

# METODI PER LA SOLUZIONE DI PROBLEMI DI PROGRAMMAZIONE NONLINEARE

Gianni Di Pillo

Dipartimento di Informatica e Sistemistica  
Università di Roma “La Sapienza”  
Via Eudossiana 18 - 00184 Roma

## 1 Definizione del problema e nozioni preliminari

In queste note vengono esposti, nei loro fondamenti, alcuni metodi per la soluzione di problemi di *Programmazione Nonlineare* della forma

$$\begin{array}{ll} \min_x f(x) & f : \mathbb{R}^n \rightarrow \mathbb{R} \\ h(x) = 0 & h : \mathbb{R}^n \rightarrow \mathbb{R}^p \quad p \leq n \\ g(x) \leq 0 & g : \mathbb{R}^n \rightarrow \mathbb{R}^m \end{array} \quad (P_0)$$

Nel problema  $(P_0)$  sono presenti vincoli di uguaglianza e disuguaglianza; spesso tuttavia per semplicità di trattazione o per specificità dei metodi proposti, si considerano problemi con soli vincoli di uguaglianza

$$\begin{array}{ll} \min_x f(x) & \\ h(x) = 0 & \end{array} \quad (P_1)$$

o con soli vincoli di disuguaglianza

$$\begin{array}{ll} \min_x f(x) & \\ g(x) \leq 0 & \end{array} \quad (P_2)$$

Da un punto di vista formale è sempre possibile trasformare il vincolo di uguaglianza  $h_i(x) = 0$  nella coppia di vincoli di disuguaglianza  $\{h_i(x) \leq 0, h_i(x) \geq 0\}$ ; in tal modo il problema  $(P_1)$  può essere riscritto nella forma

$$\begin{array}{ll} \min_x f(x) & \\ h(x) \leq 0 & \\ -h(x) \leq 0 & \end{array} \quad (\tilde{P}_1)$$



$$I_a(\tilde{x}) \triangleq \{i \in \{1, 2, \dots, m\} : g_i(\tilde{x}) = 0\},$$

e con  $g_a(\tilde{x})$  il subvettore di  $g(\tilde{x})$  corrispondente ai vincoli di disuguaglianza attivi

$$g_a(\tilde{x}) \triangleq \{g_i(\tilde{x}) : i \in I_a(\tilde{x})\}.$$

Se in  $\tilde{x}$  i gradienti dei vincoli attivi, e cioè i vettori  $\nabla h_i(\tilde{x})$ ,  $i = 1, 2, \dots, p$ ,  $\nabla g_i(\tilde{x})$ ,  $i \in I_a(\tilde{x})$ , risultano linearmente indipendenti,  $\tilde{x}$  è un *punto di regolarità* per i vincoli.<sup>1</sup>

Considerato un punto  $x^*$  di minimo locale vincolato, le condizioni necessarie e quelle sufficienti soddisfatte in  $x^*$  si semplificano notevolmente nel caso in cui  $x^*$  sia di regolarità.

Con riferimento al problema  $(P_0)$ , si definisce *funzione lagrangiana (generalizzata)* la funzione

$$L(x, \mu, \lambda) \triangleq f(x) + \mu' h(x) + \lambda' g(x), \quad \mu \in \mathbb{R}^p, \quad \lambda \in \mathbb{R}^m;$$

i vettori  $\mu, \lambda$  vengono detti *moltiplicatori*.

Ciò premesso, le condizioni di ottimalità più frequentemente utilizzate in pratica, e cui si farà qui riferimento, sono:

**Teorema 1.1** (*Condizioni necessarie del primo ordine*): se  $x^*$  è una soluzione locale del problema  $(P_0)$ , e se in  $x^*$  è soddisfatta la condizione di regolarità dei vincoli, esistono e sono unici i moltiplicatori  $\mu^*, \lambda^*$  che insieme a  $x^*$ , soddisfano le condizioni:

$$\nabla_x L(x^*, \mu^*, \lambda^*) = 0 \tag{1}$$

$$h(x^*) = 0 \tag{2}$$

$$g(x^*) \leq 0 \tag{3}$$

$$\lambda^* \geq 0 \tag{4}$$

$$\lambda^{*'} g(x^*) = 0 \tag{5}$$

□

Le condizioni (1)-(5) sono dette di Kuhn-Tucker.

Sia  $x^*$  una soluzione locale del problema  $(P_0)$  e  $\lambda^*$  il moltiplicatore dei vincoli di disuguaglianza che soddisfa le condizioni di Kuhn-Tucker; si dice che in  $x^*$  vale la condizione di *stretta complementarità* se  $\forall i \in I_a(x^*)$  risulta  $\lambda^* > 0$ .

<sup>1</sup>Per definizioni più generali di regolarità dei vincoli in un punto  $\tilde{x}$  vedi [1].

**Teorema 1.2** (Condizioni sufficienti del secondo ordine): se,  $x^*, \mu^*, \lambda^*$  soddisfano le condizioni (1) - (5); se  $x^*, \lambda^*$  soddisfano la condizione di stretta complementarità; se infine risulta

$$y' \nabla_{xx}^2 L(x^*, \mu^*, \lambda^*) y > 0 \quad \forall y \neq 0 : \left[ \begin{array}{c} \frac{\partial h(x^*)}{\partial x} \\ \frac{\partial g_a(x^*)}{\partial x} \end{array} \right] y = 0,$$

allora  $x^*$  è una soluzione locale stretta del problema ( $P_0$ ). □

I teoremi (1.1) e (1.2) si particolarizzano in modo ovvio al caso dei problemi ( $P_1$ ) e ( $P_2$ ).

Nel seguito per semplicità di enunciazione *ammetteremo sempre soddisfatte, per i punti di interesse, sia la condizione di regolarità, sia la condizione di stretta complementarità* precedentemente formulate.

## 2 Generalità sui metodi per la soluzione di problemi di programmazione nonlineare

I metodi per la soluzione di problemi di programmazione nonlineare su cui negli ultimi anni si è maggiormente concentrata l'attività di ricerca si riconducono fondamentalmente a due categorie.

La prima comprende i metodi basati sulla trasformazione del problema vincolato in un problema non vincolato o in una successione di problemi non vincolati. Fanno parte di questa categoria gli algoritmi basati su funzioni di penalità sequenziali od esatte, e su funzioni Lagrangiane aumentate sequenziali od esatte. La seconda categoria comprende i metodi basati sulla trasformazione del problema vincolato in una successione di problemi di programmazione quadratica. Nell'ambito di questa categoria si distinguono due classi: nella prima i sottoproblemi hanno i vincoli di uguaglianza e disuguaglianza, nella seconda i sottoproblemi hanno solo vincoli di uguaglianza. Nella letteratura tali metodi sono spesso indicati con le sigle RQP (recursive quadratic programming) o, con maggior dettaglio, RIQP quelli della prima classe e REQP quelli della seconda, ova I ed E significano rispettivamente "inequality" ed "equality". Da notare che nonostante la similarità dei sottoproblemi cui si riducono, i metodi RIQP e REQP si basano su presupposti del tutto differenti, e verranno quindi presentati separatamente.

Evidentemente alla base dell'utilità dei metodi considerati in queste note c'è la possibilità di disporre di algoritmi e di routines efficienti per la soluzione di problemi non vincolati o di problemi di programmazione quadratica. La validità dei metodi di tipo sequenziale si basa sulla dimostrazione analitica della convergenza delle soluzioni dei sottoproblemi risolti alla soluzione del problema originario; la convergenza va possibilmente caratterizzata in termini di rapidità. La validità dei metodi di tipo esatto si basa invece sulla dimostrazione della

corrispondenza delle soluzioni del problema vincolato e di quello non vincolato in cui si riconduce.

Per qualificare i metodi da un punto di vista pratico occorre poi valutare la complessità delle operazioni richieste (in particolare inversioni di matrici, calcolo di derivate, soluzione di sottoproblemi quadratici ecc.) e il loro numero.

### 3 Metodi non vincolati sequenziali ed esatti

#### 3.1 Funzioni di penalità sequenziali

Verranno qui considerate solo le funzioni di penalità sequenziali “esterne”, che devono questa denominazione al fatto che le soluzioni dei problemi non vincolati risultano in generale esterne all’insieme ammissibile. Esistono anche funzioni di penalità sequenziali “interne” (o “funzioni di barriera”), ma solo per problemi in cui l’insieme ammissibile abbia interno non vuoto, il che esclude il caso di presenza di vincoli di uguaglianza. Data la maggiore utilizzabilità delle funzioni di penalità esterne, ci si limiterà a queste.

L’idea base delle funzioni di penalità sequenziali esterne è molto semplice. Con riferimento al problema  $(P_0)$ , si consideri la funzione  $q(x)$  così definita:

$$q(x) = \begin{cases} 0 & x \in X \\ +\infty & x \notin X \end{cases}$$

e si costruisca la funzione

$$F(x) \triangleq f(x) + q(x).$$

È evidente che la minimizzazione nonvincolata di  $F(x)$  fornisce le soluzioni del problema  $(P_0)$ . Ovviamente, a causa della discontinuità di  $F(x)$  sulla frontiera di  $X$ , la minimizzazione di  $F(x)$  non può essere effettuata in pratica, se non in casi banali. È invece possibile costruire una successione di funzioni continuamente differenziabili che hanno come limite la  $F(x)$ , ed è possibile studiare la convergenza dei minimi di tali funzioni alle soluzioni di  $(P_0)$ . Si consideri infatti, anziché la funzione  $q(x)$  una funzione  $p(x)$  continuamente differenziabile e tale che

$$p(x) = \begin{cases} 0 & x \in X \\ > 0 & x \notin X, \end{cases}$$

e si costruisca la funzione

$$F(x; \epsilon) \triangleq f(x) + \frac{1}{\epsilon} p(x), \quad \epsilon > 0.$$

È evidente che, quando  $\epsilon \rightarrow 0$ ,  $F(x; \epsilon) \rightarrow F(x)$ .

Il termine  $(1/\epsilon)p(x)$  viene detto *termine di penalità* in quanto penalizza la violazione dei vincoli ( $p(x) > 0$ ) in modo tanto più sentito quanto più  $\epsilon$  è piccolo; la struttura più comunemente utilizzata per  $p(x)$ , con riferimento al problema  $(P_0)$  è

$$p(x) = \sum_{i=1}^p [h_i(x)]^2 + \sum_{i=1}^m [\max(0, g_i(x))]^2. \quad (6)$$

La funzione  $F(x; \epsilon)$  viene detta *funzione di penalità*; la specifica *esterna* è dovuta al fatto che, se  $\bar{x}$  è un punto di minimo di  $F(x; \epsilon)$  risulta in genere  $p(\bar{x}) \neq 0$  e quindi  $\bar{x} \notin X$ .

Data una successione  $\{\epsilon_k\}$  di valori di  $\epsilon$ , *strettamente decrescente* e tale che  $\epsilon_k \rightarrow 0$  il metodo delle funzioni di penalità sequenziali (esterne) consiste nel risolvere la successione di problemi non vincolati:

$$\min_{x \in \mathbb{R}^n} F(x; \epsilon_k).$$

I risultati più significativi sulla convergenza del metodo sono enunciati nei due teoremi seguenti, il primo relativo ai minimi globali, il secondo ai minimi locali:

**Teorema 3.1** *Siano soddisfatte le seguenti ipotesi:*

i)  $\exists \sigma > 0$  tale che l'insieme  $X_\sigma$  così definito

$$X_\sigma \triangleq \{x : |h_i(x)| \leq \sigma, i = 1, 2, \dots, p; g_i(x) \leq \sigma, i = 1, 2, \dots, m\}$$

(insieme ammissibile rilassato) sia compatto;

ii)  $\forall k$  la funzione  $F(x; \epsilon_k)$  ha un punto di minimo globale  $x^k$ .

Allora, la successione  $x^k$  ha (almeno) una sottosuccessione convergente, e il limite di ogni sottosuccessione convergente di  $\{x^k\}$  è una soluzione globale di  $(P_0)$ .  $\square$

**Teorema 3.2** *Sia  $x^*$  un punto di minimo locale stretto del problema  $(P_0)$ ; allora esistono una successione  $\{x^k\}$  e un intero  $\bar{k}$  tali che  $\{x^k\}$  converge a  $x^*$ , e  $x^k$  per  $k \geq \bar{k}$  è un punto di minimo locale per  $F(x; \epsilon_k)$ .  $\square$*

La costruzione e il calcolo della funzione  $F(x; \epsilon)$  è evidentemente molto semplice, e in ciò consiste la principale attrattiva del metodo. Non altrettanto semplice è invece la minimizzazione della  $F(x; \epsilon)$  per valori di  $\epsilon$  che tendono a zero. È infatti possibile dimostrare, e verificare anche su esempi banali, che al tendere di  $\epsilon$  a zero si produce un malcondizionamento nella matrice Hessiana della  $F(x; \epsilon)$ , là dove questa è definita, che fa sì che l'algoritmo di minimizzazione non vincolata utilizzato per minimizzare la  $F(x; \epsilon)$  converga lentamente e richieda un numero molto elevato di calcoli di funzioni. Per alleviare questo inconveniente,

in pratica si effettua una sequenza di minimizzazioni, per valori decrescenti di  $\epsilon_k$ , assumendo il punto di minimo  $x^k$  come punto iniziale per la minimizzazione di  $F(x; \epsilon_{k+1})$ , nella aspettativa che  $x^k$  sia una buona stima di  $x^{k+1}$ . Tuttavia tale aspettativa è fondata solo se  $\epsilon_k$  non differisce molto da  $\epsilon_{k+1}$ ; e poiché è possibile dimostrare che la velocità di convergenza della successione  $\{x^k\}$  dipende in modo determinante dalla velocità di convergenza della successione  $\{\epsilon_k\}$ , si comprende come evitare il malcondizionamento e la conseguente lenta convergenza nella singola minimizzazione non vincolata comporta in ogni caso una lenta convergenza complessiva, relativa alla successione  $\{x^k\}$ . È anche possibile dimostrare che i minimi locali di  $F(x; \epsilon)$  descrivono, al variare di  $\epsilon$ , traiettorie che possono essere estrapolate, e su questa proprietà si basano le tecniche più sofisticate per determinare stime iniziali dei valori  $x^k$ . Tuttavia i fenomeni di malcondizionamento e di bassa velocità di convergenza sono inerenti al metodo, e costituiscono la principale motivazione dei successivi sviluppi. Osserviamo infine che, se si adotta per  $p(x)$  l'espressione data dalla (6), la funzione  $F(x; \epsilon)$  non risulta due volte differenziabile ovunque, dovendosi escludere i punti in cui qualche vincolo di uguaglianza è attivo, e quindi presumibilmente i punti di minimo del problema  $(P_0)$ , di ciò occorre tenere conto nella scelta del metodo di minimizzazione non vincolata.

### 3.2 Funzioni Lagrangiane aumentate sequenziali

Consideriamo in questo paragrafo dapprima il problema  $(P_1)$  con soli vincoli di uguaglianza. Per esso la funzione lagrangiana diviene

$$L(x, \mu) = f(x) + \mu'h(x),$$

e le condizioni necessarie del primo ordine si riducono all'esistenza di un moltiplicatore  $\mu^*$  tale che, in corrispondenza ad  $x^*$ , soluzione locale di  $(P_1)$ , risulti

$$\nabla_x L(x^*, \mu^*) = 0 \tag{7}$$

$$\nabla_\mu L(x^*, \mu^*) = h(x^*) = 0 \tag{8}$$

La (7) potrebbe far pensare che la funzione  $L(x, \mu^*)$  abbia in  $x^*$  un minimo non vincolato; infatti in  $x^*$  è soddisfatta la condizione necessaria di minimo per  $L(x, \mu^*)$ . Tuttavia ciò non avviene, in quanto la  $L(x, \mu^*)$  non risulta convessa nell'intorno di  $x^*$ . È però possibile "convessificare" la funzione  $L(x, \mu^*)$  con l'aggiunta di un termine di penalità, in modo da ottenere una nuova funzione,

$$L_a(x, \mu^*; \epsilon) \triangleq L(x, \mu^*) + \frac{1}{\epsilon} \|h(x)\|^2$$

che, per  $\epsilon$  sufficientemente piccolo ma comunque tale che  $1/\epsilon$  sia finito, abbia, sotto opportune ipotesi, un minimo non vincolato in  $x^*$ . Ciò può essere verificato tenendo conto del seguente lemma sulle forme quadratiche:

**Lemma 3.1** *Sia  $A$  una matrice  $n \times n$ , e  $B$  una matrice  $p \times n$ . Allora risulta  $x'Ax > 0$  per ogni  $x \neq 0$  che soddisfa  $Bx = 0$  se e solo se esiste un numero  $\bar{c} > 0$  tale che, per ogni  $c \geq \bar{c}$  risulta  $x'(A + cB'B)x > 0$  per ogni  $x \neq 0$ .  $\square$*

Consideriamo ora le condizioni sufficienti del secondo ordine particolarizzate al problema  $(P_1)$ . Esse assicurano che  $x^*$  è una soluzione locale di  $(P_1)$  se esiste un  $\mu^*$  per cui valgono le (7) (8), e la  $y'\nabla_{xx}^2 L(x^*, \mu^*)y > 0 \quad \forall y \neq 0 : \frac{\partial h(x^*)}{\partial x}y = 0$ .

Se si associano rispettivamente  $\nabla_{xx}^2 L(x^*, \mu^*)$  e  $\frac{\partial h(x^*)}{\partial x}$  con le matrici  $A$  e  $B$  del precedente lemma, è immediato dimostrare il teorema:

**Teorema 3.3** *Si supponga che, in corrispondenza ai valori  $x^*, \mu^*$ , siano soddisfatte le condizioni sufficienti affinché  $x^*$  sia un punto di minimo locale stretto per  $(P_1)$  espresse dal Teorema 1.2. Allora esiste un numero  $\bar{\epsilon} > 0$  tale che, per ogni  $\epsilon \in (0, \bar{\epsilon}]$ ,  $x^*$  è un punto di minimo locale non vincolato della funzione  $L_a(x, \mu^*; \epsilon)$ . Viceversa, se per qualche  $\bar{\epsilon}$  e  $\mu^*$ , il punto  $x^*$  soddisfa le condizioni sufficienti del secondo ordine di minimo locale stretto non vincolato per la funzione  $L_a(x, \mu^*; \epsilon)$ , e inoltre risulta  $h(x^*) = 0$ , allora  $x^*$  è un punto di minimo locale stretto per  $(P_1)$ .  $\square$*

Il teorema precedente evidenzia il fatto che, nelle ipotesi specificate, la funzione  $L_a(x, \mu^*, \epsilon)$  si comporta come una funzione di penalità esterna, con la particolarità di non richiedere, per fornire soluzioni locali di  $(P_1)$ , che il coefficiente  $\epsilon$  tenda a zero, bensì che sia inferiore a un valore di soglia  $\bar{\epsilon}$ . Purtroppo l'applicazione diretta del teorema è impedita dal fatto che non è a priori noto il valore del moltiplicatore  $\mu^*$ . Si osservi però che, applicando il teorema delle funzioni implicite all'equazione  $\nabla_x L_a(x, \mu; \epsilon) = 0$ , risulta che esistono un intorno aperto di  $\mu^*$ , denotato con  $S_{\mu^*}$ , un intorno aperto di  $x^*$ , denotato con  $S_{x^*}$ , e una funzione  $x(\mu) : S_{\mu^*} \rightarrow S_{x^*}$ , continuamente differenziabile e tale che  $\nabla_x L_a(x(\mu), \mu; \epsilon) = 0 \quad \forall \mu \in S_{\mu^*}$ ; inoltre fissato  $\epsilon \in (0, \bar{\epsilon}]$ , poichè la matrice  $\nabla_{xx}^2 L_a(x^*, \mu^*; \epsilon)$  è definita positiva, si può assumere che  $\forall (x, \mu) \in S_{x^*} \times S_{\mu^*}$  anche  $\nabla_{xx}^2 L_a(x, \mu; \epsilon)$  sia definita positiva, cosicchè il valore  $x(\mu)$  è quello che, fissato  $\mu \in S_{\mu^*}$ , corrisponde all'unico valore di  $x \in S_{x^*}$  che minimizza rispetto ad  $x$  la funzione  $L_a(x, \mu; \epsilon)$ . È allora facilmente ipotizzabile che se  $\mu^k$  è una buona stima di  $\mu^*$ , la minimizzazione di  $L_a(x, \mu^k; \epsilon)$  per un valore di  $\epsilon$  sufficientemente piccolo, possa fornire una buona stima  $x^k$  di  $x^*$ ; inoltre, data la presenza del termine di penalità, ci si può aspettare che la stima di  $x^*$  sia tanto migliore quanto più  $\epsilon$  è piccolo.

La funzione  $L_a(x, \mu; \epsilon)$  viene detta *funzione lagrangiana aumentata*; e con riferimento ad essa, le considerazioni precedenti suggeriscono l'implementazione di un algoritmo sequenziale basato sullo schema:

0. dati  $x^{k-1}, \mu^k, \epsilon_k, k = 1$ ;
1. si determini un punto di minimo locale  $x^k$  di  $L_a(x, \mu^k, \epsilon_k)$ , utilizzando un algoritmo di minimizzazione nonvincolata a partire dal punto  $x^{k-1}$ ;

2. si determini una nuova stima  $\mu^{k+1}$  di  $\mu^*$ ;
3. si ponga  $\epsilon_{k+1} = \beta\epsilon_k$ , con  $\beta = 1$  se la violazione dei vincoli in  $x^k$  è sufficientemente minore che in  $x^{k-1}$ , con  $\beta < 1$  in caso contrario;
4. si torni ad 1 con  $k \leftarrow k + 1$ ;

in cui resta ancora da definire la modalità di aggiornamento di  $\mu$ , in modo da ottenere che l'algoritmo converga ad una soluzione locale di  $(P_1)$ .

Per quel che riguarda l'aggiornamento di  $\mu$ , i metodi proposti possono essere interpretati sulla base delle proprietà della *funzione duale*  $\psi(\mu)$ , data dalla

$$\psi(\mu) \triangleq L_a(x(\mu), \mu; \epsilon), \quad \mu \in S_{\mu^*}, \quad \epsilon \in (0, \bar{\epsilon}].$$

Se  $x^*$  è una soluzione locale di  $(P_1)$  e  $\mu^*$  il corrispondente moltiplicatore, risulta evidentemente  $x(\mu^*) = x^*$  e  $h(x^*) = 0$ ; valgono allora le relazioni

$$\psi(\mu^*) = L_a(x^*, \mu^*; \epsilon) = L_a(x^*, \mu; \epsilon) \geq \psi(\mu) \quad \forall \mu \in S_{\mu^*},$$

il che dimostra che  $\mu^*$  fornisce un massimo locale per la funzione duale  $\psi(\mu)$ . Queste considerazioni inducono ad effettuare la determinazione di  $\mu^*$  applicando un metodo di minimizzazione nonvincolata alla funzione  $-\psi(\mu)$ . Osserviamo a questo proposito che, per la definizione di  $x(\mu)$  si ha:

$$\nabla_x L_a(x(\mu), \mu; \epsilon) = 0 \quad \forall \mu \in S_{\mu^*} \tag{9}$$

e da questa, derivando formalmente rispetto a  $\mu$

$$\nabla_{xx}^2 L_a(x(\mu), \mu; \epsilon) \frac{\partial x(\mu)}{\partial \mu} + \nabla_{x\mu}^2 L_a(x(\mu), \mu; \epsilon) = 0; \tag{10}$$

poiché inoltre risulta

$$\nabla_{x\mu}^2 L_a(x(\mu), \mu; \epsilon) = \frac{\partial h(x(\mu))'}{\partial x}. \tag{11}$$

si ottiene dalle (10),(11)

$$\frac{\partial x(\mu)}{\partial \mu} = -[\nabla_{xx}^2 L_a(x(\mu), \mu; \epsilon)]^{-1} \frac{\partial h(x(\mu))'}{\partial \mu}. \tag{12}$$

utilizzando la (9) si ottiene

$$\nabla \psi(\mu) = \frac{\partial x(\mu)'}{\partial \mu} \nabla_x L_a(x(\mu), \mu; \epsilon) + \nabla_\mu L_a(x(\mu), \mu; \epsilon) = h(x(\mu))$$

e, derivando quest'ultima e tenendo conto delle (9) e (12) si ottiene

$$\nabla^2\psi(\mu) = -\frac{\partial h(x(\mu))}{\partial\mu} = [\nabla_{xx}^2 L_a(x(\mu), \mu; \epsilon)]^{-1} \frac{\partial h(x(\mu))'}{\partial\mu}. \quad (13)$$

La minimizzazione di  $-\psi(\mu)$  con il metodo della discesa più ripida corrisponde allora all'iterazione

$$\mu^{k+1} = \mu^k + \alpha h(x^k)$$

ove  $\alpha$  è uno spostamento che, sulla base di ulteriori considerazioni, può essere preso pari a  $2/\epsilon_k$ . Se invece la minimizzazione di  $-\psi(\mu)$  viene effettuata con il metodo di Newton, si ha l'iterazione

$$\mu^{k+1} = \mu^k - [\nabla^2\psi(\mu^k)]^{-1} h(x^k)$$

ove  $[\nabla^2\psi(\mu^k)]^{-1}$  è ottenuta tramite la (13).

Poiché nel metodo qui considerato la rapidità di convergenza di  $x^k$  della soluzione  $x^*$  è condizionata da quella di  $\mu^k$  a  $\mu^*$ , è immaginabile che, da questo punto di vista, la seconda formula di aggiornamento risulti più conveniente dalla prima: si può in effetti dimostrare che con la prima formula si ottiene una rapidità di convergenza lineare e con la seconda superlineare. D'altra parte la seconda formula richiede il calcolo di derivate seconde e l'inversione di una matrice, e quindi risulta computazionalmente molto più onerosa della prima. Una soluzione di compromesso è poi costituita da una formula di aggiornamento in cui la matrice  $[\nabla_{xx}^2 L_a(x^k, \mu^k; \epsilon_k)]^{-1}$  è approssimata con un metodo Quasi-Newton.

Le ultime considerazioni sono relative alla sola convergenza locale; per quel che riguarda la convergenza globale, viene forzata dalla riduzione di  $\epsilon$ , riduzione che non deve però essere automatica ad evitare il malcondizionamento tipico dalla funzioni di penalità. Un criterio per decidere sulla riduzione consiste nel valutare il rapporto  $\|h(x^{k+1})\|/\|h(x^k)\|$  e nel ridurre  $\epsilon$  se tale rapporto è maggiore di un valore soglia (0,25 è un valore comunemente accettato).

Quanto detto finora può estendersi al caso di problemi con vincoli di disuguaglianza. Facciamo riferimento al problema  $(P_2)$ , in quanto il caso di presenza contemporanea di vincoli di uguaglianza e disuguaglianza è poi facilmente deducibile. Utilizziamo l'artificio, cui si è fatto cenno nell'introduzione, di introdurre variabili ausiliarie  $y_i$ ,  $i = 1, 2, \dots, m$ , al fine di convertire il problema con vincoli disuguaglianza in un problema equivalente con vincoli di uguaglianza; consideriamo cioè il problema  $(\tilde{P}_2)$ .

Introdotta il vettore di moltiplicatori  $\lambda \in \mathbb{R}^p$ , la funzione lagrangiana aumentata per il problema  $(\tilde{P}_2)$  risulta

$$L_a(x, y, \lambda; \epsilon) = f(x) + \sum_{i=1}^m \lambda_i [g_i(x) + y_i^2] + \frac{1}{\epsilon} \sum_{i=1}^m [g_i(x) + y_i^2]^2. \quad (14)$$

Su tale funzione si potrebbe operare come in precedenza, minimizzando sequenzialmente, per valori fissati di  $\lambda$ , rispetto alla variabile  $z = [x' \quad y']'$ . Tuttavia l'aumento di dimensionalità, rispetto all'incognita originaria, in questa minimizzazione, può essere evitato osservando che la (14) può essere minimizzata *analiticamente* rispetto alle variabili ausiliarie  $y_i$ . Posto

$$w_i(x, \lambda; \epsilon) \triangleq g_i(x) + \frac{\epsilon}{2} \lambda_i$$

si trova che il valore delle variabili ausiliarie in corrispondenza a cui la funzione  $L_a(x, y, \lambda; \epsilon)$ , assume un minimo, per fissati valori di  $x, \lambda, \epsilon$  è dato dalla formula

$$y_i = y_i(x, \lambda; \epsilon) = [-\min(0, w_i(x, \lambda; \epsilon))]^{1/2},$$

e sostituendo  $y_i(x, \lambda; \epsilon)$  al posto di  $y_i$  nella (14) si ottiene la funzione

$$\begin{aligned} L_a(x, \lambda; \epsilon) &\triangleq \min_y L_a(x, y, \lambda; \epsilon) \\ &= f(x) + \sum_{i=1}^m \lambda_i [g_i(x) + y_i^2(x, \lambda; \epsilon)] + \frac{1}{\epsilon} \sum_{i=1}^m [g_i(x) + y_i^2(x, \lambda; \epsilon)]^2 \\ &= f(x) + \lambda' g(x) + \frac{1}{\epsilon} \|g(x)\|^2 - \frac{1}{\epsilon} \sum_{i=1}^m [\min(0, w_i(x, \lambda; \epsilon))]^2, \end{aligned}$$

che è la funzione lagrangiana aumentata associata al problema ( $P_2$ ).

Le considerazioni analitiche ed algoritmiche svolte per il problema con vincoli di uguaglianza possono allora essere estese al problema con vincoli di disuguaglianza, se si assume verificata nei punti di minimo locale l'ipotesi di stretta complementarità. L'unica particolarità specifica, di cui occorre tenere conto nella scelta del metodo di minimizzazione non vincolata, consiste nel fatto che, come si verifica facilmente, la funzione  $L_a(x, \lambda; \epsilon)$  non è derivabile rispetto ad  $x$  due volte ovunque, dovendosi escludere i punti in cui risulta per qualche  $w_i(x, \lambda; \epsilon) = 0$ ; tuttavia, nell'ipotesi di stretta complementarità, risulta  $w_i(x, \lambda; \epsilon) \neq 0 \forall i$ , e quindi nella fase terminale dell'algoritmo l'inconveniente non si verifica. Lo studio della funzione duale della  $L_a(x, \lambda; \epsilon)$  suggerisce, per l'aggiornamento del moltiplicatore  $\lambda$ , la formula del primo ordine

$$\lambda^{k+1} = \max\{0, \lambda^k + \frac{2}{\epsilon_k} g(x^k)\},$$

che può essere interpretata nel modo seguente: se risulta  $w_i(x^k, \lambda^k; \epsilon_k) > 0$  il vincolo  $i$ -esimo viene considerato attivo, e il corrispondente moltiplicatore  $\lambda_i$  aggiornato secondo la formula

$$\lambda_i^{k+1} = \lambda_i^k + \frac{2}{\epsilon_k} g_i(x^k);$$

altrimenti il vincolo viene considerato non attivo, e il corrispondente moltiplicatore nullo; ovvero, da un altro punto di vista, vengono trattati *come se corrispondenti a vincoli di uguaglianza*  $g_i(x) = 0$  i moltiplicatori per cui  $w_i(x, \lambda; \epsilon) > 0$ , e vengono posti nulli gli altri.

Si può dimostrare che questa regola vale anche per definire una formula di aggiornamento del secondo ordine, almeno in tutti i punti in cui risulta  $w_i(x, \lambda; \epsilon) \neq 0 \forall i$ .

Si può concludere questo paragrafo ribadendo che le funzioni qui considerate presentano il vantaggio di superare i problemi di malcondizionamento e di lenta convergenza tipici delle funzioni di penalità sequenziali, a seguito di un modesto aumento della complessità di costruzione dell'algoritmo. Esiste una ampia evidenza sperimentale che testimonia della convenienza di utilizzare, nell'ambito dei metodi nonvincolati sequenziali, le funzioni lagrangiane aumentate.

### 3.3 Funzioni di penalità esatte

Una funzione di penalità esatta, relativa ad un dato problema di ottimo vincolato, è una funzione delle stesse variabili indipendenti presenti nel problema, e la cui minimizzazione non vincolata fornisce le soluzioni del problema stesso. Il termine “esatto”, in alternativa a “sequenziale”, denota appunto il fatto che è richiesta una sola minimizzazione, anziché una sequenza di minimizzazioni come nei casi precedenti.

Le funzioni di penalità esatte proposte per prime risultano non continuamente differenziabili, e richiedono quindi per la loro minimizzazione i metodi specifici dell'ottimizzazione non differenziabile. In questa sede ci si limiterà a considerare le funzioni di penalità esatte continuamente differenziabili introdotte più di recente, e la cui minimizzazione può essere effettuata con i metodi standard che fanno uso del gradiente della funzione da minimizzare.

Anche in questo caso consideriamo dapprima il problema  $(P_1)$ ; sia  $x^*$  una soluzione locale e  $\mu^*$  il corrispondente moltiplicatore. L'idea di base nella costruzione di una funzione di penalità esatta per detto problema consiste nel sostituire al moltiplicatore  $\mu$  che compare nella funzione lagrangiana aumentata  $L_a(x, \mu; \epsilon)$  una funzione  $\mu(x)$ , continuamente differenziabile, e con la proprietà che  $\mu(x^*) = \mu^*$ . In tal modo si ottiene la funzione della sola  $x$

$$L_a(x, \mu(x); \epsilon) = f(x) + \mu(x)'h(x) + \frac{1}{\epsilon} \|h(x)\|^2.$$

Per individuare una formula opportuna per la funzione  $\mu(x)$ , ricordiamo che risulta

$$\nabla f(x^*) + \frac{\partial h(x^*)'}{\partial x} \mu^* = 0$$

da cui, nell'ipotesi di validità della condizione di regolarità in  $x^*$ , si ha

$$\mu^* = - \left[ \frac{\partial h(x^*)}{\partial x} \frac{\partial h(x^*)'}{\partial x} \right]^{-1} \frac{\partial h(x^*)}{\partial x} \nabla f(x^*).$$

Tale eguaglianza suggerisce di considerare come  $\mu(x)$  la funzione data dalla

$$\mu(x) \triangleq - \left[ \frac{\partial h(x)}{\partial x} \frac{\partial h(x)'}{\partial x} \right]^{-1} \frac{\partial h(x)}{\partial x} \nabla f(x),$$

definita nei punti in cui l'indicata inversione di matrice può essere effettuata, e quindi sicuramente in un intorno di  $x^*$ . Nei punti ove è definita, la funzione moltiplicatrice può essere interpretata come quella che, dato  $x$ , fornisce il valore dell'unico moltiplicatore che soddisfa la condizione  $\nabla_x L(x, \mu) = 0$  necessaria perchè  $x$  sia soluzione di  $(P_1)$ .

Si dimostra che questa scelta di  $\mu(x)$  effettivamente fornisce una funzione di penalità esatta per  $(P_1)$ . Più precisamente, denotata con  $G(x; \epsilon)$  la funzione risultante

$$G(x; \epsilon) \triangleq f(x) - h(x)' \left[ \frac{\partial h(x)}{\partial x} \frac{\partial h(x)'}{\partial x} \right]^{-1} \frac{\partial h(x)}{\partial x} \nabla f(x) + \frac{1}{\epsilon} \|h(x)\|^2,$$

definita sull'insieme

$$\tilde{X} \triangleq \left\{ x : \frac{\partial h(x)}{\partial x} \text{ ha rango } p \right\}$$

e continuamente differenziabile, si possono dimostrare i seguenti teoremi:

**Teorema 3.4** *Sia  $\bar{X}$  un sottoinsieme compatto di  $\tilde{X}$ . Si assuma che  $x^*$  sia l'unico punto di minimo globale di  $f$  sull'insieme  $X \cap \bar{X}$ , e che  $x^*$  appartenga all'interno di  $\bar{X}$ . Allora esiste un valore  $\bar{\epsilon} > 0$  tale che, per ogni  $\epsilon \in (0, \bar{\epsilon}]$ ,  $x^*$  è l'unico punto di minimo di  $G(x; \epsilon)$  in  $\bar{X}$ .  $\square$*

**Teorema 3.5** *Sia  $\bar{X}$  un sottoinsieme compatto di  $\tilde{X}$ . Allora esiste un valore  $\bar{\epsilon} > 0$  tale che, per ogni  $\epsilon \in (0, \bar{\epsilon}]$ , se  $x^*$  è un punto di minimo locale nonvincolato di  $G(x; \epsilon)$  appartenente a  $\bar{X}$ , allora  $x^*$  è un punto di minimo locale di  $(P_1)$ .  $\square$*

I precedenti teoremi quindi autorizzano a cercare le soluzioni locali di  $(P_1)$  effettuando minimizzazioni nonvincolate di  $G(x; \epsilon)$ , per valori di  $\epsilon$  sufficientemente piccoli.

Per problemi con vincoli di disuguaglianza la definizione di una funzione di penalità esatta continuamente differenziabile non si deduce direttamente dal caso con vincoli di uguaglianza, come avviene invece per le funzioni lagrangiane aumentate.

Consideriamo il problema  $(P_2)$ . Per questo problema la funzione lagrangiana è

$$L(x, \lambda) = f(x) + \lambda'g(x)$$

e per l'ottimalità di  $x^*$  esiste un vettore  $\lambda^*$  per cui valgono le condizioni necessarie di ottimalità

$$\nabla_x L(x^*, \lambda^*) = \nabla f(x^*) + \frac{\partial g(x^*)'}{\partial x} \lambda^* = 0, \quad (15)$$

$$\lambda_i^* g_i(x^*) = 0 \quad i = 1, 2, \dots, m. \quad (16)$$

Premoltiplicando la (15) per  $\partial g(x^*)/\partial x$  la (16) per  $\gamma^2 g_i(x^*)$ , e sommando, si ottiene la

$$\left[ \frac{\partial g(x^*)}{\partial x} \frac{\partial g(x^*)'}{\partial x} + \gamma^2 G^2(x^*) \right] \lambda^* + \frac{\partial g(x^*)}{\partial x} \nabla f(x^*) = 0, \quad (17)$$

ove  $G(x^*)$  denota la matrice diagonale che ha  $g_i(x^*)$  come  $i$ -esimo elemento della diagonale. Si può dimostrare che, tenendo conto dell'ipotesi di regolarità in  $x^*$ , la matrice che premoltiplica  $\lambda^*$  nella (17) è non singolare, cosicchè la  $\lambda^*$  può essere esplicitata

$$\lambda^* = - \left[ \frac{\partial g(x^*)}{\partial x} \frac{\partial g(x^*)'}{\partial x} + \gamma^2 G^2(x^*) \right]^{-1} \frac{\partial g(x^*)}{\partial x} \nabla f(x^*).$$

Questa espressione suggerisce una funzione moltiplicatrice  $\lambda(x)$  della forma

$$\lambda(x) \triangleq - \left[ \frac{\partial g(x)}{\partial x} \frac{\partial g(x)'}{\partial x} + \gamma^2 G(x) \right]^{-1} \frac{\partial g(x)}{\partial x} \nabla f(x), \quad (18)$$

definita nei punti ove l'inversa indicata esiste e quindi, in particolare, nell'intorno di ogni punto in cui i vincoli sono regolari. In questo caso il valore fornito dalla funzione  $\lambda(x)$  è, per ogni dato  $x$ , quello del moltiplicatore per cui sono soddisfatte le condizioni necessarie di ottimalità per  $(P_2)$  in  $x$  espresse dalle relazioni di uguaglianza:  $\nabla_x L(x, \lambda) = 0$  e  $g_i(x) \lambda_i = 0, i = 1, 2, \dots, m$ .

Trasformando il problema  $(P_2)$  in un problema con vincoli di uguaglianza, considerandone la funzione lagrangiana aumentata data dalla (14) e sostituendo in questa la funzione moltiplicatrice  $\lambda(x)$  al posto del moltiplicatore  $\lambda$ , si perviene alla funzione

$$L_a(x, y, \lambda(x); \epsilon) = f(x) + \sum_{i=1}^m \lambda_i(x) [g_i(x) + y_i^2] + \frac{1}{\epsilon} \sum_{i=1}^m [g_i(x) + y_i^2]^2,$$

e questa funzione può essere minimizzata *analiticamente* rispetto alle variabili ausiliarie  $y_i$ . Posto

$$U(x; \epsilon) \triangleq \min_y L_a(x, y, \lambda(x); \epsilon)$$

risulta

$$\begin{aligned}
U(x; \epsilon) &\triangleq f(x) + \sum_{i=1}^m \lambda_i(x)[g_i(x) + y_i^2(x; \epsilon)] + \frac{1}{\epsilon} \sum_{i=1}^m [g_i(x) + y_i^2(x; \epsilon)]^2 \\
&= f(x) + \lambda(x)'g(x) + \frac{1}{\epsilon} \|g(x)\|^2 - \frac{1}{\epsilon} \sum_{i=1}^m [\min[0, g_i(x) + \frac{\epsilon}{2} \lambda_i(x)]]^2,
\end{aligned}$$

ove  $\lambda_i(x)$  è definito dalla (18) e  $y_i^2(x; \epsilon)$  è dato dalla

$$y_i^2(x; \epsilon) \triangleq -\min[0, (g_i(x) + \frac{\epsilon}{2} \lambda_i(x))].$$

La funzione  $U(x; \epsilon)$  così costruita, definita sull'insieme

$$\tilde{X} \triangleq \{x : \text{i vettori } \nabla g_i(x), i \in I_a(x), \text{ sono linearmente indipendenti}\}$$

è continuamente differenziabile, ed è una funzione di penalità esatta per il problema  $(P_2)$  nello stesso senso in cui  $G(x; \epsilon)$  lo è per  $(P_1)$ . Più precisamente, sostituendo  $U(x; \epsilon)$  e  $\tilde{X}$  al posto di  $G(x; \epsilon)$  e  $\tilde{X}$  vale ancora l'enunciato del Teorema (3.4). Inoltre, con le stesse sostituzioni e, posto  $I_\pi(x) \triangleq \{i : g_i(x) \geq 0\}$ , con l'ulteriore aggiunta della IPOTESI: in ogni punto  $x \in \tilde{X}$  ove risulta  $\sum_{i \in I_\pi} \nabla g_i(x) g_i(x) = 0$  si ha  $g_i(x) = 0$  per ogni  $i \in I_\pi(x)$ , vale anche l'enunciato del Teorema 3.5.

È quindi ragionevole, sulla base dei precedenti teoremi, cercare soluzioni locali del problema  $(P_2)$  con una minimizzazione nonvincolata della funzione  $U(x; \epsilon)$  per un valore di  $\epsilon$  sufficientemente piccolo.

Per quanto attraenti da un punto di vista teorico le funzioni di penalità esatte qui considerate presentano nella pratica utilizzazione due svantaggi, dovuti al fatto che, per come sono definite le funzioni moltiplicatrici  $\mu(x)$  e  $\lambda(x)$  ogni volta che si calcola la funzione di penalità esatta occorre calcolare le derivate prime delle funzioni del problema, e invertire una matrice la cui dimensione è pari a quella dei vincoli. Tale circostanza evidentemente limita il campo di utilizzabilità del metodo ai casi in cui il numero dei vincoli non è eccessivo. Da notare inoltre che, se si vuole effettuare la minimizzazione non vincolata con algoritmi tipo Newton occorre che le funzioni del problema siano tre volte continuamente differenziabili. In questo caso la funzione  $G(x; \epsilon)$  è due volte continuamente differenziabile nel suo insieme di definizione, con la possibile eccezione dei punti in cui risulta  $g_i(x) + \frac{\epsilon}{2} \lambda_i(x) = 0$  per qualche  $i$  e quindi è due volte continuamente differenziabile nell'intorno di ogni punto  $x^*$  di minimo locale in cui valga la condizione di stretta complementarità, risultando ivi  $g_i(x^*) + \frac{\epsilon}{2} \lambda_i(x^*) \neq 0 \forall i$ .

### 3.4 Funzioni Lagrangiane Aumentate Esatte

Le funzioni lagrangiane aumentate esatte consentono di risolvere i problemi vincolati mediante la minimizzazione di una funzione nonvincolata, definita su uno spazio esteso di dimensione pari a quella delle variabili originali del problema

più la dimensione dei vincoli; in compenso per il suo calcolo non sono richieste operazioni di inversione di matrici.

Anche in questo caso facciamo dapprima riferimento al problema  $(P_1)$ . Ricordiamo che per questo problema la funzione lagrangiana aumentata sequenziale è stata ottenuta sommando alla funzione lagrangiana il termine  $\frac{1}{\epsilon}\|h(x)\|^2$ , che penalizza la violazione del vincolo, e quindi la violazione della *condizione necessaria di ottimalità*  $h(x) = 0$ . L'effetto di tale termine, per  $\epsilon$  sufficientemente piccolo, è quello di rendere convessa rispetto ad  $x$  nell'intorno di  $x^*$ , la funzione  $L_a(x, \mu^*; \epsilon)$ .

Una più completa convessificazione della funzione lagrangiana si ottiene sommando a questa oltre al primo termine di penalità anche un secondo termine di penalità relativo alla violazione dell'altra condizione necessaria  $\nabla_x L(x, \mu) = 0$ . Più in particolare si consideri la funzione  $S(x, \mu; \epsilon)$  definita sullo spazio prodotto  $\mathbb{R}^n \times \mathbb{R}^m$  dalla formula

$$S(x, \mu; \epsilon) \triangleq f(x) + \mu' h(x) + \frac{1}{\epsilon} \|h(x)\|^2 + \eta \left\| \frac{\partial g(x)}{\partial x} \nabla_x L(x, \mu) \right\|^2;$$

in cui l'ultimo addendo, con  $\eta > 0$ , rappresenta l'ulteriore termine di penalità ora considerato. Tale funzione risulta continuamente differenziabile rispetto a  $x$  e  $\mu$ , ed è una funzione lagrangiana aumentata esatta nel senso che, per valori di  $\epsilon$  sufficientemente piccoli, la soluzione del problema  $(P_1)$  e i relativi moltiplicatori possono essere messi in corrispondenza con le coppie di valori  $(x, \mu)$  che danno luogo a minimi locali di  $S(x, \mu; \epsilon)$ . Valgono in proposito i seguenti teoremi ove l'insieme  $\tilde{X}$  che compare nell'enunciato del secondo è lo stesso definito nel par. 3.3

**Teorema 3.6** *Sia  $\bar{X} \subset \mathbb{R}^n$  un insieme compatto. Si assuma che  $x^*$  sia l'unico punto di minimo globale di  $f$  sull'insieme  $X \cap \bar{X}$ , e che  $x^*$  sia interno a  $\bar{X}$ . Sia  $\mu^*$  il moltiplicatore corrispondente a  $x^*$ . Allora, per ogni insieme compatto  $\mu \subset \mathbb{R}^m$  contenente  $\mu^*$  nel suo interno, esiste un valore  $\bar{\epsilon}$  tale che, per tutti i valori di  $\epsilon \in (0, \bar{\epsilon}]$ ,  $(x^*, \mu^*)$ , è l'unico punto di minimo globale di  $S(x, \mu, \epsilon)$  nell'insieme  $\bar{X} \times \mu$ .  $\square$*

**Teorema 3.7** *Sia  $\bar{X} \times \mu$  un sottoinsieme compatto di  $\tilde{X} \times \mathbb{R}^m$ . Allora esiste un valore  $\bar{\epsilon} > 0$  tale che per tutti gli  $\epsilon \in (0, \bar{\epsilon}]$ , se  $(x^*, \mu^*)$  fornisce un minimo locale nonvincolato di  $S(x, \mu; \epsilon)$  appartenente a  $\bar{X} \times \mu$ ,  $x^*$  è una soluzione locale del problema  $(P_1)$ , e  $\mu^*$  il corrispondente moltiplicatore.  $\square$*

Alla luce dei precedenti teoremi, la ricerca delle soluzioni locali del problema  $(P_1)$  può essere effettuata minimizzando, rispetto ad  $(x, \mu)$  la funzione  $S(x, \mu; \epsilon)$  per valori di  $\epsilon$  sufficientemente piccoli. L'estensione al caso di problemi con vincoli i disuglianza può essere ottenuta secondo le linee indicate per le funzioni lagrangiane aumentate sequenziali. Si consideri il problema  $(P_2)$ , trasformato con l'aggiunta delle variabili ausiliarie  $y_i$  nel problema  $(\tilde{P}_2)$ . Conviene qui introdurre la notazione

$$Y \triangleq \text{diag}(y_i) \quad i = 1, 2, \dots, m$$

e riscrivere i vincoli del problema  $(\tilde{P}_2)$  nella forma vettoriale  $g(x) + Yy = 0$ . Denotando con  $\tilde{L}(x, y, \lambda)$  la funzione lagrangiana per il problema  $(\tilde{P}_2)$ , si ha

$$\tilde{L}(x, y, \lambda) = f(x) + \lambda'g(x) + \lambda'Yy,$$

e la condizione necessaria di annullamento del gradiente di questa funzione rispetto alle variabili del problema si esplicita nelle

$$\begin{aligned}\nabla_x \tilde{L}(x, y, \lambda) &= \nabla f(x) + \frac{\partial g(x)'}{\partial x} \lambda = 0, \\ \nabla_y \tilde{L}(x, y, \lambda) &= 2Y\lambda = 0.\end{aligned}$$

Osservando che  $\nabla_x \tilde{L}(x, y, \lambda) = \nabla_x L(x, \lambda)$  ove  $L(x, \lambda)$  è la funzione lagrangiana del problema  $(P_2)$ , le condizioni precedenti si possono sintetizzare nella

$$\nabla_z \tilde{L}(x, y, \lambda) = \begin{bmatrix} \nabla_x L(x, \lambda) \\ \dots\dots\dots \\ 2Y\lambda \end{bmatrix} = 0.$$

D'altro canto la matrice Jacobiana dei vincoli del problema  $(\tilde{P}_2)$  è data da

$$\frac{\partial(g(x) + Yy)}{\partial z} = \begin{bmatrix} \frac{\partial g(x)}{\partial x} & \vdots & 2Y \end{bmatrix};$$

sulla base di quanto detto per il problema  $(P_1)$ , si ha allora che la funzione

$$\begin{aligned}S(x, y, \lambda; \epsilon) &= f(x) + \lambda'[g(x) + Yy] + \frac{1}{\epsilon} \|g(x) + Yy\|^2 + \\ &+ \eta \left\| \begin{bmatrix} \frac{\partial g(x)}{\partial x} & \vdots & 2Y \end{bmatrix} \begin{bmatrix} \nabla_x L(x, \lambda) \\ \dots\dots\dots \\ 2Y\lambda \end{bmatrix} \right\|^2\end{aligned}$$

è una funzione lagrangiana aumentata esata per il problema  $(P_2)$ .

Anche in questo caso la minimizzazione rispetto alle variabili ausiliari  $y_i$  può essere effettuata analiticamente; posto

$$w_i(x, \lambda; \epsilon) \triangleq g_i(x) + \frac{\epsilon}{2} \left[ \lambda_i + 8\eta\lambda_i \frac{\partial g_i(x)}{\partial x} \nabla_x L(x, \lambda) \right]$$

si trova che, fissati  $x, \lambda, \epsilon$ , la funzione  $S(x, u, \lambda; \epsilon)$  è minimizzata rispetto alle variabili ausiliari quando

$$y_i = y_i(x, \lambda; \epsilon) = \left[ -\frac{\min(0, w_i(x, \lambda; \epsilon))}{1 + 16\epsilon\eta\lambda_i^2} \right]^{1/2}.$$

Sostituendo  $y_i(x, \lambda; \epsilon)$  al posto di  $y_i$  nella funzione  $S(x, y, \lambda; \epsilon)$  si ottiene la funzione  $T(x, \lambda; \epsilon)$ :

$$\begin{aligned}
T(x, \lambda; \epsilon) &\triangleq f(x) + \sum_{i=1}^m \lambda_i (g_i(x) + y_i^2(x, \lambda; \epsilon)) \\
&+ \frac{1}{\epsilon} \sum_{i=1}^m (g_i(x) + y_i^2(x, \lambda; \epsilon))^2 + \eta \sum_{i=1}^m \left( \frac{\partial g_i(x)}{\partial x} \nabla_x L(x, \lambda) + 4\lambda_i^2 y_i^2(x, \lambda; \epsilon) \right)^2 \\
&= f(x) + \lambda' g(x) + \frac{1}{\epsilon} \|g(x)\|^2 + \eta \left\| \frac{\partial g(x)}{\partial x} \nabla_x L(x, \lambda) \right\|^2 \\
&\quad - \frac{1}{\epsilon} \sum_{i=1}^m \frac{[\min[0, w_i(x, \lambda; \epsilon)]]^2}{1 + 16\epsilon\eta\lambda_i^2}
\end{aligned}$$

che è la funzione lagrangiana aumentata esatta per il problema  $(P_2)$ ; questa funzione è continuamente differenziabile rispetto ad  $x$  e  $\lambda$ .

È possibile dimostrare che, con la sostituzione di  $T(x, \lambda; \epsilon)$  al posto di  $S(x, \mu; \epsilon)$  e di  $\lambda$  al posto di  $\mu$  vale ancora l'enunciato del Teorema (3.6); e con ulteriori sostituzioni di  $\tilde{X}$  definito nel par. 3.3 al posto di  $\tilde{X}$  e di  $(P_2)$  al posto di  $(P_1)$  vale ancora l'enunciato del Teorema (3.7). Per le funzioni lagrangiane aumentate esatte valgono le considerazioni già fatte per le funzioni di penalità esatte relativamente alle conseguenze della presenza delle derivate prime delle funzioni del problema nella espressione della funzione nonvincolata da minimizzare.

## 4 Metodi di programmazione quadratica ricorsiva

### 4.1 Programmazione quadratica ricorsiva con vincoli di disuguaglianza

Facciamo inizialmente riferimento al problema  $(P_1)$ , e ammettiamo che in corrispondenza alla coppia  $(x^*, \mu^*)$  siano soddisfatte le condizioni sufficienti del Teorema (1.2).

Se si effettua uno sviluppo in serie della funzione  $L(x, \mu^*)$  nell'intorno di  $x^*$  si ottiene, troncandolo ai termini del secondo ordine

$$L(x, \mu^*) = f(x^*) + \frac{1}{2}(x - x^*)' \nabla_{xx}^2 L(x^*, \mu^*)(x - x^*) + \dots$$

Uno sviluppo del vincolo troncato ai termini del primo ordine, fornisce poi

$$0 = h(x) = h(x^*) + \frac{\partial h(x^*)}{\partial x}(x - x^*) + \dots$$

È immediato verificare per semplice ispezione, che il problema

$$\begin{aligned} \min_x \quad & f(x^*) + \frac{1}{2}(x - x^*)' \nabla_{xx}^2 L(x^*, \mu^*)(x - x^*) \\ & \frac{\partial h(x^*)}{\partial x}(x - x^*) = 0 \end{aligned} \quad (\tilde{P}Q_1)$$

ha una soluzione in  $x^*$ . Tuttavia il moltiplicatore associato a questa soluzione risulta diverso da  $\mu^*$ . Per evitare questo “bias” del moltiplicatore, basta aggiungere il termine  $\frac{\partial f(x^*)}{\partial x}(x - x^*)$  alla funzione obiettivo di  $(\tilde{P}Q_1)$ ; si ottiene così il problema

$$\begin{aligned} \min_x \quad & \frac{1}{2}(x - x^*)' \nabla_{xx}^2 L(x^*, \mu^*)(x - x^*) + \nabla f(x^*)'(x - x^*) + f(x^*) \\ & \frac{\partial h(x^*)}{\partial x}(x - x^*) = 0 \end{aligned} \quad (PQ_1)$$

che ha come soluzione  $x^*$  e come moltiplicatore corrispondente  $\mu^*$ . Questa osservazione suggerisce di considerare una successione di sottoproblemi di programmazione quadratica del tipo

$$\begin{aligned} \min_{\delta} \quad & \frac{1}{2} \delta' \nabla_{xx}^2 L(x^k, \mu^k) \delta + \nabla f(x^k)' \delta + f(x^k) \\ & \frac{\partial h(x^k)}{\partial x} \delta = -h(x^k) \end{aligned} \quad k = 0, 1, \dots \quad (\tilde{P}Q_1^{k+1})$$

ove si è posto  $\delta = x - x^k$ , e ove  $(x^k, \mu^k)$  sono le stime correnti della soluzione e del moltiplicatore; la soluzione fornisce la nuova stima  $x^{k+1}$  e il nuovo moltiplicatore  $\mu^{k+1}$ . Per giustificare la convergenza locale dell’algoritmo così delineato, si osservi che le condizioni necessarie per la soluzione del sottoproblema  $(\tilde{P}Q_1^{k+1})$  forniscono il seguente sistema di equazioni nelle incognite  $\delta, \mu$ :

$$\begin{bmatrix} \nabla_{xx}^2 L(x^k, \mu^k) & \frac{\partial h(x^k)'}{\partial x} \\ \frac{\partial h(x^k)}{\partial x} & 0 \end{bmatrix} \begin{bmatrix} \delta \\ \mu \end{bmatrix} = - \begin{bmatrix} \nabla f(x^k) \\ h(x^k) \end{bmatrix} \quad (19)$$

e tale sistema coincide con quello che si ottiene applicando il metodo di Newton alla soluzione del sistema di equazioni nelle incognite  $x, \mu$  fornito dalle condizioni necessarie del problema  $(P_1)$ . È infatti immediato verificare che il metodo di Newton, applicato al sistema di equazioni nelle incognite  $x, \mu$

$$\begin{cases} \nabla f(x) + \frac{\partial h(x)'}{\partial x} \mu = 0 \\ h(x) = 0 \end{cases}$$

fornisce, dati  $x^k$  e  $\mu^k$ , nuovi valori  $x^{k+1}, \mu^{k+1}$  che sono soluzione del sistema

$$\begin{bmatrix} \nabla_{xx}^2 L(x^k, \mu^k) & \frac{\partial h(x^k)'}{\partial x} \\ \frac{\partial h(x^k)}{\partial x} & 0 \end{bmatrix} \begin{bmatrix} x - x^k \\ \mu - \mu^k \end{bmatrix} = - \begin{bmatrix} \nabla_x L(x^k, \mu^k) \\ h(x^k) \end{bmatrix}$$

e che coincidono con quelli dati dal sistema (19) tenendo conto che i termini in  $\mu^k$  nella prima equazione si cancellano, e che la seconda equazione non dipende da  $\mu^k$ .

In sostanza le soluzioni dei sottoproblemi quadratici convergono alla soluzione del problema ( $P_1$ ) nelle stesse ipotesi, e con le stesse modalità con cui il metodo di Newton applicato alle condizioni necessarie di ( $P_1$ ) converge; in particolare, vale il seguente teorema

**Teorema 4.1** *Sia  $x^*$  un punto di minimo locale isolato per il problema ( $P_1$ ) e sia  $\mu^*$  il corrispondente moltiplicatore. Si ammetta che per la coppia  $(x^*, \mu^*)$  siano soddisfatte le condizioni sufficienti di minimo del Teorema (1.2). Allora nello spazio prodotto  $\mathbb{R}^n \times \mathbb{R}^m$  esiste un intorno aperto del punto  $(x^*, \mu^*)$  tale che, se  $(x^0, \mu^0)$  appartiene a detto intorno, la successione  $\{(x^k, \mu^k)\}$  generata dalla soluzione dei sottoproblemi di programmazione quadratica ( $\tilde{P}Q_1^{k+1}$ ),  $k = 0, 1, \dots$  converge a  $(x^*, \mu^*)$ , con velocità di convergenza superlineare (quadratica se le funzioni  $f, g$ , sono tre volte continuamente differenziabili).  $\square$*

La convenienza di risolvere la successione di sottoproblemi di programmazione quadratica, anzichè applicare il metodo di Newton alle condizioni necessarie per il problema ( $P_1$ ), consiste nel fatto di potere discriminare i punti di minimo da quelli che soddisfano condizioni solo necessarie.

Se invece del problema ( $P_1$ ) si considera il problema ( $P_2$ ) si arriva facilmente, sulla base di sviluppi del tutto analoghi e quelli già svolti, a considerare sottoproblemi di programmazione quadratica del tipo

$$\min_{\delta} \begin{aligned} & \frac{1}{2} \delta' \nabla_{xx}^2 L(x^k, \mu^k) \delta + \nabla f(x^k)' \delta + f(x^k) \\ & \frac{\partial g(x^k)}{\partial x} \delta + g(x^k) \leq 0. \end{aligned} \quad (\tilde{P}Q_2^{k+1})$$

Il fatto che quando nel problema originario sono presenti vincoli di disuguaglianza i sottoproblemi di programmazione quadratica hanno i vincoli di disuguaglianza, giustifica la denominazione di *RIQP* per i metodi basati sulle precedenti considerazioni.

Il metodo ora descritto costituisce la base per tutta una classe di metodi derivati, con varianti di diversa natura. Ad esempio un evidente svantaggio del metodo descritto consiste nel richiedere il calcolo delle derivate seconde delle funzioni del problema; queste infatti sono presenti nella matrice Hessian  $\nabla_{xx}^2 L(x, \mu)$ . Tuttavia, in maniera analoga a quanto avviene nell'ottimizzazione non vincolata,

è possibile sostituire a detta matrice delle approssimazioni ottenute con formule di aggiornamento quasi-Newton, conservando una velocità di convergenza superlineare.

Un altro aspetto che accomuna il metodo descritto al metodo di Newton per la soluzione di problemi non vincolati consiste nel fatto che il teorema (4.1) assicura la convergenza solo se la coppia  $(x_0, \mu_0)$  è sufficientemente prossima a  $(x^*, \mu^*)$ . Per allargare la regione di convergenza del metodo è allora possibile combinarlo con qualche altro metodo che goda di soddisfacenti proprietà di convergenza globale. In sostanza, si cerca di ottenere algoritmi combinati e dotati di un opportuno criterio di commutazione, tali che, se il punto corrente è abbastanza vicino all'ottimo commutano automaticamente su un metodo di *RIQP*, con convergenza locale superlineare, mentre se il punto corrente è lontano dall'ottimo commutano automaticamente su un algoritmo con proprietà di convergenza globale verso un punto di Kuhn-Tucker del problema. È evidente che candidati all'uso come metodi globali sono i metodi basati sulle funzioni di penalità e lagrangiane aumentate sequenziali.

In alternativa a questo modo di procedere, per allargare la regione di convergenza è stato anche proposto di effettuare una ricerca di linea lungo la direzione  $\delta$  determinata dalla soluzione del sottoproblema, riducendo in questa ricerca una funzione di penalità esatta. A questo proposito è stato studiato l'uso della funzione non differenziabile

$$\phi(x) \triangleq \nu f(x) + \sum_{i=1}^p |h_i(x)| + \sum_{i=1}^m \max\{0, g_i(x)\}$$

che, per  $\nu > 0$  sufficientemente piccolo è una funzione di penalità esatta; più recentemente è stato mostrato che per i problemi con vincoli di uguaglianza si presta allo stesso scopo la funzione lagrangiana aumentata esatta continuamente differenziabile  $S(x, \mu^k; \epsilon)$ , nei cui confronti la direzione  $\delta$  risulta essere di discesa nell'intorno della soluzione. Una caratteristica degli algoritmi di *RIQP*, è che, unici tra quelli qui considerati, non dipendono, almeno per il calcolo della direzione, da coefficienti di penalità.

La principale complicazione del metodo consiste proprio nel fatto di dovere risolvere ad ogni iterazione un problema di programmazione quadratica, con tanti vincoli quanti ce ne sono nel problema originario, e di cui non è detto che l'insieme ammissibile sia non vuoto.

## 4.2 Programmazione quadratica ricorsiva con vincoli di uguaglianza

I metodi di programmazione quadratica che risolvono sottoproblemi con soli vincoli di uguaglianza sono derivati da quelli di penalità sequenziale.

Facciamo riferimento al problema  $(P_0)$ . Se si utilizza la (6), la funzione di penalità può essere scritta nel seguente modo:

$$F(x; \epsilon) = f(x) + \frac{1}{\epsilon} \|c(x)\|^2,$$

ove con  $c(x)$  si è indicato il vettore formato dai vincoli di uguaglianza e dai vincoli di disuguaglianza non soddisfatti. Si è già accennato al fatto che i minimi locali della  $F(x; \epsilon)$  descrivono, al variare di  $\epsilon$ , una traiettoria la cui determinazione è resa difficile, nella fase terminale, dal malcondizionamento della matrice Hessiana  $\nabla^2 F(x; \epsilon)$ .

L'idea base del metodo di *REQP* è quella di generare una successione di punti la cui traiettoria approssimi, ed eguagli al limite per  $\epsilon \rightarrow 0$ , una traiettoria di minimi locali della  $F(x; \epsilon)$ , la cui determinazione comporti la soluzione di problemi ben condizionati.

A tale scopo si indichi con  $x^k$  la stima attuale della soluzione del problema ( $P_0$ ); la stima successiva può essere ottenuta minimizzando rispetto a  $\delta$ , e con un valore opportuno di  $\epsilon$ , la funzione  $F(x^k + \delta; \epsilon)$ . Tale minimizzazione a sua volta può essere effettuata *in modo approssimato* determinando il valore di  $\delta$  per cui uno sviluppo del gradiente  $\nabla F(x^k + \delta; \epsilon)$  troncato ai termini del primo ordine si annulla, e cioè, supponendo che in  $x^k$  e in  $x^k + \delta$  i vincoli da includere nel vettore  $c(x)$  siano gli stessi, determinando il valore di  $\delta$  per cui si ha

$$\nabla F(x^k; \epsilon) + \nabla^2 F(x^k; \epsilon)\delta = 0. \quad (20)$$

Tale equazione in  $\delta$  coincide evidentemente con l'equazione di Newton, calcolata in  $x^k$  relativa al sistema di equazioni nonlineari  $\nabla F(x; \epsilon) = 0$ . La sua soluzione per piccoli valori di  $\epsilon$  è però resa difficile dal malcondizionamento della matrice  $\nabla^2 F$ . Se si esplicita l'espressione di  $\nabla^2 F$ :

$$\nabla F(x^k; \epsilon) = \nabla^2 f(x^k) + \frac{2}{\epsilon} \frac{\partial c(x^k)'}{\partial x} \frac{\partial c(x^k)}{\partial x} + \frac{2}{\epsilon} \sum_i [\nabla^2 c_i(x^k)] c_i(x^k)$$

si evidenzia il fatto che il malcondizionamento è dovuto sostanzialmente al secondo addendo, in quanto nel terzo addendo a piccoli valori di  $\epsilon$  dovrebbero corrispondere piccoli valori di  $c_i(x^k)$ .

Per ovviare alla difficoltà del malcondizionamento, cominciamo col riscrivere la (20), utilizzando la notazione

$$W(x^k; \epsilon) \triangleq \nabla^2 f(x^k) + \frac{2}{\epsilon} \sum_i [\nabla^2 c_i(x^k)] c_i(x^k)$$

ed esplicitando il termine  $\nabla F(x^k; \epsilon)$ :

$$\nabla f(x^k) + \frac{2}{\epsilon} \frac{\partial c(x^k)'}{\partial x} c(x^k) + W(x^k; \epsilon)\delta + \frac{2}{\epsilon} \frac{\partial c(x^k)'}{\partial x} \frac{\partial c(x^k)}{\partial x} \delta = 0. \quad (21)$$

Supponendo che  $W^{-1}$  esista, premoltiplicando la (21) per

$$\frac{\epsilon}{2} \frac{\partial c(x^k)}{\partial x} [W(x^k; \epsilon)]^{-1}$$

si ottiene:

$$\begin{aligned}
& \left( \frac{\epsilon}{2}I + \frac{\partial c(x^k)}{\partial x} [W(x^k; \epsilon)]^{-1} \frac{\partial c(x^k)'}{\partial x} \right) \frac{\partial c(x^k)}{\partial x} \delta + \\
& + \frac{\epsilon}{2} \frac{\partial c(x^k)}{\partial x} [W(x^k; \epsilon)]^{-1} \nabla f(x^k) + \\
& + \frac{\partial c(x^k)}{\partial x} [W(x^k; \epsilon)]^{-1} \frac{\partial c(x^k)'}{\partial x} c(x^k) = 0.
\end{aligned}$$

Aggiungendo ad ambo i membri il termine  $(\epsilon/2)c(x^k)$ , e riordinando, si ottiene:

$$\begin{aligned}
& \left( \frac{\epsilon}{2}I + \frac{\partial c(x^k)}{\partial x} [W(x^k; \epsilon)]^{-1} \frac{\partial c(x^k)'}{\partial x} \right) \left( \frac{\partial c(x^k)}{\partial x} \delta + c(x^k) \right) \\
& = -\frac{\epsilon}{2} \left( \frac{\partial c(x^k)}{\partial x} [W(x^k; \epsilon)]^{-1} \nabla f(x^k) - c(x^k) \right).
\end{aligned} \tag{22}$$

Se si introduce la variabile ausiliaria  $u$ , si riconosce che il valore  $\delta$  che soddisfa l'equazione (22) soddisfa anche il sistema

$$\frac{\partial c(x^k)}{\partial x} \delta = -\frac{\epsilon}{2}u - c(x^k), \tag{23}$$

$$\begin{aligned}
& \left( \frac{\epsilon}{2}I + \frac{\partial c(x^k)}{\partial x} [W(x^k; \epsilon)]^{-1} \frac{\partial c(x^k)'}{\partial x} \right) u \\
& = \frac{\partial c(x^k)}{\partial x} [W(x^k; \epsilon)]^{-1} \nabla f(x^k) - c(x^k).
\end{aligned} \tag{24}$$

Notiamo che la matrice che premoltiplica  $u$  nella (24) non presenta problemi di malcondizionamento dovuti al fatto che  $\epsilon \rightarrow 0$ . Tuttavia, dato  $u$ , la (23) non specifica in modo unico  $\delta$  se, come avviene usualmente, l'equazione omogenea in  $\delta' \frac{\partial c(x^k)}{\partial x} \delta = 0$  ha soluzioni non nulle.

In particolare, una soluzione delle (23) può essere determinata risolvendo il problema di programmazione quadratica.

$$\begin{aligned}
& \min_{\delta} \quad \frac{1}{2} \delta' W(x^k; \epsilon) \delta + \nabla f(x^k)' \delta \\
& \quad \frac{\partial c(x^k)}{\partial x} \delta = -\frac{\epsilon}{2}u - c(x^k) \leq 0.
\end{aligned} \tag{EQP}^{k+1}$$

la cui soluzione, se  $W$  è *definita positiva*, esiste, è unica ed è data, tenendo conto del fatto che  $u$  soddisfa la (24), dalla

$$\delta = [W(x^k; \epsilon)]^{-1} \left( \frac{\partial c(x^k)'}{\partial x} U - \nabla f(x^k) \right); \tag{25}$$

inoltre si verifica facilmente che tale soluzione è *anche* soluzione della (20). In definitiva quindi la soluzione della (20) può essere determinata, se  $W$  è definita positiva, risolvendo la (24) rispetto a  $u$  ed effettuando poi le operazioni indicate nella (25); è evidente che operando in questo modo i problemi di malcondizionamento causati dal tendere di  $\epsilon$  a zero nella matrice  $\nabla^2 F$  vengono superati. Alla base degli algoritmi di *REQP* vi è quindi l'espedito ora illustrato. Nella effettiva implementazione, un algoritmo *REQP* deve incorporare:

- una procedura per decidere quali vincoli includere nel vettore  $c$ , tenendo conto che l'ipotesi che in  $x^k$  e in  $x^k + \delta$  i vincoli violati siano gli stessi può non essere soddisfatta; in particolare dato che il vettore  $u$  può essere considerato una approssimazione del vettore dei moltiplicatori di Lagrange per il problema  $(P_0)$  viene consigliato, su base euristica, di includere nel vettore  $c$  oltre ai vincoli di uguaglianza e a quelli di disuguaglianza non soddisfatti anche i vincoli di disuguaglianza soddisfatti, nel caso che, presenti in  $c$  nella iterazione precedente, abbiano dato luogo a componenti di  $u$  negative;
- un criterio per aggiornare il coefficiente di penalità  $\epsilon$ , tenendo conto che questo aggiornamento può essere effettuato prima della soluzione di un nuovo sottoproblema quadratico, anche se nell'iterazione precedente la funzione di penalità  $F(x; \epsilon)$  non era stata ancora minimizzata (ricordiamo che il valore di  $\delta$  che risolve  $(EQP^{k+1})$  è solo una approssimazione del valore che minimizza  $F(x^k + \delta; \epsilon)$ );
- una ricerca di linea lungo la direzione individuata da  $\delta$ , tenendo conto che data la natura Newtoniana dell'iterazione, non necessariamente lo spostamento  $\delta$  è tale che  $F(x^k + \delta; \epsilon) \leq F(x^k; \epsilon)$ ; ovviamente in questo caso la funzione da considerare nella ricerca di linea è la funzione di penalità stessa;
- una strategia di comportamento nel caso in cui  $W$  non è definita positiva; in particolare quest'ultimo punto può essere affrontato sostituendo a  $W$  una sua approssimazione di tipo Quasi-Newton, che presenta anche il vantaggio di non richieder l'uso di derivate seconde.

I risultati di convergenza degli algoritmi di *REQP* sono troppo articolati per potere essere qui citati testualmente; in sintesi è possibile dimostrare la convergenza di detti algoritmi ad un punto stazionario di una funzione di penalità esatta per il problema  $(P_0)$ , con velocità di convergenza superlineare.

Il principale vantaggio degli algoritmi di *REQP* nei confronti di quelli di *RIQP*, consiste nel fatto che, essendo presenti nel sottoproblema quadratico solo alcuni dei vincoli del problema originario, il carico computazionale associato al calcolo della direzione è di norma molto ridotto; inoltre è possibile dimostrare che la presenza del termine in  $\epsilon$  nel vincolo lineare consente di rendere non vuoto l'insieme ammissibile del sottoproblema di programmazione quadratica. D'altra parte il fatto di basarsi ad ogni iterazione su *una stima* dell'insieme dei vincoli attivi alla soluzione può causare tipici fenomeni di zigzagging.

## 5 Considerazioni conclusive

A conclusione di quanto detto in precedenza, ricapitoliamo vantaggi e svantaggi relativi dei metodi presentati, ed accenniamo ad ulteriori problemi comuni alla loro utilizzazione.

Per quel che riguarda le funzione di penalità sequenziali, alla semplicità di implementazione fa riscontro il malcondizionamento inevitabile per piccoli valori del coefficiente di penalità. Ciò limita oggi l'impiego delle funzioni di penalità sequenziali a casi in cui non è richiesta molta precisione nella soluzione del problema e nel soddisfacimento dei vincoli; ovvero alla ricerca di stime della soluzione, da utilizzare come valori iniziali per altri algoritmi. A questo proposito è da notare che oltre ad una stima della soluzione, si può ottenere una stima dei moltiplicatori di Lagrange richiesta al passo iniziale degli altri metodi considerati; ad esempio, per il problema ( $P_1$ ) si ha:

$$\nabla F(x^k; \epsilon_k) = \nabla f(x^k) + \frac{2}{\epsilon_k} \frac{\partial h(x^k)'}{\partial x} h(x^k) = 0,$$

e dovendo risultare

$$\nabla f(x^*) + \frac{\partial h(x^*)'}{\partial x} \mu^* = 0,$$

si riconosce che il termine  $(2/\epsilon_k)h(x^k)$  fornisce una stima di  $\mu^*$ .

Le funzioni lagrangiane aumentate sequenziali sono meno soggette a fenomeni di malcondizionamento, e consentono una maggiore velocità di convergenza, rispetto alle funzioni di penalità sequenziali; ciò è confermato in modo non controverso sia dall'analisi teorica che dalla sperimentazione numerica.

Le funzioni di penalità esatte evitano di dover risolvere una successione di sottoproblemi, ma poichè richiedono ad ogni calcolo della funzione non vincolata l'inversione di una matrice di dimensione pari a quella dei vincoli, il loro uso è di fatto limitato al caso in cui il numero dei vincoli non è eccessivo. Le funzioni lagrangiane aumentate esatte sono definite su uno spazio esteso, di dimensione pari a quello delle variabili del problema più quello dei vincoli; ed anche per il loro calcolo non è richiesta l'inversione di matrici, è in tale spazio che va effettuata la minimizzazione. Ricordiamo inoltre che nelle funzioni esatte considerate compaiono le derivate prime delle funzioni del problema, e quindi una minimizzazione non vincolata che faccia uso del gradiente richiede il calcolo delle derivate seconde delle stesse funzioni.

Gli algoritmi di programmazione quadratica ricorsiva richiedono, ad ogni iterazione, la soluzione di un sottoproblema di programmazione quadratica; nel caso *RIQP* il numero di vincoli è sempre pari a quello del problema originario, e l'insieme ammissibile può risultare vuoto; nel caso della *REQP* intervengono solo i vincoli "di lavoro", che possono cambiare da iterazione a iterazione producendo fenomeni di zig-zagging. Trattandosi di metodi sequenziali, è d'obbligo una valutazione comparativa con i metodi basati su funzioni lagrangiane aumentate sequenziali; a questo riguardo non esiste una teoria precisa, ma solo indicazioni di massima, confermate dall'esperienza numerica, che fanno preferire

i metodi di *RQP* quando i vincoli del problema siano in prevalenza lineari, o quando la dimensione complessiva del problema (numero di variabili più numero di vincoli) sia al di sotto di un valore di soglia; al di sopra di tale valore alla complicazione di dovere risolvere il sottoproblema di programmazione quadratica fa preferire l'impiego di funzioni lagrangiane aumentate.

L'ovvia conclusione di questo paragrafo è che non esiste un metodo che risulti sempre migliore degli altri; la convenienza di utilizzare un metodo rispetto ad altri è condizionata sia dalla specificità del problema da risolvere, e in particolare della dimensione complessiva, dalla semplicità di calcolo delle funzioni del problema e delle derivate richieste, dalla struttura dei vincoli; sia dalla disponibilità di routines efficienti per effettuare le minimizzazioni non vincolate o per risolvere i sottoproblemi quadratici.

## 6 Note Bibliografiche

Nell'elenco che conclude queste note sono inclusi alcuni riferimenti di rilevanza per un approfondimento delle problematiche considerate.

Per quel che riguarda i fondamenti analitici della programmazione nonlineare un testo ormai classico è quello di Mangasarian [1] cui si rimanda per le dimostrazioni dei Teoremi (1.1)(1.2); tra le monografie più recenti, e che considerano anche alcuni aspetti algoritmici, vanno citate quella di Avriel [2], Giannessi [3], Mc Cormick [4]. Più specificamente orientate agli aspetti algoritmici e computazionali sono le opere di Fletcher [5], e di Gill et al. [6].

L'idea di ottenere la soluzione di un problema di ottimo vincolato come limite di una successione di problemi nonvincolati appare per la prima volta nel 1943 in un lavoro di Courant, ma è solo nel 1968 che trova una compiuta estensione, con il testo di Fiacco e Mc Cormick [7]. Tale testo è stato di importanza fondamentale, ed ha costituito per anni un riferimento base per la soluzione numerica di problemi di programmazione nonlineare. In particolare la dimostrazione del teorema (3.2) può essere dedotta, con facili varianti, dai risultati riportati nel capitolo relativo alle funzioni di penalità esterne. Per la dimostrazione del teorema (3.1) si rinvia invece a [2].

Le funzioni lagrangiane aumentate sequenziali sono state introdotte nella forma qui presentata, da Hestenes nel 1969 [8]; Hestenes denominò il suo approccio *metodo dei moltiplicatori* per il ruolo che vi giocano i moltiplicatori di Lagrange, e tale dizione è ricorrente in molta letteratura. Nello stesso anno Powell ha proposto l'impiego di cosiddette *funzioni di penalità traslate* (sequenziali) [9], che pur basandosi su presupposti diversi, risultano equivalenti a quelle di Hestenes; per cui la paternità delle funzioni lagrangiane aumentate sequenziali è comunemente attribuita ad entrambi Hestenes e Powell. L'estensione al caso di vincoli di disuguaglianza è stata proposta nel 1973 da Rockafellar [10]. Il testo di Hestenes del 1975 [11] tratta in maniera estesa la teoria della "aumentabilità", ed ad esso si rimanda per il teorema (3.3); il testo di Pierre e Lowe [12], sempre del 1975, riporta una documentazione estensiva su aspetti algoritmici e numerici, corredata dei listings dei programmi di calcolo sviluppati. La dimostrazione della

superiorità delle funzioni lagrangiane aumentate sequenziali, rispetto a quelle di penalità, è avvenuta principalmente ad opera di Bertsekas in una serie di lavori pubblicati tra il 1973 e il 1979, che costituiscono parte del materiale esposto in forma organica nella monografia [13], pubblicata nel 1982 dallo stesso Bertsekas; a tale riferimento si rinvia per i teoremi 3.4 e 3.5. Altri importanti contributi, in particolare relativi alla possibilità di definire, utilizzando funzioni lagrangiane aumentate sequenziali, algoritmi che godano delle stesse attrattive di cui godono gli algoritmi Quasi-Newton nella ottimizzazione nonvincolata, sono dovuti a Fletcher [14], Tapia [15], Han [16].

Le funzioni di penalità esatte continuamente differenziabili sono state introdotte da Fletcher nel 1970, per problemi con vincoli di uguaglianza [17], e da Di Pillo e Grippo nel 1982 per problemi con vincoli di disuguaglianza [18]; per la dimostrazione dei teoremi conviene fare riferimento a [13]. L'impiego nella definizione di algoritmi di tipo Newton o Quasi-Newton per problemi vincolati è stato discusso rispettivamente in [15] e [16].

Le funzioni lagrangiane aumentate esatte sono state introdotte da Di Pillo e Grippo nel 1979 per problemi con vincoli di uguaglianza [19] e nel 1982 per problemi con vincoli di disuguaglianza [20]; in particolare in [19] sono dimostrati i teoremi (3.6) e (3.7). L'impiego di queste funzioni nella costruzione di algoritmi di tipo Newton e Quasi-Newton è stato discusso da Bertsekas in [21], da Di Pillo, Grippo e Lampariello in [22], [23]; la possibilità di utilizzarle come funzioni da minimizzare nella ricerca unidimensionale associata ad algoritmi di *RQP* è stata dimostrata da Dixon in [24].

La già citata monografia di Bertsekas dedica alle funzioni esatte continuamente differenziabili un capitolo in cui sono sistematizzati i contributi apparsi fino al 1981, e parte di quelli apparsi nel 1982. L'idea di risolvere problemi di programmazione nonlineare con un metodo basato su sottoproblemi di programmazione quadratica appare in una tesi del 1963 ad opera di Wilson [25], che propose l'algoritmo di *RIQP* noto in letteratura come *SOLVER*. Ulteriori sviluppi, nell'ambito dei metodi di *RIQP*, sono dovuti a Han [26] e a Fletcher [27], per quel che riguarda la definizione di algoritmi con convergenza globale, e a Powell [28], per quel che riguarda la possibilità di garantire la velocità di convergenza superlineare senza utilizzare derivate seconde. I metodi di *REQP* originati dalle funzioni di penalità sequenziali sono stati introdotti da Murray nel 1969 [29]; i successivi sviluppi sono dovuti principalmente a Biggs, in lavori pubblicati tra il 1972 e il 1978: si veda ad esempio [30], [31]. Le connessioni tra i metodi di *RQP* e quelli basati su funzioni di penalità o lagrangiane aumentate esatte sono illustrate nella monografia [13], cui si rimanda anche per il teorema 4.1.

I metodi descritti sono implementati in numerose routines, alcune delle quali disponibili in librerie di larga diffusione (ad esempio quelle della *NAG* e quella di Harwell), altre di meno facile reperimento. Una rassegna del software circolante al 1981 è stata fatta da Schittkowski in [32]; lo stesso autore ha sottoposto le routines disponibili ad una ampia sperimentazione numerica i cui risultati, riportati in [33], convalidano le considerazioni del paragrafo 6.

È di interesse in queste note ricordare che i metodi per la soluzione di problemi di programmazione nonlineare possono essere utilizzati nella soluzione numerica

di problemi di controllo ottimo, le cui approssimazioni con problemi a dimensione finita risultano spesso problemi di programmazione nonlineare dotati di particolare struttura. L'impiego di funzioni di penalità sequenziali nei problemi di controllo ottimo è stato proposto da Balakrishnan nel 1968 [34], ed è noto con la denominazione di "metodo  $\epsilon$ "; l'impiego di funzioni lagrangiane aumentate sequenziali è stato studiato da Di Pillo, Grippo e Lampariello nel 1974 [35]; gli stessi autori hanno recentemente mostrato come la particolare struttura dei problemi di controllo ottimo a tempo discreto o discretizzato può essere sfruttata in connessione all'impiego di funzioni lagrangiane aumentate esatte [36], [37]. L'impiego della *REQP* nella determinazione di traiettorie ottime di velivoli spaziali è stato illustrato da Bartolomew-Biggs nel 1980 [38].

## References

- [1] O. L. Mangasarian. *Nonlinear Programming*. Prentice-Hall, 1969.
- [2] M. Avriel. *Nonlinear Programming: Analysis and Methods*. Prentice-Hall, 1976.
- [3] F. Giannessi. *Metodi Matematici della Programmazione*. Pitagora Editrice, 1982.
- [4] G. P. Mc Cormick. *Nonlinear Programming: Theory, Algorithms, and Applications*. J. Wiley and Sons, 1983.
- [5] R. Fletcher. *Practical Methods of Optimization: Constrained Optimization*, volume 2. J. Wiley and Sons, 1981.
- [6] W. Murray P.E. Gill and M. H. Wright. *Practical Optimization*. Academic Press, 1981.
- [7] A. V. Fiacco and G. P. Mc Cormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. J. Wiley and Sons, 1968.
- [8] M. R. Hestenes. Multiplier and Gradient Methods. *JOTA*.
- [9] M. J. D. Powell. A method for Nonlinear Constraints in Minimization Problems. In R. Fletcher, editor, *Optimization*. Academic Press, 1969.
- [10] R. T. Rockafellar. The Multiplier Method of Hestenes and Powell Applied to Convex Programming. *JOTA*, 12:555–562, 1973.
- [11] M. R. Hestenes. *Optimization Theory*. J. Wiley and Sons, 1975.
- [12] D. A. Pierre and M. J. Lowe. *Mathematical Programming via Augmented Lagrangians*. Addison-Wesley, 1975.
- [13] D. P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, 1982.

- [14] R. Fletcher. An Ideal Penalty Function for Constrained Optimization. *J. Inst. Maths. Applics.*, 5:319–342, 1975.
- [15] R. A. Tapia. Diagonalized Multiplier Methods and Quasi-Newton Method for Constrained Optimization. *JOTA*, 22:135–194, 1977.
- [16] S. P. Han. Penalty Lagrangian Methods via a Quasi-Newton Approach. *Math. of Op. Res.*, 4:291–302, 1979.
- [17] R. Fletcher. A Class of Methods for Nonlinear Programming with Termination and Convergence Properties. In J. Abadie, editor, *Integer and Nonlinear Programming*. North Holland, 1970.
- [18] G. Di Pillo and L. Grippo. A Continuously Differentiable Exact Penalty Function for Nonlinear Programming Problem with Inequality Constraints. *SIAM J. on Contr. and Opt.*, 23:72–84.
- [19] G. Di Pillo and L. Grippo. A New Class of Augmented Lagrangians in Nonlinear Programming. *SIAM J. on Contr. and Opt.*, 17:618–628, 1979.
- [20] G. Di Pillo and L. Grippo. A New Augmented Lagrangian Function for Inequality Constraints in Nonlinear Programming. *JOTA*, 36:495–519, 1982.
- [21] D. P. Bertsekas. Enlarging the Region of Convergence of Newton’s Method for Constrained Optimization. *JPTA*, 36:221–252, 1982.
- [22] L. Grippo G. Di Pillo and L. Lampariello. A Method for Solving Equality Constrained Optimization Problems by Unconstrained Minimization. In K. Malanowski K. Iracki and S. Walukiewicz, editors, *Optimization Techniques - 9 th IFIP Conf.*, pages 167–182. Springer Verlag, 1980.
- [23] L. Grippo G. Di Pillo and L. Lampariello. A Class of Algorithms for the Solution of Optimization Problems with Inequalities. In R. F. Drenick and F. Kozin, editors, *System Modelling and Optimization - 10 th IFIP Conf.* Springer Verlag, 1982.
- [24] L. C. W. Dixon. Exact Penalty NFunction Methods in Nonlinear Programming. Tech. Rep. n.103/1979, NOC, The Hatfield Polytechnic, Evanston, IL, USA, 1979.
- [25] R. B. Wilson. *A Simplicial Algorithm for Concave Programming*. PhD thesis, Harvard University, Harvard, 1963.
- [26] S. P. Han. A Globally Convergent Method for Nonlinear Programming. *JOTA*, 22:297–309, 1977.
- [27] R. Fletcher. Numerical Experiments with an Exact  $L_1$  Penalty Function Method. In S. M. Robinson eds R. R. Meyer, editor, *Nonlinear Programming 4*. Academic Press, 1981.

- [28] M. J. D. Powell. The Convergence of Variable Metric Methods for Non-linearly Constrained Optimization Calculations. In S. M. Robinson eds R. R. Meyer, editor, *Nonlinear Programming 3*. Academic Press, 1978.
- [29] W. Murray. An algorithm for constrained minimization. In R. Fletcher eds, editor, *Optimization*. Academic Press, 1969.
- [30] M. C. Biggs. Constrained minimization using recursive equality quadratic programming. In F. A. Lootsma Ed., editor, *Numerical Methods for Non-linear Optimization*. Academic Press, 1972.
- [31] M. Biggs. On The Convergence of Some Constrained Minimization Algorithms Based on Recursive Quadratic Programming. *J. Inst. Maths. Applic.*, 21:67–81, 1978.
- [32] K. Schittkowski. The Current State of Constrained Optimization Software. In M. J. D. Powell Ed., editor, *Nonlinear Optimization*. Academic Press, 1982.
- [33] K. Schittkowski. *Nonlinear Programming Codes*. Springer Verlag, 1980.
- [34] A. V. Balakrishnan. On a New Computing Technique in Optimal Control. *SIAM J. on Contr. and Opt. Series A*, 6:149–173, 1968.
- [35] F. Lampariello G. Di Pillo, L. Grippo. The Multiplier Method for Optimal Control Problems. 5:133–157, 1974.
- [36] F. Lampariello G. Di Pillo, L. Grippo. A Newton-type Computing Technique for Optimal Control Problems. Tech. Rep. n.33/1982, IASI/CNR, Roma, 1979.
- [37] F. Lampariello G. Di Pillo, L. Grippo. A Class of Structured Quasi-Newton Algorithms for Optimal Control Problems. In H. E. Rauch Ed., editor, *Control Applications of NLP and Optimization*. Pergamon Press, 1984.
- [38] M. C. Bartholomew-Biggs. The Optimization of Spacecraft Orbital Manoeuvres. In L. C. W. Dixon and G. P. Szego editors, *Numerical Optimization of Dynamic Systems*. 1980.