{x}_{21} = 5/6$, $\bar{x}_{22} = 1/6$, $\bar{x}_{31} = 1$, $\bar{x}_{41} = 1/4$, $\bar{x}_{43} = 5/4$, $\bar{x}_{53} = 1$ con valore ottimo $\bar{z} = 7.25$.

L'ottimo duale di (a) è $\hat{u} = (0,\ 0,\ 0.25)$. I valori $c_{ij} + \hat{u}_j\, a_{ij}$ sono

$$\begin{pmatrix} 4 & 1 & 3.25 \\ 2 & 2 & 5.75 \\ 1 & 2 & 1.5 \\ 2 & 4 & 2 \\ 4 & 3 & 2.75 \end{pmatrix}$$

quindi $X(\hat{u})$ è dato dai quattro valori

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

che danno luogo ai valori $g(X(\hat{u}))$

$$(\,3 \quad 1 \quad -3\,) \qquad (\,-5 \quad -5 \quad -3\,) \qquad (\,0 \quad 1 \quad 1\,) \qquad (\,8 \quad 5 \quad 1\,),$$

i quali, combinati convessamente con i coefficienti 1/12, 1/6, 3/4 e 0 danno luogo al subgradiente $(-7/12, 0, 0)$ che soddisfa le condizioni di ottimalità con il vincolo $u \geq 0$.

L'ottimo duale di (b) è $\hat{v} = (-7,\ -9,\ -1,\ -6,\ -7)$. I valori $c_{ij} + \hat{v}_i$ sono

$$c = \begin{pmatrix} -3 & -6 & -5 \\ -7 & -7 & -7 \\ 0 & 1 & 0 \\ -4 & -2 & -5 \\ -3 & -4 & -5 \end{pmatrix}$$

Con questi valori l'insieme $X(\hat{v})$ contiene 72 elementi (6 soluzioni per il primo knapsack, 2 per il secondo e 6 per il terzo). Uno di questi è proprio $\hat{x}$, per il quale si ha $g(\hat{x}) = 0$ e quindi $\hat{v}$ è ottimo. Si noti che $d_b = 12$ e la dualità forte è verificata. Ovviamente risulta $d_c = d_b$. ∎

Non tragga in inganno il fatto che in molti esempi si sia verificata la dualità forte pur essendo problemi NP-completi. Si tratta di una circostanza volutamente creata con dati di tipo particolare. Normalmente la verifica della dualità forte per problemi di questo genere deve essere considerata un caso raro.

5.4. Condizioni di ottimalità nel caso convesso

5.28 DEFINIZIONE. *Si definisce* problema perturbato *il seguente problema*

$$\begin{aligned} v(y) = \inf\, & f(x) \\ & g(x) \le y \\ & x \in X \end{aligned}$$

∎

Si è quindi modificato il vincolo esplicito. Per $y = 0$ si ha esattamente il problema dato, quindi $v(0) = v$. Si noti ancora che, indicando con $F(y)$ l'insieme ammissibile in corrispondenza di una certa y, si ha che $y' \ge y \Rightarrow F(y') \supset F(y)$ e quindi il valore ottimo del problema perturbato è una funzione, detta di *perturbazione*, $v : \mathbb{R}^m \to \mathbb{R}$ monotona decrescente in y (cioè $y' \ge y \Rightarrow v(y') \le v(y)$).

Costruiamo in $\mathbb{R}^{m+1}$ il seguente insieme:

$$Y = \{(y_0, y) \in \mathbb{R}^{m+1} : y_0 \ge f(x), y \ge g(x) \text{ per qualche } x \in X\}$$

o alternativamente

$$Y = \begin{pmatrix} f \\ g \end{pmatrix} (X) + \mathbb{R}^{m+1}_+$$

L'insieme Y si può anche caratterizzare nel seguente modo:

5.29 TEOREMA. $\overline{\operatorname{epi} v} = \overline{Y}$.

DIMOSTRAZIONE. Se $(y_0, y) \in \operatorname{epi} v$, $y_0 \ge v(y)$, cioè esistono x tali che $f(x)$ è arbitrariamente vicino a $v(y)$ e $g(x) \le y$, quindi $(v(y), y)$ appartiene alla chiusura di Y, ovvero $\operatorname{epi} v \subset \overline{Y}$, quindi $\overline{\operatorname{epi} v} \subset \overline{Y}$.

Viceversa se $(y_0, y) \in Y$ allora esiste x tale che $f(x) \le y_0$ e $g(x) \le y$ e quindi $v(y) \le y_0$ cioè $(y_0, y) \in \operatorname{epi} v$, ovvero $Y \subset \operatorname{epi} v$, da cui $\overline{Y} \subset \overline{\operatorname{epi} v}$. ∎

Normalmente è quasi sempre vero che Y e $\operatorname{epi} v$ sono chiusi e quindi coincidono. Tuttavia ciò non è vero in generale come dimostra il seguente esempio:

5.30 ESEMPIO.

$$\begin{aligned} v(y) = \inf\, & e^{x_1 x_2} \\ & x_2^2 \le y \qquad (g) \\ & x_1 \ge 0 \qquad (X) \end{aligned} \quad \Longrightarrow \quad v(y) = \begin{cases} +\infty & \text{se } y \le 0 \\ 1 & \text{se } y = 0 \\ 0 & \text{se } y > 0 \end{cases}$$

$$Y = \{(y_0, y) : y_0 > 0\,,\ y > 0\} \cup \{(y_0, y) : y_0 \ge 1\,,\ y = 0\}$$

e le inclusioni $Y \subset \operatorname{epi} v \subset \overline{Y}$ sono strette. ∎

L'equazione di un piano di supporto ad Y è:

$$u_0\, y_0 + u\, y = \inf_{(\eta_0, \eta) \in Y} u_0\, \eta_0 + u\, \eta$$

Però si noti che $u_0 \ge 0$ e $u \ge 0$, dato che Y è illimitato verso l'alto. Se $u_0 = 0$ il piano di supporto è detto verticale. Prendiamo in esame i piani di supporto non verticali per i quali si può porre $u_0 = 1$ senza perdità di generalità. Per essi l'equazione è del tipo:

$$y_0 + u\, y = \inf_{(\eta_0, \eta) \in Y} \eta_0 + u\, \eta = \inf_{x \in X} f(x) + u\, g(x) = L(u)$$

I fatti rilevanti sono:

a) $L(u)$ è il funzionale di supporto di Y e quindi anche di epi v;
b) Il piano di supporto di equazione $y_0 + uy = L(u)$ interseca l'asse coordinato $(1, 0)$ nel punto $y_0 = L(u)$;
c) Il semispazio di supporto $y_0 + u\,y \geq L(u)$ è l'epigrafo della funzione affine $L(u) - u\,y$.

Da queste osservazioni si vede che la risoluzione del problema duale corrisponde a trovare quel piano di supporto per cui l'intersezione con l'asse coordinato $(1, 0)$ è massima. Il piano di supporto deve essere non verticale, altrimenti $u \to +\infty$ e quindi non esiste l'ottimo del duale.

L'osservazione c) è cruciale. Infatti ogni funzione convessa, propria e chiusa si può caratterizzare come il supremo puntuale di quelle funzioni affini i cui epigrafi sono i semispazi di supporto non verticali del suo epigrafo. Quindi possiamo immediatamente stabilire:

5.31 Teorema. *Se $v(y)$ è chiusa, convessa e propria, allora:*

$$v(y) = \sup_{u \geq 0} L(u) - u\,y$$

e in particolare

$$v = v(0) = \sup_{u \geq 0} L(u) = d$$

∎

Quindi lo scarto di dualità è nullo se la funzione di perturbazione risulta chiusa, convessa e propria. Il fatto che lo scarto di dualità sia nullo non implica tuttavia la dualità forte. Infatti anche se il primale ammette una soluzione ottima potrebbe darsi che il valore supremo di $L(u)$ venga ottenuto al tendere di u ad infinito, cioè con piani di supporto sempre più verticali. Siamo perciò interessati da un lato a vedere sotto quali condizioni v risulta chiusa, convessa e propria e dall'altro quali altre ipotesi siano necessarie affinché valga la dualità forte.

5.32 Teorema. *Se $X \subset \mathbb{R}^n$ è convesso e f e g sono funzioni convesse su X allora $v(y)$ è una funzione convessa.*

Dimostrazione. Siano $y, y' \in \operatorname{dom} v$. Per ogni $\varepsilon > 0$ esistono $x_\varepsilon, x'_\varepsilon \in X$ tali che $f(x_\varepsilon) - \varepsilon \leq v(y)$, $g(x_\varepsilon) \leq y$ e $f(x'_\varepsilon) - \varepsilon \leq v(y')$, $g(x'_\varepsilon) \leq y'$. Quindi

$$\begin{aligned} \alpha\, v(y) + (1-\alpha)\, v(y') &\geq \alpha\,(f(x_\varepsilon) - \varepsilon) + (1-\alpha)\,(f(x'_\varepsilon) - \varepsilon) = \\ &\alpha\, f(x_\varepsilon) + (1-\alpha)\, f(x'_\varepsilon) - \varepsilon \geq f(\alpha\, x_\varepsilon + (1-\alpha)\, x'_\varepsilon) - \varepsilon \end{aligned}$$

Il punto $\alpha\, x_\varepsilon + (1-\alpha)\, x'_\varepsilon \in X$ perché X è convesso; inoltre dalla convessità di g e da $\alpha\, g(x_\varepsilon) \leq \alpha\, y$, $\quad (1-\alpha)\, g(x'_\varepsilon) \leq (1-\alpha)\, y'$ discende:

$$g(\alpha\, x_\varepsilon + (1-\alpha)\, x'_\varepsilon) \leq \alpha\, g(x_\varepsilon) + (1-\alpha)\, g(x'_\varepsilon) \leq \alpha\, y + (1-\alpha)\, y'$$

Quindi $\alpha\, x_\varepsilon + (1-\alpha)\, x'_\varepsilon$ è ammissibile per il problema perturbato con perturbazione $\alpha\, y + (1-\alpha)\, y'$. E allora deve valere $f(\alpha\, x_\varepsilon + (1-\alpha)\, x'_\varepsilon) \geq v(\alpha\, y + (1-\alpha)\, y')$, e quindi si è ottenuto complessivamente

$$\alpha\, v(y) + (1-\alpha)\, v(y') \geq v(\alpha\, y + (1-\alpha)\, y') - \varepsilon$$

Siccome la diseguaglianza è valida per $\forall \varepsilon > 0$ deve anche valere

$$\alpha\, v(y) + (1-\alpha)\, v(y') \geq v(\alpha\, y + (1-\alpha)\, y')$$

cioè la convessità di $v(y)$. ∎

Si noti che la condizione data dal teorema è soltanto sufficiente ma non necessaria. Come controesempio si riconsideri l'esempio 5.30 dove v è convessa anche se la funzione obiettivo $\exp(-x_1 x_2)$ non lo è.

Per ciò che riguarda la chiusura di v, si può far rilevare che l'immagine continua di un insieme compatto è compatta, e quindi Y è chiuso se X è compatto. Come si vede dalla dimostrazione del teorema 5.29 la chiusura di Y implica la chiusura di $\operatorname{epi} v$ e quindi si è dimostrato il seguente teorema:

5.33 TEOREMA. *Se $X \subset \mathbb{R}^n$ è compatto e convesso, f e g sono convesse e continue su X, allora lo scarto di dualità è nullo.* ∎

L'ipotesi di compattezza di X non è così restrittiva come sembra, in quanto ciò che interessa è il comportamento locale e in molti casi il problema perturbato può essere ristretto in un intorno di $y = 0$ senza perdere le sue caratteristiche locali.

Quindi le condizioni del teorema 5.33 sono sufficienti ad annullare lo scarto di dualità. Sono tuttavia anche sufficienti a garantire la dualità forte? Cioè a garantire che anche il duale possieda una soluzione ottima? (infatti il primale la possiede in base alle ipotesi). La risposta è negativa come si può vedere dal seguente esempio:

5.34 ESEMPIO.

$$\begin{aligned} v(y) \;=\; & \inf x \\ & \quad x^2 \le y \qquad (g) \\ -1 \le & \quad x \le 1 \qquad (X) \end{aligned} \qquad \Longrightarrow \qquad v(y) = \begin{cases} +\infty & \text{se } y < 0 \\ 0 & \text{se } y = 0 \\ -\sqrt{y} & \text{se } 0 \le y \le 1 \\ -1 & \text{se } y \ge 1 \end{cases}$$

Le ipotesi del teorema 5.33 sono verificate. Consideriamo il duale:

$$L(u) = \inf_{-1 \le x \le 1} x + u\,x^2 = \begin{cases} u - 1 & \text{se } 0 \le u \le \frac{1}{2} \\ -\frac{1}{4u} & \text{se } u > \frac{1}{2} \end{cases}$$

quindi $\sup L(u) = 0$, ma non esiste l'ottimo duale in quanto u tende ad infinito. Il problema sorge dal fatto che il piano di supporto all'epigrafo di v nel punto $(v(0), 0)$ è verticale e quindi non corrisponde ad una funzione affine. ∎

Per garantire la dualità forte bisogna quindi introdurre una condizione che impedisca al piano di supporto di essere verticale. A tal fine consideriamo la seguente definizione:

5.35 DEFINIZIONE. *Sia v convessa. Il problema primale si dice* stabile *se esiste x tale che $f(x) = v(0) < +\infty$ ed esiste una costante K tale che $v(0) - v(y) \le K\,||y||$, $\forall y$.* ∎

La definizione corrisponde pertanto ad una limitazione superiore del rapporto incrementale di v. Si badi però che $v(0) - v(y)$ compare senza valore assoluto e quindi la limitazione è automaticamente soddisfatta se $y \le 0$ dato che v è monotona. Possiamo ora enunciare il seguente fondamentale teorema:

5.36 TEOREMA. *Sia $X \subset \mathbb{R}^n$ convesso e siano f e g funzioni convesse su X. Allora vale la dualità forte se e solo se il problema è stabile.*

DIMOSTRAZIONE. La necessità si dimostra facilmente. Infatti vale sempre, indicando con $\hat{u}$ l'ottimo duale,

$$v(y) = \sup_{u \ge 0} L(u) - u\,y \ge L(\hat{u}) - \hat{u}\,y = v(0) - \hat{u}\,y$$

da cui $v(0) - v(y) \leq \hat{u}\, y \leq ||\hat{u}||\, ||y||$ e quindi il problema è stabile con costante data dalla norma dell'ottimo duale.

Per la sufficienza si costruiscano in $\mathbb{R}^{m+1}$ i due insiemi (vedi figura 5.11):

$$A = \operatorname{epi} v - \begin{pmatrix} v(0) \\ 0 \end{pmatrix}$$
$$B = \{(y_0, y) : y_0 < -K\, ||y||\}$$

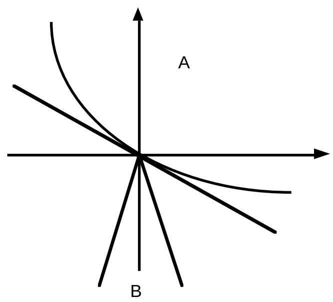


FIGURA 5.11

È immediato verificare che A e B sono convessi, non vuoti e disgiunti (per l'ipotesi di stabilità). Quindi esiste un piano separatore (w_0, w), tale che:

$$w_0\, y_0 + w\, y \geq w_0\, y_0' + w\, y' \qquad \forall (y_0, y) \in A \quad \forall (y_0'.y') \in B$$

Siccome $0 \in \overline{B}$ si ha $w_0\, y_0 + w\, y \geq 0$, $\forall (y_0, y) \in A$ e siccome $0 \in A$ si ha $w_0\, y_0' + w\, y' \leq 0$, $\forall (y_0', y') \in B$. Siccome $A = A + \mathbb{R}_+^{m+1}$, $w_0 \geq 0$, $w \geq 0$. Tuttavia non può avvenire che $w_0 = 0$. Infatti si avrebbe $w\, y' \leq 0$, $\forall (y_0', y') \in B$ e cioè $w\, y' \leq 0$, $\forall y' \in \mathbb{R}^m$ e quindi $w = 0$. Ma siccome $(w_0, w) \neq 0$ dobbiamo avere $w_0 > 0$. Poniamo allora $u = w/w_0$ e quindi abbiamo

$$y_0 + u\, y \geq 0 \qquad \forall (y_0, y) \in A$$

cioè

$$f(x) + u\, g(x) - v(0) \geq 0 \qquad \forall x \in X$$

e quindi

$$\inf_{x \in X} f(x) + u\, g(x) \geq v(0)$$

da cui

$$L(u) \geq v(0)$$

e siccome vale sempre $v(0) \geq L(u)$, abbiamo $v(0) = L(u)$. ∎

Il fatto notevole enunciato dal teorema è che la stabilità è equivalente all'esistenza dell'ottimo duale in presenza di convessità. Una verifica diretta della stabilità è tuttavia difficile in generale, ragion per cui sono state elaborate diverse condizioni sufficienti di stabilità. Queste condizioni vanno sotto il generico nome di *qualificazioni dei vincoli*. Riportiamo qui di seguito solo una di queste qualificazioni in quanto è l'unica che si presta ad una rapida verifica.

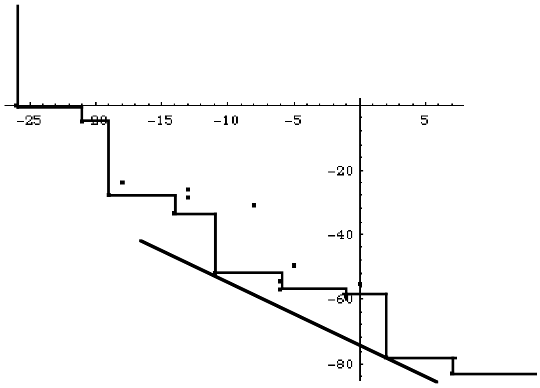


FIGURA 5.12

5.37 TEOREMA. *Sia $X \subset \mathbb{R}^n$ convesso e siano f e g funzioni convesse. Se esiste $\bar{x} \in X$ tale che $g(\bar{x}) < 0$ allora il problema è stabile.*

DIMOSTRAZIONE. L'ipotesi $g(\bar{x}) < 0$ implica $0 \in \overset{\circ}{\text{dom}}\, v$. Quindi v essendo convessa è senz'altro continua nel punto 0 e il piano di supporto a $\text{epi}\, v$ nel punto $(v(0), 0)$ non può essere verticale. ∎

Se le condizioni di convessità non sono rispettate, $\text{epi}\, v$ non sarà in generale convesso. Siccome il problema duale corrisponde alla caratterizzazione esterna di $\text{epi}\, v$ tramite piani di supporto, vediamo che risolvere il duale corrisponde a risolvere una versione convessificata del primale, cioè un problema in cui la funzione di perturbazione ha epigrafo $\overline{\text{conv}}(\text{epi}\, v)$. Un caso particolare interessante si ha quando f e g sono funzioni affini. Infatti in questo caso si ha:

$$\begin{pmatrix} f \\ g \end{pmatrix} \big(\sum \alpha_k\, x_k \big) = \sum \alpha_k \begin{pmatrix} f \\ g \end{pmatrix} (x_k)$$

e quindi $\overline{\text{conv}}\, \text{epi}\, v = \overline{\text{conv}}\, Y = \begin{pmatrix} f \\ g \end{pmatrix} (\overline{\text{conv}}\, X)$.

Quindi se f e g sono funzioni affini la risoluzione del duale corrisponde alla convessificazione del vincolo implicito.

5.38 ESEMPIO. Si riconsideri l'esempio 5.15. In figura 5.12 sono riportati l'immagine di X tramite f e g e l'insieme Y. Le soluzioni ammissibili sono quelle la cui immagine si trova a sinistra dell'asse y_0. Come si vede Y non è convesso e l'immagine dell'ottimo si trova "dentro" l'inviluppo convesso di Y. Pertanto nessun piano di supporto potrà passarvi. In figura è anche riportato il piano di supporto corrispondente all'ottimo duale. L'intersezione di questo piano con l'asse y_0 è il valore ottimo duale; si noti lo scarto di dualità. ∎

5.39 ESERCIZIO. Si debba minimizzare $f(x)$ con $f(x) := \max_{i \in I} \{g_i(x)\}$, I insieme finito e $g_i(x)$ convessa per ogni i. Osservando che minimizzare $f(x)$ è equivalente a risolvere

$$\begin{aligned} \min \quad & v \\ & v \geq g_i(x) \qquad \forall i \end{aligned} \tag{5.9}$$

si calcoli il duale di (5.9) e si confronti il risultato con il teorema 4.119. ∎

5.5. Dualità e differenziabilità nel caso convesso

Uno dei significati più importanti della variabile duale ottima è dato dal fatto che essa misura la variazione del valore ottimo al variare del vincolo.

Ci sono diverse ragioni pratiche per cui questa variazione è importante. In molti casi ad esempio il vincolo esplicito rappresenta una limitazione sulle risorse disponibili in un processo di massimizzazione del profitto e quindi la variazione del profitto ottimo rispetto ad uno scostamento della disponibilità di risorse rivela il valore intrinseco della risorsa rispetto all'obiettivo. In altri casi il vincolo esplicito può rappresentare il bisogno di soddisfare la domanda del mercato e l'obiettivo è la minimizzazione dei costi di produzione. Anche in questo caso la variazione del costo ottimo rispetto ad una variazione della richiesta del mercato rappresenta il costo intrinseco di produzione del bene. In questi casi in cui l'obiettivo è di natura monetaria, la variazione si presenta come un prezzo intrinseco della risorsa relativamente alla sua potenzialità di utilizzo. Dimostreremo fra poco che, in presenza di convessità, le variabili duali ottime sono uguali a tale variazione e per tale motivo le variabili duali ottime prendono anche il nome di *prezzi ombra* in un contesto economico.

Vogliamo ora dimostrare anche che le variabili duali ottime, sotto opportune ipotesi e in un modello ovviamente molto semplificato, rappresentano non solo il prezzo di un bene determinato in base al costo intrinseco di produzione, ma anche il prezzo determinato dal processo di libera competizione del mercato. In altre parole il prezzo determinato dalla libera concorrenza si 'aggiusta' sui costi intrinseci di produzione.

Per far vedere quest'aspetto della dualità si consideri un'azienda che produce dei beni. Il livello di attività produttiva sia rappresentato da una variabile $x \in \mathbb{R}^n$. I costi di produzione siano definiti da una funzione $f(x)$, che supponiamo convessa. Questa è un'ipotesi abbastanza forte, in quanto in realtà i costi hanno un andamento dapprima concavo, poi lineare e infine convesso (cioè inizialmente è efficiente aumentare di scala, ma oltre una certa soglia i costi tendono ad aumentare molto). L'attività produttiva si realizza fornendo al mercato m beni. Le quantità richieste dal mercato per ogni bene sono b_i. La quantità del bene i che si riesce a produrre con un livello produttivo x è rappresentata da una funzione $g_i(x)$, che supponiamo concava (questa è un'ipotesi più realistica della precedente). Il livello di attività deve inoltre sottostare a dei vincoli che possiamo rappresentare come $x \in X$. Quindi il problema che deve affrontare il responsabile dell'azienda è quello di minimizzare i costi soddisfacendo contemporaneamente alla domanda esterna di beni, cioè deve risolvere

$$
\begin{aligned}
\inf\ & f(x) \\
& g(x) \geq b \\
& x \in X
\end{aligned}
$$

Tuttavia, per il responsabile, il problema non si esaurisce qui. Bisogna ancora stabilire il prezzo dei beni da lanciare sul mercato. Si indichino con u_i i vari prezzi. Siccome il mercato richiede le quantità b_i (e non di più) si potrebbe pensare di massimizzare una funzione del tipo $\sum_i u_i\, b_i$. Però questo non è corretto. Tale massimizzazione, oltre ad essere illimitata (e quindi inutilizzabile), non è corretta dal punto di vista economico in quanto i prezzi devono anche essere competitivi.

Per modellare l'aspetto competitivo del problema si consideri il mercato come antagonista dell'azienda. In altri termini ogni profitto dell'azienda è una perdita per il mercato e viceversa. Se il mercato produce per proprio conto gli stessi beni dell'azienda realizza un profitto, al prezzo u, pari a $\sum_i u_i\, g_i(x)$ (valore dei beni) $-$ $f(x)$ (costi di produzione). In questo caso x rappresenta il livello di attività del mercato e stiamo naturalmente assumendo che lo stesso modello produttivo dell'azienda sia applicabile anche al mercato (cioè le funzioni

f e g siano le medesime). Diamo inoltre per scontato che il mercato sia efficiente, per cui il profitto realizzato dal mercato è in realtà

$$V(u) = \sup_{x \in X} \sum_i u_i\, g_i(x) - f(x) = - \inf_{x \in X} f(x) - \sum_i u_i\, g_i(x)$$

Questo profitto realizzato dal mercato è per l'azienda una perdita. Quindi l'azienda, nel fissare i prezzi, deve tenere conto di questa perdita e quindi deve risolvere

$$\max_{u \geq 0} \sum_i u_i\, b_i - V(u)$$

che è esattamente il problema duale. Può essere sorprendente notare che, in base alla teoria, se c'è un surplus di produzione di un certo bene, il prezzo ottimo debba essere nullo. In effetti se un vincolo non è attivo, questo significa che i costi di produzione non sono toccati da una eventuale variazione (piccola) della domanda. Il prezzo intrinseco, cioè quanto costa produrre quel particolare bene, è effettivamente nullo, e se l'azienda proponesse al mercato un prezzo positivo, il mercato reagirebbe immediatamente producendo per proprio conto tutto il bene richiesto ad un prezzo inferiore, cosa effettivamente possibile.

Si noti ancora che $V(\hat{u}) = \hat{u}\, b - f(\hat{x})$ in base alle CGO con $\hat{x}$ e $\hat{u}$ ottimi e quindi $V(\hat{u})$ rappresenta il profitto sia del mercato che dell'azienda ad indicare che si è raggiunto un equilibrio fra domanda e offerta. Il profitto $V(\hat{u})$ è positivo se il costo marginale dei beni da produrre (rappresentato da $\hat{u}$) supera il costo globale.

Da un altro punto di vista la validazione di un modello matematico di ottimizzazione è anche affidata alla sua sensibilità parametrica. Le variabili duali misurano appunto questa sensibilità rispetto ai vincoli. Valori molto elevati delle variabili duali vanno attentamente studiati e bisogna verificare se si tratta di un modello inadeguato o se invece è la natura stessa del problema reale a presentare un'intrinseca instabilità.

Questa discussione informale si precisa meglio nei risultati che verranno ora esposti e che legano la funzione di perturbazione alle variabili duali.

5.40 Teorema. *Sia X convesso e siano f e g funzioni convesse su X. Allora la dualità forte vale se e solo se $-\hat{u}$, con $\hat{u}$ ottimo duale, è un subgradiente di $v(y)$ per $y = 0$ e il primale ammette soluzione ottima.*

Dimostrazione. Per la necessità si veda la dimostrazione del teorema 5.36, da cui si era trovato $v(y) - v(0) \geq (-\hat{u})\, y$, per qualsiasi ottimo duale $\hat{u}$ e qualsiasi y.

Per la sufficienza, se $\gamma \in \partial v(0)$ si ha per definizione $v(y) - \gamma y \geq v(0)$, $\forall y$. Se $(y_0, y) \in \operatorname{epi} v$, vale quindi anche $y_0 - \gamma y \geq v(0)$, $\forall (y_0, y) \in \operatorname{epi} v$ ed anche $y_0 - \gamma y \geq v(0)$, $\forall (y_0, y) \in Y$, cioè

$$L(-\gamma) = \inf_{x \in X} f(x) - \gamma\, g(x) \geq v(0)$$

Si noti ora che $\gamma \leq 0$ dalla monotonia di $v(y)$. Quindi $-\gamma$ è una variabile duale e deve valere $L(-\gamma) \leq v(0)$ per la dualità debole. Quindi $L(-\gamma) = v(0)$, cioè vale la dualità forte e $-\gamma$ è ottimo duale. ∎

In altre parole, se indichiamo con $\hat{U}$ l'insieme dei duali ottimi abbiamo che l'eguaglianza

$$-\hat{U} = \partial v(0)$$

vale se e solo se il problema è stabile oppure vale la dualità forte. La relazione che vale comunque nel caso convesso è $\partial v(0) \subset -\hat{U}$ (si riveda l'esempio 5.30 dove $\partial v(0) = \emptyset$ e $\hat{U} = [0, +\infty)$). Ricordando le proprietà della derivata direzionale possiamo pertanto concludere:

5.41 Corollario. *Sia X convesso e siano f e g funzioni convesse. Allora se vale la dualità forte*

$$Dv(0,y) = \sup_{u \in \hat{U}} -u\,y$$

∎

Ovviamente se l'ottimo duale è unico, la dualità forte (ovvero la stabilità) implica che $v(y)$ è differenziabile in 0 e che $Dv(0) = -\hat{u}$ cioè l'ottimo duale è l'opposto del gradiente di $v(y)$ nell'origine. Si noti ancora che si può scrivere

$$Dv(0,y) = -\inf_{u \in \hat{U}} u\,y = -\varphi(y, \hat{U})$$

cioè, se vale la dualità forte, la derivata direzionale della funzione di perturbazione corrisponde all'opposto della funzione di supporto dell'insieme delle variabili duali ottime.

Queste considerazioni ci permettono di 'completare' il teorema 5.11, ovvero di rispondere anche alla seguente domanda: dato un problema primale e i problemi duali ottenuti ripartendo in modi diversi i vincoli in impliciti ed espliciti, possiamo dire che questi problemi duali sono 'equivalenti' fra loro nel senso che la variabile duale ottima relativa ad un certo vincolo esplicito è invariante rispetto alle diverse formulazioni del problema duale? Nel caso convesso, purché si assuma la stabilità, la risposta è affermativa. In generale invece non si può affermare che la dualità forte rimanga verificata se si 'aggiungono' vincoli espliciti, perché la stabilità potrebbe non essere verificata per i vincoli aggiunti.

5.42 Teorema. *Sia $X \subset \mathbb{R}^n$ convesso. Sia f convessa e siano g^1 e g^2 funzioni convesse con $g^1 : \mathbb{R}^n \to \mathbb{R}^{m_1}$, $g^2 : \mathbb{R}^n \to \mathbb{R}^{m_2}$. Siano*

$$L(u^1, u^2) := \inf_{x \in X} f(x) + u^1\,g^1(x) + u^2\,g^2(x) \qquad \tilde{L}(u^1) := \inf_{\substack{x \in X \\ g^2(x) \le 0}} f(x) + u^1\,g^1(x)$$

Sia $(\hat{x}, \hat{u}^1)$ punto di sella per $\tilde{L}(u^1)$. Inoltre valga la dualità forte anche per $L(u^1, u^2)$. Allora esiste $\hat{u}^2$ tale che $(\hat{x}, \hat{u}^1, \hat{u}^2)$ è punto di sella per $L(u^1, u^2)$.

Dimostrazione. Si indichi con $\Pi : \mathbb{R}^{m_1+m_2} \to \mathbb{R}^{m_1}$ la proiezione canonica. Sia $U \subset \mathbb{R}^{m_1+m_2}$ l'insieme degli ottimi di $L(u^1, u^2)$ e $\tilde{U} \subset \mathbb{R}^{m_1}$ l'insieme degli ottimi di $\tilde{L}(u^1)$. Sia $v(y^1, y^2)$ la funzione di perturbazione per il problema con vincolo esplicito g^1, g^2 e $\tilde{v}(y^1)$ con vincolo esplicito g^1. Naturalmente $v(y^1, 0) = \tilde{v}(y^1)$ per ogni y^1. Quindi anche $Dv_0(h, 0) = D\tilde{v}_0(h)$ per ogni h. Allora

$$Dv_0(h,0) = \sup_{(\gamma^1,\gamma^2) \in \partial v_0} \gamma^1\,h + \gamma^2\,0 = \sup_{\gamma \in \Pi\,\partial v_0} \gamma\,h = -\inf \varphi(h, \Pi\,U)$$

$$D\tilde{v}_0(h) = \sup_{\gamma \in \partial \tilde{v}_0} \gamma\,h = -\inf \varphi(h, \tilde{U})$$

Essendo i subdifferenziali chiusi e convessi anche $\tilde{U}$ e $\Pi\,U$ sono chiusi e convessi e si può allora applicare il teorema 4.104 da cui $\tilde{U} = \Pi\,U$, che implica immediatamente la tesi. ∎

Nel caso convesso le CGO possono essere riformulate in termini differenziali. Sia $D(x)$ il cono delle direzioni ammissibili in X a partire da x. In base al teorema 4.121 la prima delle CGO è equivalente a:

$$(\partial f(\hat{x}) + \hat{u}\,\partial g(\hat{x})) \cap -D^*(\hat{x}) \neq \emptyset \tag{5.10}$$

Un caso particolare di interesse si ha quando $X = \mathbb{R}^n_+$, per il quale:

$$-D^*(\hat{x}) = \{w \in \mathbb{R}^n : w_j \geq 0 \text{ se } \hat{x}_j = 0;\ w_j = 0 \text{ se } \hat{x}_j > 0\}$$

e quindi, nel caso f e g siano differenziabili, (5.10) diventa:

$$\begin{array}{ll} Df_j(\hat{x}) + \hat{u}\, Dg_j(\hat{x}) \geq 0 & \text{se} \quad \hat{x}_j = 0 \\ Df_j(\hat{x}) + \hat{u}\, Dg_j(\hat{x}) = 0 & \text{se} \quad \hat{x}_j > 0 \end{array}$$

Indicando sinteticamente $D_x L(\hat{x}, \hat{u}) := Df(\hat{x}) + \hat{u}\, Dg(\hat{x})$ e $D_u L(\hat{x}, \hat{u}) := g(\hat{x})$, le CGO si traducono nelle seguenti condizioni.

5.43 Teorema. (Condizioni di Karush-Kuhn-Tucker) *Un punto* $(\hat{x}, \hat{u})$ *è ottimo se soddisfa le seguenti condizioni*

$$\begin{array}{llll} 1) & D_x L(\hat{x}, \hat{u}) \geq 0 & 1') & D_u L(\hat{x}, \hat{u}) \leq 0 \\ 2) & D_x L(\hat{x}, \hat{u}) \cdot \hat{x} = 0 & 2') & \hat{u} \cdot D_u L(\hat{x}, \hat{u}) = 0 \\ 3) & \hat{x} \geq 0 & 3') & \hat{u} \geq 0 \end{array} \tag{5.11}$$

Se il problema è stabile le condizioni (5.11) sono anche necessarie. ∎

5.44 Esempio. Sia da risolvere:

$$\begin{aligned} \min \quad & c\, x + \frac{1}{2} x^T Q\, x \\ & A\, x \geq b \\ & x \geq 0 \end{aligned}$$

dove Q è una matrice simmetrica positiva semidefinita. Allora

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

e le condizioni di Karush-Kuhn-Tucker diventano

$$\begin{array}{llll} 1) & c + x^T Q - u\, A \geq 0 & 1') & b - A\, x \leq 0 \\ 2) & (c + x^T Q - u\, A) \cdot x = 0 & 2') & u \cdot (b - A\, x) = 0 \\ 3) & x \geq 0 & 3') & u \geq 0 \end{array}$$

ovvero

$$\begin{gathered} \begin{pmatrix} Q & -A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ u^T \end{pmatrix} - \begin{pmatrix} -c^T \\ b \end{pmatrix} = \begin{pmatrix} r^T \\ s \end{pmatrix} \\ (x,\ u^T,\ r^T,\ s) \geq 0 \\ x^T\, r^T + u\, s = 0 \end{gathered}$$

Si tratta di un problema di complementarità lineare (vedi capitolo seguente) che nel caso di matrice Q positiva semidefinita (altrimenti la funzione obiettivo non è convessa) può essere risolto polinomialmente (vedi anche il capitolo 17). ∎

5.6. Dualità e differenziabilità nel caso non convesso

Si è visto che, se il problema da minimizzare non è convesso, non necessariamente la funzione Lagrangiana fornisce informazione utile per il calcolo dei minimi. Tuttavia se si tratta di un problema non discreto con funzioni differenziabili la funzione Lagrangiana può ancora essere usata con vantaggio anche se si ottengono soltanto risultati di tipo locale. A titolo di esempio si consideri il seguente problema:

$$\begin{aligned} \min \quad & -e^{-x_1^2-x_2^2} - x_1 - x_2 \\ & x_1 + x_2 \leq 2 \end{aligned}$$

L'insieme $Y = (f, g)(\mathbb{R}^2) + \mathbb{R}^2_+$ è rappresentato in figura 5.13. Come si vede Y non è convesso e c'è scarto di dualità per cui sembrerebbe inutile applicare la dualità. Però se consideriamo il valore $u = \hat{u}$ per cui la retta $y_0 + u\,y = f((1,1))$ risulta tangente al punto di minimo $(f(1,1), g(1,1))$ (indicata in figura) si vede che il punto $\hat{x} = (1,1)$, oltre ad essere il minimo, è stazionario per la Lagrangiana $L(x, \hat{u}) = f(x) + \hat{u}\,g(x)$, cioè $D_x L(\hat{x}, \hat{u}) = 0$. Quindi applicando le condizioni (5.11) di Karush-Kuhn-Tucker (non considerando in questo caso il vincolo $x \geq 0$) otteniamo ugualmente l'ottimo $\hat{x}$.

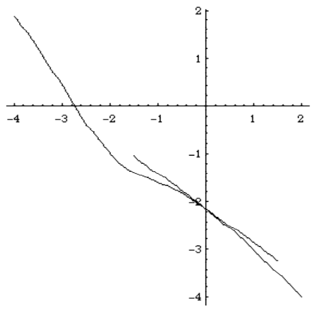


FIGURA 5.13

Per derivare formalmente delle condizioni del tipo (5.11) dobbiamo seguire dei procedimenti diversi da quelli finora impiegati. Consideriamo dapprima il caso di vincoli d'eguaglianza. Sia dato il problema

$$\begin{aligned} \min \quad & f(x) \\ & g(x) = 0 \end{aligned}$$

dove $f : \mathbb{R}^n \to \mathbb{R}$ e $g : \mathbb{R}^n \to \mathbb{R}^m$ sono due volte differenziabili. Ci proponiamo di ricavare delle condizioni necessarie affinché un punto $\hat{x}$ sia minimo. Sia $x(t)$ una curva due volte differenziabile in $\mathbb{R}^n$, parametrizzata in modo che $x(0) = \hat{x}$ e $g(x(t)) = 0$ per ogni t. Chiameremo tale curva ammissibile. I vettori tangenti a $x(t)$ nel punto $x(\bar{t})$ sono per definizione i vettori $\alpha dx(t)/dt|_{t=\bar{t}}$ con $\alpha > 0$. Useremo la notazione $\dot{x}(\bar{t}) := dx(t)/dt|_{t=\bar{t}}$. Il fatto che $g(x(t))$ sia identicamente nulla implica che anche la derivata rispetto a t è identicamente nulla e quindi

$$\left.\frac{dg(x(t))}{dt}\right|_{t=0} = Dg(\hat{x})\,\dot{x}(0) = 0$$

cioè la direzione $\dot{x}(0)$, tangente alla curva $x(t)$ nel punto $\hat{x}$, appartiene a $\mathcal{N}(Dg(\hat{x}))$. Siccome caratterizzeremo le tangenti a $x(t)$ tramite $\mathcal{N}(Dg(\hat{x}))$ vogliamo essere sicuri che effettivamente ogni elemento di $\mathcal{N}(Dg(\hat{x}))$ sia a sua volta tangente a $x(t)$. In generale ciò non è vero.

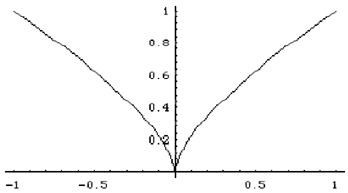


FIGURA 5.14

Si consideri ad esempio $g(x) = x_1^2 - x_2^3$. I punti che soddisfano $g(x) = 0$ sono rappresentati in figura 5.14.

Solo le direzioni $(0,1)$ sono tangenti alla curva nell'origine mentre $\mathcal{N}(Dg(0,0)) = \mathbb{R}^2$. Un criterio che garantisca la tangenza a $x(t)$ in $\hat{x}$ per ogni elemento di $\mathcal{N}(Dg(\hat{x}))$ è dato dalla seguente definizione e dal successivo teorema:

5.45 DEFINIZIONE. *Un punto x tale che* $\operatorname{rank} Dg(x) = m$ *si dice* regolare. ∎

5.46 TEOREMA. *Sia $\hat{x}$ regolare. Allora $h \in \mathcal{N}(Dg(\hat{x}))$ se e solo se esiste una curva ammissibile $x(t)$ tale che $h = \dot{x}(0)$ (cioè h è tangente alla curva in $\hat{x}$).* ∎

Se $\hat{x}$ è minimo locale la derivata di f rispetto a t deve annullarsi per $t = 0$, cioè $df(x(t)/dt|_{t=0} = Df(x(0))\,\dot{x}(0) = 0$, e quindi $Df(\hat{x})\,h = 0$ per ogni h tangente a $x(t)$ e, se $\hat{x}$ è regolare, per ogni $h \in \mathcal{N}(Dg(\hat{x}))$. Questa condizione si può esprimere anche dicendo che $Df(\hat{x})$ deve essere ortogonale a $\mathcal{N}(Dg(\hat{x}))$, ovvero $Df(\hat{x})$ deve appartenere a $\mathcal{R}(Dg^T(\hat{x}))$. In altre parole $Df(\hat{x})$ si deve poter esprimere come combinazione lineare dei vari gradienti $Dg_i(\hat{x})$. Abbiamo pertanto dimostrato che:

5.47 TEOREMA. (Condizione necessaria di primo ordine) *Sia $\hat{x}$ regolare. Allora se $\hat{x}$ è minimo locale esiste $\hat{u}$ tale che $Df(\hat{x}) + \hat{u}\,Dg(\hat{x}) = 0$.* ∎

Fissato un valore di u il Lagrangiano è stazionario nel punto $x(u)$ se il piano $y_0 + u\,y = L(x(u), u)$ è tangente all'insieme $(f,g)(\mathbb{R}^n)$ (anziché di supporto come nel caso convesso). Trattandosi di minimi locali la situazione tuttavia è un po' più complessa che non nel caso globale. Si consideri ad esempio il problema

$$\begin{aligned} \min \quad & 1/3\,x_1^3 - x_1 - x_2 \\ & 1/3\,x_1^3 - x_1 + x_2 = 1 \end{aligned}$$

Non è difficile vedere che $(f,g)(\mathbb{R}^2) = \mathbb{R}^2$. Tuttavia

$$Df(x) + u\,Dg(x) = (x_1^2 - 1\,,\,-1) + u\,(x_1^2 - 1\,,\,1) = 0$$

porta a $\hat{u} = 1$ e $x_1^2 = 1$ da cui $\hat{x} = (1, 5/3)$ oppure $\hat{x} = (-1, 1/3)$. Che significato dare a questi valori che sono interni a $(f,g)(\mathbb{R}^2) \cap \{0, \mathbb{R}\}$? Per comprendere la situazione si supponga di mappare il piano $\mathbb{R}^2$ in $\mathbb{R}^3$ anziché in $\mathbb{R}^2$ tramite la mappa $(f(x), g(x), x_1)$. In figura 5.15 viene rappresentata l'immagine del rettangolo $\{x : -2.2 \leq x_1 \leq 2.2,\ -1 \leq x_2 \leq 2\}$ tramite la mappa.

I punti di stazionarietà corrispondono alle 'pieghe' della superficie in $\mathbb{R}^3$. Ovviamente $(f,g)(\mathbb{R}^2)$ si ottiene semplicemente proiettando tale superficie sul piano delle prime due coordinate e si vede come i punti di stazionarietà risultino interni. Del resto una situazione simile accade dovendo trovare i minimi della funzione $\mathbb{R} \to \mathbb{R}$, $x\,(x^2 - 1)$.

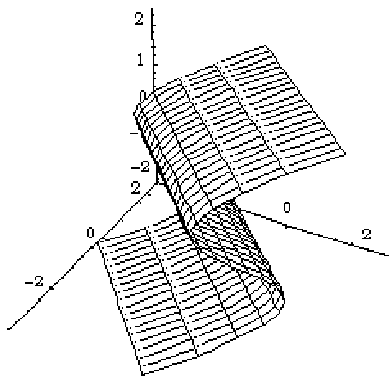


Figura 5.15

Come si vede da quest'ultimo esempio l'applicazione delle condizioni di stazionarietà del Lagrangiano può fornire diversi punti e non è detto che tutti questi punti siano minimi, dato che la condizione di primo ordine del teorema 5.47 è solo necessaria. Dei due punti dell'esempio il punto $(1, 5/3)$ è minimo locale mentre $(-1, 1/3)$ è massimo locale come si vede dalla figura 5.15. In generale, per poter decidere se un punto che soddisfa la condizione di primo ordine è minimo bisogna ricorrere all'informazione fornita dalle derivate seconde.

5.48 Teorema. (Condizione necessaria di secondo ordine) *Sia $\hat{x}$ regolare. Allora se $\hat{x}$ è minimo locale $D_{xx}L(\hat{x}, \hat{u})$ è positivo semidefinito su $\mathcal{N}(Dg(\hat{x}))$ per ogni $\hat{u}$ tale che $Df(\hat{x}) + \hat{u}\, Dg(\hat{x}) = 0$.*

Dimostrazione. Siccome $g(x(t))$ è identicamente nulla per ogni t anche la derivata seconda rispetto a t di $g(x(t))$ è nulla per ogni t e in particolare per $t = 0$. Quindi

$$\frac{d^2\, g_i(x(0))}{dt^2} = \sum_{j,k} \frac{\partial^2 g_i(\hat{x})}{\partial x_j\, \partial x_k}\, \dot{x}_j(0)\, \dot{x}_k(0) + \sum_j \frac{\partial g_i(\hat{x})}{\partial x_j}\, \ddot{x}_j(0) = 0 \qquad \forall i \tag{5.12}$$

e, moltiplicando ogni equazione per $\hat{u}_i$ e sommando, si ottiene

$$\sum_{j,k} \sum_i \hat{u}_i\, \frac{\partial^2 g_i(\hat{x})}{\partial x_j\, \partial x_k}\, \dot{x}_j(0)\, \dot{x}_k(0) + \sum_j \sum_i \hat{u}_i\, \frac{\partial g_i(\hat{x})}{\partial x_j}\, \ddot{x}_j(0) = 0$$

cioè con notazione compatta

$$\dot{x}(0)^T\, D_{xx}(\hat{u}\, g(\hat{x}))\, \dot{x}(0) + D(\hat{u}\, g(\hat{x}))\, \ddot{x}(0) = 0$$

Siccome $\hat{x}$ è minimo, $d^2 f(x(t))/dt^2|_{t=0} \geq 0$ e quindi

$$\frac{d^2\, f(\hat{x})}{dt^2} = \sum_{j,k} \frac{\partial^2 f(\hat{x})}{\partial x_j\, \partial x_k}\, \dot{x}_j(0)\, \dot{x}_k(0) + \sum_j \frac{\partial f(\hat{x})}{\partial x_j}\, \ddot{x}_j(0) = \dot{x}(0)^T\, D_{xx} f(\hat{x})\, \dot{x}(0) + Df(\hat{x})\, \ddot{x}(0) \geq 0$$

e quindi, combinando assieme le due relazioni e sfruttando la condizione di primo ordine

$$\dot{x}(t)^T\, D_{xx}(f(x) + \hat{u}\, g(x))\, \dot{x}(t) \geq 0$$

ovvero $D_{xx}L(\hat{x}, \hat{u})$ è positivo semidefinito su $\mathcal{N}(Dg(\hat{x}))$. ∎

Per capire il senso delle espressioni appena scritte, si consideri dapprima (5.12). Si tratta di una relazione che lega i vettori $\dot{x}(t)$ e $\ddot{x}(t)$ a $g(x)$, quando $x(t)$ soddisfa il vincolo $g(x) = 0$. L'espressione (5.12) rivela che una quantità calcolata sul sottospazio ortogonale a $Dg(x)$, cioè $\dot{x}(0)^T D_{xx} g(\hat{x}) \dot{x}(0)$ è uguale all'opposto di un'altra quantità calcolata su $Dg(x)$, cioè $Dg(\hat{x})\,\ddot{x}(0)$. Queste due quantità misurano ambedue la curvatura e sono nulle se e solo se g è affine. La valutazione della derivata seconda di f fa anch'essa uso di termini in $\dot{x}$ e $\ddot{x}$. Però i termini in $\ddot{x}$ possono essere eliminati grazie alla relazione precedente per $g(x)$ e alla condizione di primo ordine. Infatti solo se la curva di $g(x)$ e le curve di livello di $f(x)$ sono tangenti la sostituzione può essere effettuata. Il risultato finale si esprime unicamente sul sottospazio ortogonale a $Dg(x)$ dato che, a meno di infinitesimi di ordine superiore, f e g si comportano allo stesso modo su $Dg(x)$.

La condizione del teorema 5.48 può essere leggermente rafforzata e diventare così sufficiente:

5.49 Teorema. (Condizione sufficiente di secondo ordine) *Un punto $\hat{x}$ è minimo locale stretto se esiste $\hat{u}$ tale che $Df(\hat{x}) + \hat{u}\,Dg(\hat{x}) = 0$ e $D_{xx}L(\hat{x}, \hat{u})$ è positivo definito su $\mathcal{N}(Dg(\hat{x}))$.*

Dimostrazione. Si supponga che $\hat{x}$ non sia minimo locale stretto e quindi esista una successione $h^k \to 0$ tale che $f(\hat{x} + h^k) \leq f(\hat{x})$ e $g(\hat{x} + h^k) = 0$ per ogni k. La successione $h^k/||h^k||$ è limitata e ammette una sottosuccessione convergente ad una direzione $\bar{h}$ con $||\bar{h}|| = 1$. Senza perdita di generalità possiamo supporre che la successione stessa tenda a $\bar{h}$. Allora si ha

$$g_i(\hat{x} + h^k) - g_i(\hat{x}) = Dg_i(\hat{x})\,h^k + o(||h^k||^2)$$

I due termini di sinistra sono nulli dato che i punti sono ammissibili. Dividendo per $||h^k||$ e passando al limite si ottiene $Dg_i(\hat{x})\,\bar{h} = 0$ cioè $\bar{h} \in \mathcal{N}(Dg(\hat{x}))$. Inoltre si ha

$$f(\hat{x} + h^k) - f(\hat{x}) = Df(\hat{x})\,h^k + \frac{1}{2} D^2 f(\hat{x})(h^k, h^k) + o(||h^k||^3) \leq 0$$
$$g_i(\hat{x} + h^k) - g_i(\hat{x}) = Dg_i(\hat{x})\,h^k + \frac{1}{2} D^2 g_i(\hat{x})(h^k, h^k) + o(||h^k||^3) = 0$$

Da cui

$$\left(Df(\hat{x}) + \sum_i \hat{u}_i\, Dg_i(\hat{x}) \right) h^k + \frac{1}{2} \left(D^2 f(\hat{x}) + \sum_i \hat{u}_i D^2 g_i(\hat{x}) \right) (h^k, h^k) + o(||h^k||^3) =$$
$$\frac{1}{2}(D^2 f(\hat{x}) + \sum_i \hat{u}_i D^2 g_i(\hat{x}))(h^k, h^k) + o(||h^k||^3 \leq 0$$

Dividendo per $||h^k||$ e passando al limite si ottiene

$$\left(D^2 f(\hat{x}) + \sum_i \hat{u}_i D^2 g_i(\hat{x}) \right) (\bar{h}, \bar{h}) \leq 0$$

che contraddice l'ipotesi di positiva definitezza. ∎

Anche se i teoremi 5.48 e 5.49 non caratterizzano completamente gli ottimi locali vincolati tramite condizioni differenziali (possono cioè esistere punti che soddisfano le condizioni necessarie ma non quelle sufficienti, e quindi abbiamo incertezza in merito alla loro ottimalità), pur tuttavia nella maggior parte dei casi è sufficiente limitarsi alle condizioni di secondo ordine.

Affrontiamo ora il caso di vincoli di diseguaglianza. Sia dato quindi

$$\begin{array}{ll} \min & f(x) \\ & g(x) \leq 0 \end{array}$$

Siccome stiamo considerando minimi locali, un vincolo per cui si abbia $g_i(\hat{x}) < 0$ è come se non esistesse per il punto $\hat{x}$, e pertanto può essere trascurato nella caratterizzazione dell'ottimalità di $\hat{x}$. Supponiamo quindi senza perdita di generalità che tutti i vincoli siano attivi nel punto $\hat{x}$ che stiamo considerando, cioè $g(\hat{x}) = 0$.

Anche in questo caso usiamo archi di curva ammissibili, ne valutiamo le tangenti e caratterizziamo queste tangenti tramite lo Jacobiano dei vincoli. Sia $x(t)$ una curva due volte differenziabile in $\mathbb{R}^n$ parametrizzata in modo che $x(0) = \hat{x}$, $g(x(0)) = 0$, $g(x(t)) \leq 0$ per ogni t. Chiameremo tale curva *ammissibile*. Si ha

$$g(x(t)) = g(x(0)) + Dg(\hat{x})\,\dot{x}(0)\,t + o(t^2) \leq 0$$

da cui $Dg(\hat{x})\,\dot{x}(0) \leq 0$. In generale un vettore h che soddisfa $Dg(\hat{x})\,h \leq 0$ non è necessariamente tangente a qualche arco ammissibile. Si consideri ad esempio $g_1(x_1, x_2) := -x_1^3 - x_2$ e $g_2(x_1, x_2) := -x_1^3 + x_2$ e $\hat{x} = (0, 0)$. Si può dimostrare che

5.50 TEOREMA. *Sia $\hat{x}$ regolare (nel senso della definizione 5.45). Allora $Dg(\hat{x})\,h \leq 0$ se e solo esiste una curva ammissibile $x(t)$ tale che $h = \dot{x}(0)$.* ∎

Analogamente al caso con vincoli di uguaglianza, con l'unica differenza riguardante la non negatività delle variabili duali, abbiamo i seguenti teoremi.

5.51 TEOREMA. (Condizione necessaria di primo ordine) *Sia $\hat{x}$ regolare. Allora se $\hat{x}$ è minimo locale esiste $\hat{u} \geq 0$ tale che $Df(\hat{x}) + \hat{u}\,Dg(\hat{x}) = 0$.*

DIMOSTRAZIONE. Sia $\hat{x}$ minimo. Consideriamo una curva ammissibile $x(t)$. Per t sufficientemente piccolo si ha $0 \leq f(x(t)) - f(\hat{x}) = Df(\hat{x})\,\dot{x}(0)\,t + o(t^2)$ da cui $Df(\hat{x})\,\dot{x}(0) \geq 0$. Quindi se h è tale che $Dg(\hat{x})\,h \leq 0$, per la regolarità $h = \dot{x}(0)$ per qualche curva ammissible, e quindi $Df(\hat{x})\,h \geq 0$. Dal lemma 4.74 si ha che esistono coefficienti $\hat{u}_i \geq 0$ tali che $-Df(\hat{x}) = \sum_i \hat{u}_i\,Dg_i(\hat{x})$. ∎

Una dimostrazione alternativa del precedente teorema può esser fatta considerando il problema

$$\begin{array}{ll} \min & Df(\hat{x})\,h \\ & Dg(\hat{x})\,h \leq 0 \end{array}$$

che risulta essere convesso. Per poter affermare l'esistenza delle variabili duali bisogna però far vedere che il problema è stabile.

5.52 TEOREMA. (Condizione necessaria di secondo ordine) *Sia $\hat{x}$ regolare. Allora se $\hat{x}$ è minimo locale $D_{xx}L(\hat{x}, \hat{u})$ è positivo semidefinito su $\mathcal{N}(Dg(\hat{x}))$ per ogni $\hat{u} \geq 0$ tale che $Df(\hat{x}) + \hat{u}\,Dg(\hat{x}) = 0$.*

DIMOSTRAZIONE. Ovviamente $\hat{x}$ è minimo anche rispetto al vincolo $g(x) = 0$ e quindi si applicano i risultati del teorema 5.48. ∎

5.53 Teorema. (Condizione sufficiente di secondo ordine) *Un punto $\hat{x}$ è minimo locale stretto se esiste $\hat{u} > 0$ tale che $Df(\hat{x}) + \hat{u}\, Dg(\hat{x}) = 0$ e $D_{xx}L(\hat{x}, \hat{u})$ è positivo definito su $\mathcal{N}(Dg(\hat{x}))$.*

Dimostrazione. Come nel teorema 5.49 supponiamo l'esistenza di una successione h^k tale che $f(\hat{x} + h^k) \le f(\hat{x})$ e $g(x + h^k) \le 0$. Sia $\bar{h} = \lim_{k\to\infty} h^k/||h^k||$. Si ottiene $Df(\hat{x})\,\bar{h} \le 0$ e $Dg(\hat{x})\,\bar{h} \le 0$. Vogliamo far vedere che le ipotesi implicano la più forte condizione $\bar{h} \in \mathcal{N}(Dg(\hat{x}))$. Sia $\bar{h}_0$ la proiezione di $\bar{h}$ su $\mathcal{N}(Dg(\hat{x}))$ e $\bar{h}_1$ la proiezione di $\bar{h}$ su $\mathcal{N}(Dg(\hat{x}))^\perp = \mathcal{R}(Dg(\hat{x})^T)$. Quindi esiste $r \ne 0$ tale $\bar{h}_1 = Dg(\hat{x})^T\, r$. Allora

$$Dg(\hat{x})\,\bar{h} = Dg(\hat{x})\,(\bar{h}_0 + \bar{h}_1) = Dg(\hat{x})\,\bar{h}_1 = Dg(\hat{x})\,Dg(\hat{x})^T\, r \le 0 \tag{5.13}$$

Siccome $\hat{x}$ è regolare $Dg(\hat{x})\,Dg(\hat{x})^T$ è non singolare e quindi $Dg(\hat{x})\,Dg(\hat{x})^T\, r \ne 0$. Da questo, da (5.13) e da $\hat{u} > 0$ si ricava $\hat{u}\, Dg(\hat{x})\, Dg(\hat{x})^T\, r < 0$. Per ipotesi $Df(\hat{x}) = -\hat{u}\, Dg(\hat{x})$ e quindi

$$Df(\hat{x})\,\bar{h} = -\hat{u}\, Dg(\hat{x})\,\bar{h} = -\hat{u}\, Dg(\hat{x})\,\bar{h}_1 = -\hat{u}\, Dg(\hat{x})\, Dg(\hat{x})^T\, r \le 0$$

che contraddice il precedente risultato. Quindi $\bar{h} \in \mathcal{N}(Dg(\hat{x}))$. Il resto della dimostrazione procede come nel teorema 5.49. ∎

Il tipo d'informazione che possono dare le condizioni enunciate può essere schematizzato come in figura 5.16. I tre insiemi indicati in figura come cerchi rappresentano l'insieme dei punti che soddisfano le condizioni sufficienti, i minimi e quelli che soddisfano le condizioni necessarie. Si tenga presente che è richiesta la regolarità per le condizioni necessarie e quindi possono esistere minimi che soddisfano le condizioni sufficienti senza soddisfare quelle necessarie. I casi in cui le condizioni danno un'informazione conclusiva sono quelli tratteggiati, quando cioè le condizioni sufficienti sono verificate e quindi siamo certi dell'ottimalità oppure quando il punto è regolare e le condizioni necessarie non sono soddisfatte e quindi siamo certi che il punto non è ottimo.

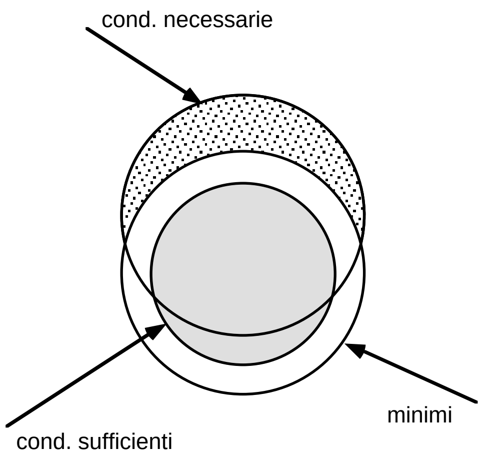


Figura 5.16

5.54 Esempio. Sia dato il seguente problema da risolvere a seconda del valore del parametro a (per comodità di notazione usiamo le variabili x e y invece di x_1 e x_2 e le variabili duali u e v invece di u_1 e u_2).

$$\min \quad a\,x^2 + y$$

$$x^2 + (y-1)^2 \geq 1 \tag{1}$$
$$x^2 + (y-2)^2 \leq 4 \tag{2}$$

$$L(x,y,u,v) = a\,x^2 + y + u\,(1 - x^2 - (y-1)^2) + v\,(x^2 + (y-2)^2 - 4);$$

$$DL_x = 2ax - 2ux + 2vx \qquad DL_y = 1 - 2u(y-1) + 2v(y-2);$$

$$D^2L = 2\begin{pmatrix} a-u+v & 0 \\ 0 & -u+v \end{pmatrix}$$

Le condizioni al primo ordine danno

$$x(a-u+v) = 0 \qquad 1 - 2u(y-1) + 2v(y-2) = 0$$

Consideriamo i vari casi di vincoli attivi:

— (1) e (2) non attivi $\Longrightarrow u = v = 0 \Longrightarrow DL_y = 1$;

— (1) non attivo e (2) attivo: quindi $u = 0$ e $DL_y = 0 \Longrightarrow y = 2 - 1/(2v)$. Imponiamo $DL_x = 0$. Se $x = 0$ allora $(y-2)^2 = 4$ (imponendo (2) attivo) e quindi o $y = 0$, $v = 1/4$ oppure $y = 4$, $v = -1/4$. Il primo caso è da escludere perché in realtà (1) risulta attivo, e il secondo caso è da escludere perché $v < 0$.

Se invece imponiamo $a - u + v = 0$ (sempre da $DL_x = 0$) abbiamo $v = -a \geq 0$ e quindi $\hat{y} = 2 + 1/(2a)$. Imponendo (2) attivo si ottiene $\hat{x}^2 = (16a^2 - 1)/(4a^2)$ da cui $a \leq -1/4$ oppure $a \geq 1/4$. Siccome $a \leq 0$ (da $v \geq 0$) $a \leq -1/4$. Imponendo l'ammissibilità di (1) si ha

$$\frac{16a^2 - 1}{4a^2} + \left(\frac{1+4a}{2a} - 1\right)^2 \geq 1 \Longrightarrow a(4a+1) \geq 0$$

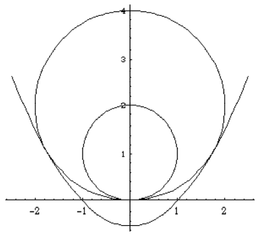
che è soddisfatta per $a \leq -1/4$. Lo Jacobiano dei vincoli attivi vale (considerando solo la soluzione positiva $\hat{x} = \sqrt{16a^2 - 1}/|2a|$)

$$\left.\begin{pmatrix} 2x & 2(y-2) \end{pmatrix}\right|_{(\hat{x},\hat{y})} = \begin{pmatrix} \sqrt{16a^2-1}/|a| & 1/a \end{pmatrix}$$

e quindi $\mathcal{N}(Dg) = (1, \sqrt{16a^2 - 1})$ e D^2L su $\mathcal{N}(Dg)$ vale $v(16a^2 - 1)$ che è positivo per $a < -1/4$. Quindi i punti

$$\left(\pm\frac{\sqrt{16a^2-1}}{2|a|}, 2 + \frac{1}{2a}\right)$$

sono minimi locali se $a < -1/4$ (nella figura i due minimi locali, di fatto anche globali, nel caso $a = -1/2$, sono i punti d'intersezione della parabola con la circonferenza).

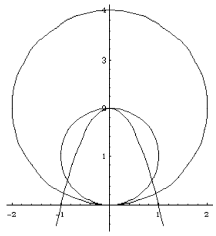


$a = -1/2$

— (1) attivo e (2) non attivo. Quindi $v = 0$ e $DL_y = 0 \Longrightarrow y = 1 + 1/(2u)$. Se $x = 0$ (da $DL_x = 0$) allora $(y-1)^2 = 1$ (imponendo (1) attivo) e quindi o $y = 0$, $u = -1/2$ oppure $y = 2$, $u = 1/2$. Il primo caso è da escludere perché $u < 0$. Consideriamo il secondo caso per il quale (2) risulta effettivamente non attivo. Lo Jacobiano dei vincoli attivi vale:

$$\begin{pmatrix} -2x & -2(y-1) \end{pmatrix}\Big|_{(0,2)} = \begin{pmatrix} 0 & -2 \end{pmatrix}$$

e quindi $\mathcal{N}(Dg) = (1,0)$. Le condizioni necessarie sono applicabili perché $(0,2)$ è regolare. La condizione sull'Hessiano della Lagrangiana dà $a - u + v > 0$ cioè $1/2 < a$. Quindi per $a > 1/2$ sono soddisfatte le condizioni sufficienti di secondo ordine e il punto $(0,2)$ è minimo locale (in figura si vede il minimo locale, questa volta non globale, nel caso $a = 2$).



$a = 2$

Se invece imponiamo $a - u + v = 0$ (sempre da $DL_x = 0$) abbiamo $u = a \geq 0$ e quindi $\hat{y} = 1 + 1/(2a)$. Imponendo (1) attivo si ottiene $\hat{x}^2 = (4a^2-1)/(4a^2)$ da cui $a \leq -1/2$ oppure $a \geq 1/2$. Siccome $a \geq 0$ (da $u \geq 0$) si ha $a \geq 1/2$. Imponendo l'ammissibilità di (2) si ha

$$\frac{4a^2-1}{4a^2} + \left(\frac{1+2a}{2a} - 2\right)^2 \leq 4 \Longrightarrow a(2a+1) \geq 0$$

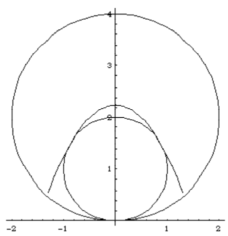
che è soddisfatta per $a \geq 1/2$. Lo Jacobiano dei vincoli attivi vale (considerando solo la soluzione positiva $\hat{x} = \sqrt{4a^2-1}/|2a|$)

$$\begin{pmatrix} -2x & -2(y-1) \end{pmatrix}\Big|_{(\hat{x},\hat{y})} = \begin{pmatrix} -\sqrt{4a^2-1}/|a| & -1/a \end{pmatrix}$$

e quindi $\mathcal{N}(Dg) = (-1, \sqrt{16a^2-1})$ e D^2L su $\mathcal{N}(Dg)$ vale $-u(4a^2-1)$ che è negativo per $a > 1/2$. Quindi i punti

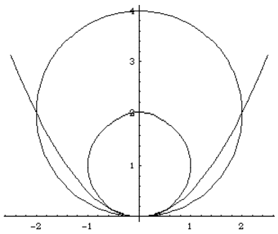
$$\left(\pm\frac{\sqrt{4a^2-1}}{2|a|}, 1 + \frac{1}{2a}\right)$$

non sono minimi locali se $a > 1/2$ in quanto le condizioni necessarie di ottimalità non sono verificate. Se $a = 1/2$ non si può dire nulla in base alle condizioni di secondo ordine (tali punti non sono ottimi comunque).

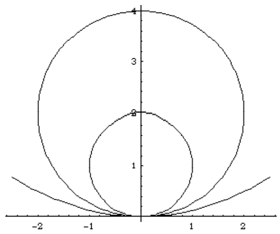


$a = 1$

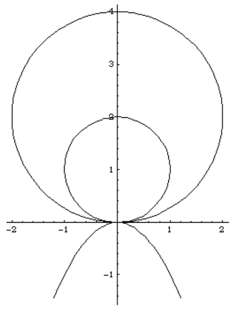
— (1) e (2) attivi. Quindi $x = y = 0$. Le condizioni al primo ordine danno $1+2u-4v = 0$. Le condizioni necessarie non sono applicabili perché il punto non è regolare. Le condizioni sufficienti del secondo ordine portano a $(a - u + v) > 0$. Siccome $u > 0$ e $v > 0$, da $1+2u-4v = 0$ si ricava $v > 1/4$. Da $(a-u+v) > 0$ e $1+2u-4v = 0$ si ricava $2a+1 > 2v$ quindi $a > -1/4$. Si noti che le variabili duali non sono univocamente determinate. Basta che sia $v = 1/4 + 1/2\, u$.



$a = -1/2$



$a = -1/8$



$a = 1$

Riassumendo, gli ottimi locali sono:

$a < -1/4$	$\Longrightarrow$	$\left(\pm\dfrac{\sqrt{16a^2-1}}{2\lvert a\rvert}, 2+\dfrac{1}{2a}\right)$
$a = -1/4$	$\Longrightarrow$	$(0,0)$ non decidibile con le condizioni di I e II ordine
$-1/4 < a < 1/2$	$\Longrightarrow$	$(0,0)$
$a = 1/2$	$\Longrightarrow$	$(0,0)$, $(0,2)$ non decidibile
$a > 1/2$	$\Longrightarrow$	$(0,0)$, $(0,2)$

∎