

METODI DI OTTIMIZZAZIONE VINCOLATA

Giampaolo Liuzzi

April 10, 2019

1 Introduzione

In questa dispensa ci occuperemo di alcune classi di metodi per la soluzione di problemi di ottimizzazione vincolata nonlineare. In particolare, verranno esaminati alcuni dei metodi basati sull'uso di funzioni di merito *sequenziali* o *esatte* ovvero che si basano sul tentativo di trasformare il problema vincolato di partenza in una successione di problemi non vincolati o in un unico problema non vincolato. A tal scopo consideriamo il problema

$$\min_{x \in \mathcal{F}} f(x), \quad (1)$$

ove \mathcal{F} è un sottoinsieme chiuso di \mathbb{R}^n , detto *insieme ammissibile*, e $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Molto spesso l'insieme \mathcal{F} è definito mediante equazioni e disequazioni nonlineari ovvero $\mathcal{F} = \{x \in \mathbb{R}^n : h(x) = 0, g(x) \leq 0\}$ con $h : \mathbb{R}^n \rightarrow \mathbb{R}^q$ e $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$. In questo caso il problema (1) diviene:

$$\begin{aligned} \min f(x) \\ h(x) = 0 \\ g(x) \leq 0. \end{aligned} \quad (2)$$

Nel seguito supporremo f, g, h almeno due volte continuamente differenziabili. La funzione Lagrangiana associata al problema (2) è la funzione $L : \mathbb{R}^n \times \mathbb{R}^q \times \mathbb{R}^m \rightarrow \mathbb{R}$ data da

$$L(x, \mu, \lambda) = f(x) + \mu^\top h(x) + \lambda^\top g(x). \quad (3)$$

Diciamo che $(\bar{x}, \bar{\mu}, \bar{\lambda})$ è una tripla di Kuhn-Tucker (KT) per il problema (2) se soddisfa le seguenti condizioni:

$$\nabla_x L(\bar{x}, \bar{\mu}, \bar{\lambda}) = 0, \quad h(\bar{x}) = 0, \quad g(\bar{x}) \leq 0, \quad \bar{\lambda} \geq 0, \quad \bar{\lambda}^\top g(\bar{x}) = 0, \quad (4)$$

Considerato che $\max\{y, -z\} = 0$ se e solo se $y \leq 0$, $z \geq 0$ e $yz = 0$, le condizioni (4) possono essere equivalentemente riscritte come

$$\nabla_x L(\bar{x}, \bar{\mu}, \bar{\lambda}) = 0, \quad h(\bar{x}) = 0, \quad \max\{g(\bar{x}), -\bar{\lambda}\} = 0.$$

Dato un qualunque punto $x \in \mathfrak{R}^n$, indichiamo con

$$\begin{aligned} I_o(x) &= \{i \in \{1, \dots, m\} : g_i(x) = 0\}, \\ I_+(x) &= \{i \in \{1, \dots, m\} : g_i(x) \geq 0\} \end{aligned}$$

gli insiemi degli indici dei vincoli di disuguaglianza attivi in x e, rispettivamente, l'insieme degli indici dei vincoli di disuguaglianza attivi o violati in x .

(LICQ)

Linear Independence Constraint Qualification

Un punto $x \in \mathfrak{R}^n$ soddisfa la (LICQ) se non è possibile trovare dei numeri v_j , $j = 1, \dots, q$ e u_i , $i \in I_o(x)$ tali che

$$\begin{aligned} \sum_{i \in I_o(x)} u_i \nabla g_i(x) + \sum_{j=1}^q v_j \nabla h_j(x) &= 0, \\ (u_i, i \in I_o(x), v_j, j = 1, \dots, q) &\neq 0. \end{aligned}$$

(SOSC)

Second Order Sufficient Condition

Una tripla di KT, (x^*, μ^*, λ^*) , per il problema (2) soddisfa la (SOSC) se

$$w^\top \nabla_x^2 L(x^*, \mu^*, \lambda^*) w > 0, \quad \forall w \neq 0, w \in Y(x^*, \lambda^*),$$

con

$$Y(x^*, \lambda^*) = \left\{ \begin{array}{l} w \in \mathfrak{R}^n : \nabla h_j(x^*)^\top w = 0, j = 1, \dots, m, \\ \nabla g_i(x^*)^\top w \leq 0, i \in I_o(x^*), \lambda_i^* = 0, \\ \nabla g_i(x^*)^\top w = 0, i \in I_o(x^*), \lambda_i^* > 0 \end{array} \right\}$$

Se $\bar{x} \in \mathcal{F}$ è un punto di minimo del problema (2) che soddisfa la (LICQ), allora esistono dei moltiplicatori $\bar{\mu} \in \mathfrak{R}^q$ e $\bar{\lambda} \in \mathfrak{R}^m$ tali che $(\bar{x}, \bar{\mu}, \bar{\lambda})$ è una tripla di KT per il problema (2).

Se (x^*, μ^*, λ^*) è una tripla di KT per il problema (2) che soddisfa la (SOSC) allora il punto x^* è un minimo locale stretto per il problema (2).

Dato un insieme $\mathcal{C} \subseteq \mathfrak{R}^n$, indichiamo con $\mathcal{G}(\mathcal{C})$ e con $\mathcal{L}(\mathcal{C})$ rispettivamente l'insieme dei punti di minimo globale e l'insieme dei punti di minimo locale del problema (2) su \mathcal{C} ovvero

$$\mathcal{G}(\mathcal{C}) = \{x \in \mathcal{C} \cap \mathcal{F} : f(x) \leq f(y), \forall y \in \mathcal{C} \cap \mathcal{F}\},$$

$$\mathcal{L}(\mathcal{C}) = \{x \in \mathcal{C} \cap \mathcal{F} : \exists \rho > 0, f(x) \leq f(y), \forall y \in \mathcal{C} \cap \mathcal{F} \cap \mathcal{B}(x; \rho)\},$$

ove,

$$\mathcal{B}(x; \rho) = \{y \in \mathfrak{R}^n : \|y - x\| \leq \rho\}.$$

Di tanto in tanto e per semplicità ci riferiremo a problemi che hanno vincoli o solo di uguaglianza o solo di disuguaglianza ovvero

$$\begin{aligned} \min f(x) \\ h(x) = 0. \end{aligned} \quad (5)$$

$$\begin{aligned} \min f(x) \\ g(x) \leq 0. \end{aligned} \quad (6)$$

È conveniente a questo punto notare che il problema (6) può essere ricondotto ad un problema di tipo (5) mediante l'introduzione di m variabili ausiliarie y_i , $i = 1, \dots, m$ come segue

$$\begin{aligned} \min f(x) \\ g(x) + Yy = 0, \end{aligned} \quad (7)$$

ove $Y = \text{diag}(y_i)_{i=1, \dots, m}$. Prima di entrare nel merito dei metodi per la soluzione del problema (1), è conveniente considerare la seguente funzione $\varphi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$

$$\varphi(y, z) = z \max\{y, -z\} + \frac{1}{2} \max\{y, -z\}^2, \quad (8)$$

che, come vedremo, comparirà nell'espressione di tutte le *funzioni di merito* considerate in questi appunti. La funzione $\varphi(y, z)$ gode di alcune proprietà estremamente importanti che sono riassunte nella proposizione seguente ([3]).

Proposizione 1 *Sia φ la funzione definita in (8) allora:*

(a) $\varphi(y, z)$ è una funzione continuamente differenziabile per ogni $(y, z) \in \mathbb{R} \times \mathbb{R}$, con derivate

$$\frac{\partial \varphi(y, z)}{\partial y} = z + \max\{y, -z\}, \quad (9)$$

$$\frac{\partial \varphi(y, z)}{\partial z} = \max\{y, -z\}; \quad (10)$$

(b) se $z = 0$ allora si ha che $\varphi(y, 0) = \max\{y, 0\}^2/2$ per ogni $y \in \mathbb{R}$;

(c) se $y \leq 0$, $z \geq 0$ e $yz = 0$ allora $\varphi(y, z) = 0$;

(d) se $y \leq 0$ allora $\varphi(y, z) \leq 0$ per ogni $z \in \mathbb{R}$;

(e) se $y \leq 0$ e $\varphi(y, z) = 0$ allora $y \leq 0$, $z \geq 0$ e $yz = 0$.

Dim. I punti (a) e (b) discendono direttamente dalla definizione di $\varphi(y, z)$. Il punto (c) deriva dall'espressione di $\varphi(y, z)$ e dalle proprietà della funzione $\max\{y, -z\}$ che si annulla se e solo se $y \leq 0$, $z \geq 0$ e $yz = 0$.

Consideriamo ora il punto (d) e supponiamo quindi che $y \leq 0$. Sono possibili due casi: nel primo caso, $\max\{y, -z\} = y \leq 0$, risulta $\varphi(y, z) = y(z + y/2)$. Dato che $y \geq -z$,

segue che $z + y \geq 0$ e quindi a maggior ragione (dato che $y \leq 0$) $z + y/2 \geq 0$. Pertanto $\varphi(y, z) \leq 0$. Nel secondo caso, $\max\{y, -z\} = -z$, risulta $\varphi(y, z) = -z^2/2 \leq 0$. Per finire consideriamo il punto (e) e ammettiamo quindi che

$$\varphi(y, z) = \max\{y, -z\} \left(z + \frac{1}{2} \max\{y, -z\} \right) = 0,$$

che implica che vale almeno una delle due seguenti relazioni

$$\max\{y, -z\} = 0, \quad (11)$$

$$\left(z + \frac{1}{2} \max\{y, -z\} \right) = 0. \quad (12)$$

Se vale la (11), la prova segue banalmente dalle proprietà del $\max\{y, -z\}$. Supponiamo quindi che valga la (12). In questo caso abbiamo intanto che

$$-z \leq \max\{y, -z\} = -2z, \quad (13)$$

da cui segue intanto che $z \leq 0$. Se $z = 0$, ritorniamo al caso in cui vale la (11). Viceversa, supponiamo $z < 0$. Dato che, per ipotesi, $y \leq 0$, abbiamo che $\max\{y, -z\} = -z$ che con la (13) implica $z = 0$, il che contraddice l'assunzione $z < 0$. \square

2 Metodi con funzioni di Penalità sequenziali esterne

Si consideri il Problema (1) e la funzione

$$P_{\mathcal{F}}(x) = \begin{cases} f(x) & \text{se } x \in \mathcal{F} \\ +\infty & \text{altrimenti.} \end{cases} \quad (14)$$

Ovviamente, le soluzioni del problema vincolato (1) coincidono con i punti di minimo del problema non vincolato

$$\min_{x \in \mathbb{R}^n} P_{\mathcal{F}}(x).$$

Tuttavia la funzione $P_{\mathcal{F}}(x)$ non è continua sulla frontiera dell'insieme ammissibile \mathcal{F} del problema originario. Una sua minimizzazione non vincolata è quindi, generalmente, molto difficile. La gran parte dei metodi per la minimizzazione non vincolata richiedono, infatti, almeno la continuità della funzione obiettivo. Allo scopo di porre parzialmente rimedio alla precedente difficoltà, è possibile considerare una successione di funzioni continue che approssimano la $P_{\mathcal{F}}(x)$ al limite. In particolare, sia $\phi_S(x)$ una funzione con le seguenti proprietà:

- (i) $\phi_S(x)$ è una funzione continuamente differenziabile,
- (ii) $\phi_S(x) = 0$ se $x \in \mathcal{F}$,
- (iii) $\phi_S(x) > 0$ se $x \notin \mathcal{F}$.

Un esempio di funzione $\phi_S(x)$ per il problema (2) è il seguente:

$$\begin{aligned}\phi_S(x) &= \sum_{i=1}^m \varphi(g_i(x), 0) + \sum_{j=1}^q h_j(x)^2 \\ &= \sum_{i=1}^m \max\{g_i(x), 0\}^2 + \sum_{j=1}^q h_j(x)^2,\end{aligned}\tag{15}$$

ove $\varphi(y, z)$ è la funzione definita in (8). In virtù delle proprietà della funzione $\varphi(y, z)$ e, in particolare, visti i punti (a) e (b) di Proposizione 1, è possibile dimostrare che la funzione $\phi_S(x)$ soddisfa le proprietà (i)–(iii) viste precedentemente.

È immediato verificare che la funzione $P(x; \epsilon)$ così definita ([4],[5])

$$P(x; \epsilon) = f(x) + \frac{1}{\epsilon} \phi_S(x),\tag{16}$$

per valori di $\epsilon \downarrow 0$, approssima la (14) cioè, per ogni $x \in \mathfrak{R}^n$ fissato,

$$\lim_{\epsilon \downarrow 0} P(x; \epsilon) = P_{\mathcal{F}}(x).$$

La funzione $P(x; \epsilon)$ è detta funzione di penalità esterna per il problema (2). È possibile mettere subito in evidenza un primo importante aspetto della funzione di penalità esterna $P(x; \epsilon)$ e cioè il fatto di essere “*non esatta*” ovvero non esiste nessun valore del parametro di penalità ϵ per cui è possibile dimostrare che il problema vincolato di partenza è equivalente alla minimizzazione non vincolata della funzione $P(x; \epsilon)$.

Una possibile spiegazione intuitiva, ma certamente efficace, di questo aspetto può essere ottenuta osservando che il richiedere alla funzione $\phi_S(x)$ le proprietà (i)–(iii) equivale ad imporre che in tutti i punti ammissibili si abbia $\nabla \phi_S(x) = 0$ e, quindi, che

$$\nabla P(x; \epsilon) = \nabla f(x), \quad \forall x \in \mathcal{F}.$$

Tuttavia, in un punto di minimo vincolato x^* del problema (2), si ha generalmente che $\nabla f(x^*) \neq 0$ da cui segue che $\nabla P(x^*; \epsilon) \neq 0$ per ogni valore del parametro ϵ . Questo fatto implica che, qualunque sia il parametro di penalità ϵ , il minimo x^* del problema vincolato non può essere un minimo non vincolato di $P(x; \epsilon)$. Un altro importante aspetto della funzione $P(x; \epsilon)$ è che, per ogni valore del parametro di penalità ϵ , una sua minimizzazione non vincolata produce in genere un punto $\bar{x} \notin \mathcal{F}$ cioè tale per cui $\phi_S(\bar{x}) > 0$.

Uno schema di algoritmo per la soluzione di un problema vincolato mediante l'uso di funzioni di penalità esterne è il seguente.

Algoritmo SEQPEN**Dati:** $\{\epsilon_k\} \downarrow 0, \tau > 0$ **Step 0:** Poni $k = 0$.**Step 1:** Trova una soluzione $(x(\epsilon_k))$ del problema

$$\min_{x \in \mathfrak{R}^n} P(x; \epsilon_k). \quad (17)$$

Step 2: Se $\phi_S(x(\epsilon_k)) \leq \tau$ allora STOP,
altrimenti poni $k = k + 1$ e torna allo Step 1.

Un primo risultato teorico riguardo le proprietà di convergenza dell'algoritmo SEQPEN, quando $\tau = 0$, si basa sull'assunzione che allo Step 1 sia possibile determinare una soluzione globale del problema (17).

Proposizione 2 ([6]) *Sia $f(x)$ limitata inferiormente sull'insieme ammissibile \mathcal{F} ovvero esista $f^* = \inf_{x \in \mathcal{F}} f(x)$. Sia, inoltre $\{\epsilon_k\}$ una successione monotona non crescente e convergente a 0 e, per ogni k , $P(x; \epsilon_k)$ ammetta un punto di minimo globale $x(\epsilon_k)$. Allora:*

- (i) *la successione $\{P(x(\epsilon_k); \epsilon_k)\}$ è monotona non decrescente;*
- (ii) *la successione $\{\phi_S(x(\epsilon_k))\}$ è monotona non crescente;*
- (iii) *la successione $\{f(x(\epsilon_k))\}$ è monotona non decrescente.*

Inoltre, $\lim_{k \rightarrow \infty} \phi_S(x(\epsilon_k)) = 0$ e ogni punto limite x^ di $\{x(\epsilon_k)\}$ è un punto di minimo globale del problema (2).*

Dim. Consideriamo due interi consecutivi k e $k + 1$. Dalla ipotesi $\epsilon_k \geq \epsilon_{k+1}$ e dal fatto che il termine di penalità $\phi_S(x)$ è sempre positivo o nullo segue che

$$P(x; \epsilon_k) \leq P(x; \epsilon_{k+1}) \quad (18)$$

per ogni $x \in \mathfrak{R}^n$.

Siccome $x(\epsilon_k)$ è, per ipotesi, un punto di minimo globale di $P(x; \epsilon_k)$, risulta banalmente

$$P(x(\epsilon_k); \epsilon_k) \leq P(x(\epsilon_{k+1}); \epsilon_k), \quad (19)$$

$$P(x(\epsilon_{k+1}); \epsilon_{k+1}) \leq P(x(\epsilon_k); \epsilon_{k+1}). \quad (20)$$

Dalle (18) e (19) segue

$$P(x(\epsilon_k); \epsilon_k) \leq P(x(\epsilon_{k+1}); \epsilon_{k+1}) \quad (21)$$

che dimostra il punto (i). Dalle (19) e (20) seguono, rispettivamente

$$\frac{1}{\epsilon_k} (\phi_S(x(\epsilon_k)) - \phi_S(x(\epsilon_{k+1}))) \leq f(x(\epsilon_{k+1})) - f(x(\epsilon_k)), \quad (22)$$

$$-\frac{1}{\epsilon_{k+1}}(\phi_S(x(\epsilon_k)) - \phi_S(x(\epsilon_{k+1}))) \leq f(x(\epsilon_k)) - f(x(\epsilon_{k+1})). \quad (23)$$

Sommando le (22) e (23) otteniamo

$$\left(\frac{1}{\epsilon_k} - \frac{1}{\epsilon_{k+1}}\right)(\phi_S(x(\epsilon_k)) - \phi_S(x(\epsilon_{k+1}))) \leq 0,$$

da cui segue il punto (ii). Infine, dal punto (ii) e dalla (22) segue anche il punto (iii). Ricordando la definizione (15) del termine di penalità $\phi_S(x)$, segue che

$$P(x(\epsilon_k); \epsilon_k) \leq \inf_{x \in \mathcal{F}} P(x; \epsilon_k) = \inf_{x \in \mathcal{F}} f(x) = f^*. \quad (24)$$

La successione $\{P(x(\epsilon_k); \epsilon_k)\}$ è pertanto monotona non decrescente (dalla proprietà (i)) e limitata superiormente e quindi è convergente cioè risulta $\lim_{k \rightarrow \infty} P(x(\epsilon_k); \epsilon_k) = \bar{P}$. Siccome inoltre $\epsilon_k \downarrow 0$, segue che $\lim_{k \rightarrow \infty} \phi_S(x(\epsilon_k)) = 0$.

Sia ora x^* un qualunque punto limite della successione $\{x(\epsilon_k)\}$ ovvero

$$\lim_{k \rightarrow \infty, k \in K} x(\epsilon_k) = x^*,$$

allora, per continuità, $\phi_S(x(\epsilon_k)) \rightarrow_K \phi_S(x^*) = 0$ ovvero $x^* \in \mathcal{F}$. Ma se x^* è un punto ammissibile allora

$$f(x^*) \geq \inf_{x \in \mathcal{F}} f(x) = f^*. \quad (25)$$

D'altra parte, dalla (24) abbiamo che $f(x^*) \leq f^*$ e quindi considerando la (25) otteniamo che $f(x^*) = f^*$ ovvero x^* è una soluzione globale del problema (2). \square

Esempio 1 Si consideri il seguente problema non lineare

$$\begin{aligned} \min \quad & x^2 \\ & x = 1, \end{aligned}$$

che ammette, ovviamente, $x^* = 1$ come unica soluzione che è quindi anche l'unico punto di minimo globale del problema.

La funzione di Penalità esterna per questo problema è la seguente.

$$P(x; \epsilon) = x^2 + \frac{1}{\epsilon}(x - 1)^2.$$

In Figura 1 sono riportati gli andamenti della $P(x; \epsilon)$ per valori decrescenti del parametro di penalità ϵ . Dalla medesima figura si vede come i punti di minimo globale della penalità esterna (tutti inammissibili) tendano alla soluzione del problema per $\epsilon \downarrow 0$.

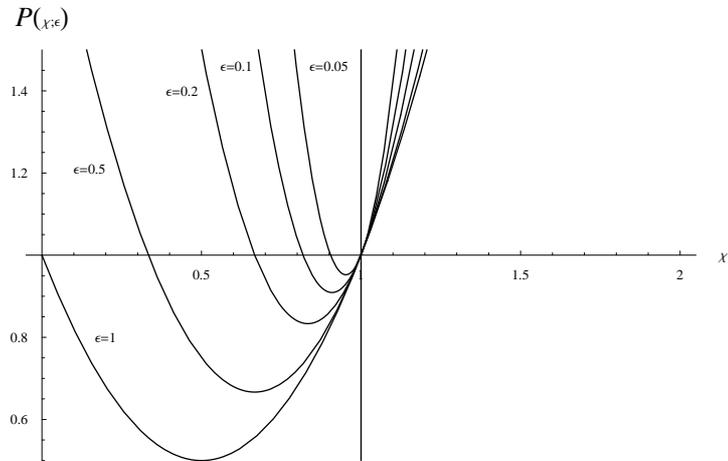


Figure 1: Grafici di $P(x; \epsilon)$ ($\epsilon = 1, 0.5, 0.2, 0.1, 0.05$) per il problema dell'Esempio 1.

Esempio 2 Si consideri il seguente problema non lineare

$$\begin{aligned} \min \quad & x^2 \\ & x \geq 1, \end{aligned}$$

che ammette, nuovamente, $x^* = 1$ come unico punto di minimo globale. La funzione di Penalità esterna per questo problema è la seguente.

$$P(x; \epsilon) = x^2 + \frac{1}{\epsilon} \max\{1 - x, 0\}^2.$$

Notiamo che, per $x \notin \mathcal{F}$, $P(x; \epsilon)$ coincide con quella dell'esempio 1, mentre per punti $x \in \mathcal{F}$, $P(x; \epsilon) = f(x)$.

In Figura 2 sono riportati gli andamenti della $P(x; \epsilon)$ per valori decrescenti del parametro di penalità ϵ . Valgono le medesime osservazioni dell'esempio 1.

La Proposizione 2 si basa sull'ipotesi, molto difficile da soddisfare in pratica, che per ogni valore del parametro di penalità ϵ_k , $x(\epsilon_k)$ sia una soluzione globale "esatta" del problema non vincolato (17). Tuttavia, come mostrato dall'esempio che segue, determinare l'ottimo globale della funzione di penalità $P(x; \epsilon)$ può essere molto difficile se non del tutto impossibile (e.g. quando la $P(x; \epsilon)$ è illimitata inferiormente).

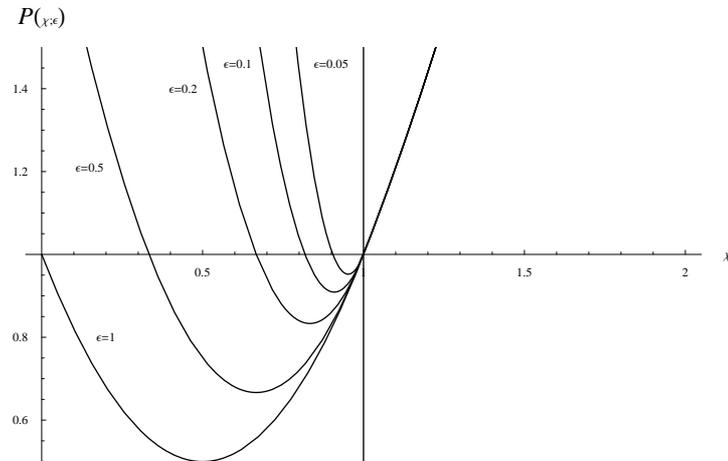


Figure 2: Grafici di $P(x; \epsilon)$ ($\epsilon = 1, 0.5, 0.2, 0.1, 0.05$) per il problema dell'Esempio 2.

Esempio 3 Si consideri il seguente problema non lineare

$$\begin{aligned} \min \quad & -x^4 \\ & x = 1, \end{aligned}$$

che ammette, naturalmente, $x^* = 1$ come unico punto di minimo globale. La funzione di Penalità esterna per questo problema è la seguente.

$$P(x; \epsilon) = -x^4 + \frac{1}{\epsilon}x^2,$$

che, come si vede dalla Figura 3, è illimitata inferiormente su \mathfrak{R}^n . In figura sono riportati gli andamenti della $P(x; \epsilon)$ per valori decrescenti del parametro di penalità ϵ . Notiamo che per alcuni valori del parametro di penalità ϵ la funzione $P(x; \epsilon)$ non ammette nessun punto di minimo locale.

Oltre alla precedente difficoltà, è necessario considerare il fatto che la maggior parte degli algoritmi per la minimizzazione non vincolata sono in grado di certificare la stazionarietà di un punto ovvero di determinare punti che soddisfano le condizioni necessarie di ottimalità del primo ordine.

A tal proposito, consideriamo il problema con soli vincoli di uguaglianza (5). In questo caso il termine di penalità si semplifica in $\phi_S(x) = \|h(x)\|^2$ e quindi

$$P(x; \epsilon) = f(x) + \frac{\|h(x)\|^2}{\epsilon}.$$

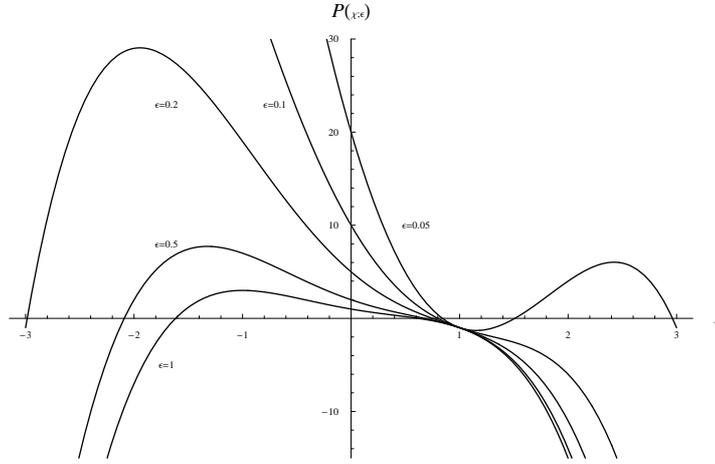


Figure 3: Grafici di $P(x; \epsilon)$ ($\epsilon = 1, 0.5, 0.2, 0.1, 0.05$) per il problema dell'Esempio 3.

Proposizione 3 ([7]) Siano $\{\epsilon_k\}$ una successione monotona non crescente e convergente a 0, $\{\tau_k\}$ una successione di numeri non negativi e convergenti a 0 e sia $\{x_k\}$ una successione di vettori tali che

$$\|\nabla_x P(x_k; \epsilon_k)\| \leq \tau_k. \quad (26)$$

Sia $\{x_k\}_K$ una sottosuccessione convergente ad un vettore x^* tale che $\nabla h(x^*)$ ha rango q . Allora,

$$\nabla_x L(x^*, \mu^*) = 0, \quad h(x^*) = 0,$$

dove μ^* è tale che

$$\left\{ \frac{1}{\epsilon_k} h(x_k) \right\}_K \rightarrow \mu^*.$$

Dim. Senza perdita di generalità possiamo assumere che l'intera successione $\{x_k\}$ converga a x^* . Per ogni k definiamo

$$\mu_k = \frac{1}{\epsilon_k} h(x_k).$$

Derivando rispetto ad x l'espressione della funzione $P(x; \epsilon)$ otteniamo

$$\nabla_x P(x_k; \epsilon_k) = \nabla f(x_k) + \frac{1}{\epsilon_k} \nabla h(x_k) h(x_k) = \nabla f(x_k) + \nabla h(x_k) \mu_k. \quad (27)$$

Dal momento che $\nabla h(x^*)$ ha rango q , per continuità segue che per valori di k sufficientemente elevati anche $\nabla h(x_k)$ ha rango q . Senza perdita di generalità, assumiamo che

$\nabla h(x_k)$ ha rango q per ogni k e quindi che $\nabla h(x_k)^\top \nabla h(x_k)$ è invertibile. Premoltiplicando l'equazione (27) per

$$(\nabla h(x_k)^\top \nabla h(x_k))^{-1} \nabla h(x_k)^\top,$$

otteniamo

$$\mu_k = (\nabla h(x_k)^\top \nabla h(x_k))^{-1} \nabla h(x_k)^\top (\nabla_x P(x_k; \epsilon_k) - \nabla f(x_k)). \quad (28)$$

Dalle ipotesi della proposizione segue che $\nabla_x P(x_k; \epsilon_k) \rightarrow 0$ e quindi dalla (28) otteniamo che $\mu_k \rightarrow \mu^*$ essendo

$$\mu^* = -(\nabla h(x^*)^\top \nabla h(x^*))^{-1} \nabla h(x^*)^\top \nabla f(x^*). \quad (29)$$

Usando nuovamente il fatto che $\nabla_x P(x_k; \epsilon_k) \rightarrow 0$ con la (27), abbiamo che

$$\nabla f(x^*) + \nabla h(x^*) \mu^* = 0.$$

Siccome $\mu_k \rightarrow \mu^*$, $\mu_k = (1/\epsilon_k)h(x_k)$ e $\epsilon_k \downarrow 0$, segue che $h(x_k) \rightarrow 0$ che, per continuità implica che $h(x^*) = 0$ che dimostra la proposizione. \square

Consideriamo ora il problema con soli vincoli di disuguaglianza (6). Come visto all'inizio di questa sezione, una funzione di penalità esterna in questo caso è

$$P(x; \epsilon) = f(x) + \frac{1}{\epsilon} \sum_{i=1}^m \max\{g_i(x), 0\}^2.$$

Notiamo che

$$P(x; \epsilon) = \min_y \bar{P}(x, y; \epsilon) \quad (30)$$

ove $\bar{P}(x, y; \epsilon)$ è la funzione di penalità esterna per il problema equivalente (7). Essendo il problema (7) un problema con soli vincoli di uguaglianza, la funzione $\bar{P}(x, y; \epsilon)$ ha la seguente espressione

$$\bar{P}(x, y; \epsilon) = f(x) + \frac{1}{\epsilon} \|g(x) + Yy\|^2.$$

La funzione $\bar{P}(x, y; \epsilon)$ risulta dunque quadratica e convessa nelle variabili y pertanto è possibile determinare analiticamente, fissate le x , il valore delle variabili y che minimizzano $\bar{P}(x, y; \epsilon)$. Data la convessità della funzione $\bar{P}(x, y; \epsilon)$, y^* risolve il problema (30) se e solo se

$$\nabla_y \bar{P}(x, y^*; \epsilon) = 2Y^*(g(x) + Yy^*) = 0,$$

da cui risulta che y^* è soluzione del problema (30) se e solo se

$$(y_i^*)^2 = \max\{0, -g_i(x)\}, \quad i = 1, \dots, m.$$

Possiamo quindi scrivere che

$$P(x; \epsilon) = \bar{P}(x, y^*; \epsilon),$$

e dimostrare la seguente proprietà (analoga della Proposizione 3)

Proposizione 4 *Siano $\{\epsilon_k\}$ una successione monotona non crescente e convergente a 0, $\{\tau_k\}$ una successione di numeri non negativi e convergenti a 0 e sia $\{x_k\}$ una successione di vettori tali che*

$$\|\nabla_x P(x_k; \epsilon_k)\| \leq \tau_k. \quad (31)$$

Sia $\{x_k\}_K$ una sottosuccessione convergente ad un vettore x^ tale che $\nabla g_i(x^*)$, $i \in I_0(x^*)$ ha rango pieno. Allora,*

$$\nabla_x L(x^*, \lambda^*) = 0, \quad g(x^*)^\top \lambda^* = 0, \quad g(x^*) \leq 0,$$

dove λ^ è tale che*

$$\left\{ \frac{1}{\epsilon_k} \max\{g(x_k), 0\} \right\}_K \rightarrow \lambda^* \geq 0.$$

Dim. La Proposizione 3 consente di stabilire un risultato di convergenza per l'algoritmo SEQPEN quando sia applicato alla soluzione del problema (7). In particolare, essendo per ipotesi x_k tale che

$$\|\nabla_x P(x_k; \epsilon_k)\| \leq \tau_k,$$

se poniamo $y_{ik}^2 = \max\{0, -g_i(x_k)\}$ abbiamo certamente che

$$\|\nabla_{x,y} \bar{P}(x_k, y_k; \epsilon_k)\| \leq \tau_k$$

risultando

$$\nabla_y \bar{P}(x_k, y_k; \epsilon_k) = 0 \quad \text{e} \quad \nabla_x \bar{P}(x_k, y_k; \epsilon_k) = \nabla_x P(x_k; \epsilon_k).$$

Dal fatto che $\{x_k\}_K \rightarrow x^*$ e per continuità, segue che $\{y_{ik}^2\}_K \rightarrow \max\{0, -g_i(x^*)\} = (y_i^*)^2$.

Facciamo ora vedere che, nelle ipotesi poste, la matrice

$$\begin{bmatrix} \nabla g(x^*) \\ 2Y^* \end{bmatrix} \quad (32)$$

ha rango pieno. A tal fine, notiamo che la matrice (32) può essere riscritta, riordinando se necessario le colonne, come

$$\begin{bmatrix} \nabla g(x^*)_{I_0(x^*)} & \nabla g(x^*)_{\bar{I}_0(x^*)} \\ 0 & 2Y_{I_0(x^*)}^* \end{bmatrix},$$

essendo $\bar{I}_0(x^*) = 1, \dots, m \setminus I_0(x^*)$. Quindi, se la matrice $\nabla g(x^*)_{I_0(x^*)}$ ha rango pieno, ovvero sono linearmente indipendenti i gradienti dei vincoli attivi in x^* , anche la matrice (32) ha rango pieno.

Quindi, valgono le ipotesi della Proposizione 3. Pertanto, possiamo concludere esiste un vettore λ^* tale che

$$\begin{bmatrix} \nabla f(x^*) \\ 0 \end{bmatrix} + \begin{bmatrix} \nabla g(x^*) \\ 2Y^* \end{bmatrix} \lambda^* = 0, \quad g(x^*) + Y^* y^* = 0 \quad (33)$$

con λ^* è tale che

$$\left\{ \frac{1}{\epsilon_k} (g(x_k) + Y_k y_k) \right\}_K \rightarrow \lambda^*. \quad (34)$$

Dalla (33) segue intanto che

$$\nabla f(x^*) + \nabla g(x^*) \lambda^* = 0, \quad (35)$$

$$y_i^* \lambda_i^* = 0. \quad (36)$$

La (36) insieme alla (33) danno

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

ovvero le condizioni di complementarità. Infine, notiamo che dalla (34) considerato che $y_{ik}^2 = \max\{0, -g_i(x_k)\}$, segue che

$$\left\{ \frac{1}{\epsilon_k} \max\{g(x_k), 0\} \right\}_K \rightarrow \lambda^*.$$

Inoltre, siccome ogni elemento della successione $\{(1/\epsilon_k) \max\{g(x_k), 0\}\}$ è positivo o nullo, segue anche che

$$\lambda^* \geq 0,$$

il che conclude la prova. \square

Dall'analisi condotta fin'ora, sembrerebbe emergere che l'algoritmo **SEQPEN**, a meno di alcune situazioni patologiche, è un algoritmo tutto sommato robusto, ammesso che per ogni indice k al passo 3 si è in grado di determinare una soluzione del problema (17) (anche non esatta ma con una precisione sempre maggiore). Purtroppo questa conclusione è falsa. Infatti, trovare una soluzione del problema (17) al passo 3 dell'algoritmo **SEQPEN** diventa sempre più difficile mano a mano che il parametro di penalità $\epsilon_k \downarrow 0$. Questa crescente complessità è dovuta essenzialmente al crescente mal condizionamento della funzione di penalità come mostrato dall'esempio che segue.

Esempio 4 Si consideri il seguente problema non lineare

$$\min \quad -x - y \\ x^2 + y^2 = 1.$$

Questo problema ammette $(x^*, y^*) = (\sqrt{2}/2, \sqrt{2}/2)$ come unico punto di minimo globale.

La funzione di Penalità esterna per questo problema è la seguente.

$$P(x, y; \epsilon) = -x - y + \frac{1}{\epsilon} (1 - x^2 - y^2)^2.$$

In Figura 4 sono riportate le curve di livello della $P(x, y; \epsilon)$ per $\epsilon = 1, 1/2, 1/20$. Valgono le medesime osservazioni dell'Esempio 1.

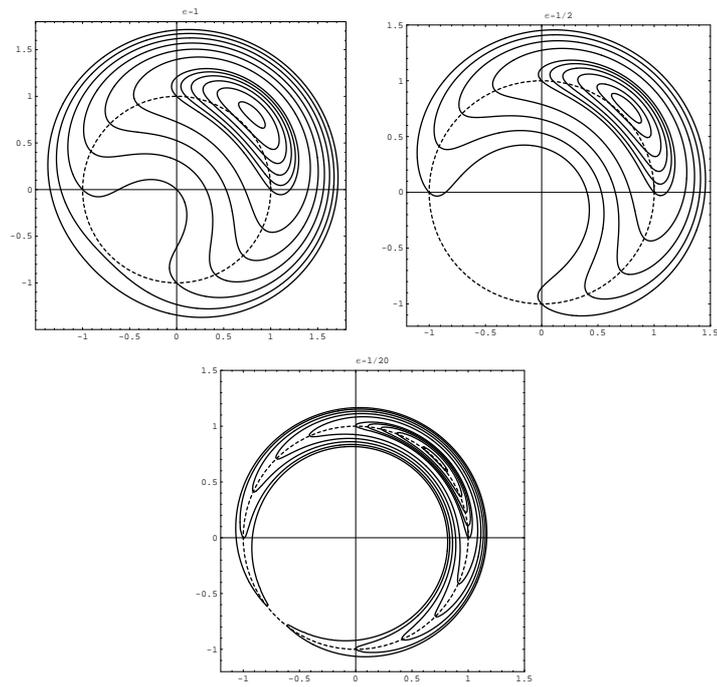


Figure 4: Curve di livello di $P(x; \epsilon)$ ($\epsilon = 1, 0.5, 0.05$) per il problema dell'Esempio 4. Il cerchio tratteggiato rappresenta il vincolo del problema.

Il problema del crescente mal condizionamento per $\epsilon_k \downarrow 0$ può essere almeno in parte risolto scegliendo, per ogni iterazione k , $x(\epsilon_{k-1})$ come punto iniziale da passare alla routine di minimizzazione non vincolata usata per risolvere il problema (17). Affinchè, però, il punto iniziale $x(\epsilon_{k-1})$ sia abbastanza vicino ad una soluzione di (17) in modo da ottenere un problema non vincolato relativamente facile da risolvere, è necessario che ϵ_k sia abbastanza vicino ad ϵ_{k-1} e questo vuol dire che la velocità con cui la successione $\{\epsilon_k\}$ tende a zero deve essere sufficientemente bassa. Una buona scelta della successione $\{\epsilon_k\}$ è, per esempio, $\epsilon_k = \epsilon_{k-1}/\beta$ con $\beta \in [2, 10]$.

Dati un punto iniziale \bar{x} , un valore del parametro di penalità ϵ ed una tolleranza τ , definiamo l'operatore $\mathcal{M}_P[\bar{x}, \epsilon, \tau]$ come quell'operatore che dato il punto iniziale \bar{x} fornisce un nuovo punto $x(\epsilon)$ tale che

$$\|\nabla P(x(\epsilon); \epsilon)\| \leq \tau.$$

In questo modo possiamo definire l'algoritmo $\text{SEQPEN}_{\text{mod}}$ come segue:

Algoritmo $\text{SEQPEN}_{\text{mod}}$

Dati: $\epsilon_0 > 0$, $\beta \in [2, 10]$, $\{\tau_k\} \rightarrow 0$, $\rho > 0$ e \bar{x}

Step 0: Poni $k = 0$ e $\bar{x} = \bar{x}$.

Step 1: Calcola $x(\epsilon_k) = \mathcal{M}_P[\bar{x}, \epsilon_k, \tau_k]$

Step 2: Se $\phi_S(x(\epsilon_k)) < \rho$ allora STOP,

altrimenti poni $\bar{x} = x(\epsilon_k)$, $\epsilon_{k+1} = \epsilon_k/\beta$, $k = k + 1$

e vai allo Step 1.

3 Metodi con funzioni Lagrangiane aumentate

Come abbiamo visto nel corso della precedente sezione, le funzioni di penalità esterna soffrono del fatto di essere “*non esatte*” ovvero non esiste nessun valore del parametro di penalità ϵ per cui è possibile dimostrare che il problema vincolato di partenza è equivalente alla minimizzazione non vincolata della funzione $P(x; \epsilon)$. Questa limitazione può essere imputata al fatto che il termine $\phi_S(x)$ non è in grado di offrire una penalizzazione sufficientemente esauriente della violazione delle condizioni di ottimalità. La funzione $\phi_S(x)$, in particolare, è in grado di pesare solo l'ammissibilità del punto corrente ma non la sua stazionarietà ovvero il soddisfacimento delle condizioni necessarie di ottimo (KT). Per rimediare a questa carenza teorica, è possibile ridefinire il termine di penalità come segue.

Sia $\phi_E(x, \mu, \lambda; \epsilon)$ una funzione della tripla (x, μ, λ) e dipendente dal parametro di penalità ϵ tale che ([8],[9],[10]):

- (i) $\phi_E(x, \mu, \lambda; \epsilon)$ è una funzione *continuamente differenziabile* nella tripla (x, μ, λ) ;

(ii) per ogni μ e λ si ha

$$\lim_{\epsilon \rightarrow 0} \phi_E(x, \mu, \lambda; \epsilon) = \phi_S(x);$$

(iii) se $x \in \mathcal{F}$, allora $\phi_E(x, \mu, \lambda; \epsilon) = 0$ se e solamente se $\lambda \geq 0$ e $g(x)^\top \lambda = 0$;

(iv) se (x^*, μ^*, λ^*) è una tripla di Kuhn-Tucker allora per ogni ϵ si ha:

$$\frac{1}{\epsilon} \nabla_x \phi_E(x^*, \mu^*, \lambda^*; \epsilon) = -\nabla f(x^*).$$

In particolare, la funzione $\phi_E(x, \mu, \lambda; \epsilon)$ proposta da Rockafellar è la seguente:

$$\begin{aligned} \phi_E(x, \mu, \lambda; \epsilon) &= \sum_{i=1}^m \varphi(g_i(x), \epsilon \lambda_i) + \sum_{j=1}^q \left(\epsilon \mu_j h_j(x) + \frac{1}{2} h_j(x)^2 \right) \quad (37) \\ &= \epsilon \lambda^\top \max\{g(x), -\epsilon \lambda\} + \frac{1}{2} \|\max\{g(x), -\epsilon \lambda\}\|^2 \\ &\quad + \epsilon \mu^\top h(x) + \frac{1}{2} \|h(x)\|^2, \end{aligned}$$

ove la funzione $\varphi(y, z)$ è definita in (8). In virtù delle proprietà della funzione φ ed in particolare visti i punti (a), (b) (c) ed (e) di Proposizione 1, è possibile dimostrare che la funzione $\phi_E(x, \mu, \lambda; \epsilon)$ soddisfa le proprietà (i)–(iv) di sopra.

Utilizzando il termine ϕ_E è possibile definire la funzione $L_a : \mathbb{R}^n \times \mathbb{R}^q \times \mathbb{R}^m \rightarrow \mathbb{R}$

$$L_a(x, \mu, \lambda; \epsilon) = f(x) + \frac{1}{\epsilon} \phi_E(x, \mu, \lambda; \epsilon), \quad (38)$$

che è detta *funzione Lagrangiana aumentata* per il problema (2). In particolare, la proprietà (i) della funzione $\phi_E(x, \mu, \lambda; \epsilon)$ assicura ovviamente che la funzione L_a sia continuamente differenziabile. La proprietà (ii) implica che, quando il parametro di penalità $\epsilon \rightarrow 0$, la funzione $L_a(x, \mu, \lambda; \epsilon)$ tende alla funzione di penalità sequenziale esterna vista nella precedente sezione. La proprietà (iii) mostra che il termine ϕ_E è in grado di forzare la positività dei moltiplicatori λ_i , $i = 1, \dots, m$ e le condizioni di complementarità $g_i(x) \lambda_i = 0$, $i = 1, \dots, m$. Infine, l'ultima proprietà assicura che ogni tripla di KT (x^*, μ^*, λ^*) sia un punto stazionario della funzione Lagrangiana aumentata. Più in dettaglio vale la seguente proposizione ([3]).

Proposizione 5 *Sia $L_a(x, \mu, \lambda; \epsilon)$ la funzione definita in (38), allora per ogni valore $\epsilon > 0$, ogni punto stazionario di $L_a(x, \mu, \lambda; \epsilon)$ è una tripla di KT del problema (2) e viceversa.*

Dim. La prova segue direttamente dall'espressione del gradiente rispetto ad x , μ e λ di $L_a(x, \mu, \lambda; \epsilon)$. Infatti, ricordando il punto (a) della Proposizione 1, abbiamo

$$\nabla_x L_a(x, \mu, \lambda; \epsilon) = \nabla f(x) + \nabla g(x)(\lambda + \max\{g(x), -\epsilon\lambda\}/\epsilon) + \nabla h(x)(\mu + h(x)/\epsilon), \quad (39)$$

$$\nabla_\mu L_a(x, \mu, \lambda; \epsilon) = h(x), \quad (40)$$

$$\nabla_\lambda L_a(x, \mu, \lambda; \epsilon) = \max\{g(x), -\epsilon\lambda\}. \quad (41)$$

Quindi, se $(\bar{x}, \bar{\mu}, \bar{\lambda})$ è un punto stazionario di $L_a(x, \mu, \lambda; \epsilon)$ dalle (39)–(41) segue immediatamente che

$$\nabla_x L(\bar{x}, \bar{\mu}, \bar{\lambda}) = 0, \quad h(\bar{x}) = 0, \quad \max\{g(\bar{x}), -\epsilon\bar{\lambda}\} = 0.$$

Da queste, ricordando che $\max\{g_i(x), -\epsilon\lambda_i\} = 0$ se e solo se $g_i(x) \leq 0$, $\lambda_i \geq 0$ e $g_i(x)\lambda_i = 0$, $i = 1, \dots, m$, segue che $(\bar{x}, \bar{\mu}, \bar{\lambda})$ è una tripla di KT per il problema (2). Vice versa, se $(\bar{x}, \bar{\mu}, \bar{\lambda})$ è una tripla di KT del problema (2), segue che

$$\nabla_x L(\bar{x}, \bar{\mu}, \bar{\lambda}) = 0, \quad h(\bar{x}) = 0, \quad \max\{g(\bar{x}), -\epsilon\bar{\lambda}\} = 0,$$

e, quindi, dalle (39)–(41), che $(\bar{x}, \bar{\mu}, \bar{\lambda})$ è un punto stazionario di $L_a(x, \mu, \lambda; \epsilon)$. \square

Esempio 5 Si consideri il seguente problema non lineare

$$\begin{aligned} \min \quad & x^3 \\ & x \geq 0, \end{aligned}$$

che ammette, ovviamente, $x^* = 0$ come unico punto di minimo globale del problema con moltiplicatore associato $\lambda^* = 0$.

La funzione Lagrangiana aumentata per questo problema è la seguente.

$$L_a(x, \lambda; \epsilon) = x^3 + \lambda \max\{-x, -\epsilon\lambda\} + \frac{1}{2\epsilon} \max\{-x, -\epsilon\lambda\}^2.$$

In Figura 5 sono riportate le curve di livello della funzione $L_a(x, \lambda; \bar{\epsilon})$ per $\bar{\epsilon} = 0.09$. Si vede come in $(0, 0)$ la funzione $L_a(x, \lambda; \bar{\epsilon})$ ha un punto stazionario. In Figura 6 sono riportati invece gli andamenti della funzione $L_a(x, \lambda^*; \epsilon)$ per $\epsilon = 0.09, 0.15$ con $\lambda^* = 0$ moltiplicatore associato al punto di minimo globale del problema. Si vede come x^* sia un punto di minimo globale di $L_a(x, \lambda^*; \epsilon)$ sul compatto $[-5, 5]$ per $\epsilon = 0.09$ ma non per $\epsilon = 0.15$.

La proposizione che segue, consente di stabilire un legame tra minimi locali stretti del problema vincolato e punti di minimo locale stretto della funzione Lagrangiana aumentata. Più in particolare, vale il seguente risultato ([7],[11]).

Proposizione 6 *Sia \bar{x} una soluzione locale del problema (2) che soddisfa la (LICQ). Supponiamo, inoltre, che la tripla di KT $(\bar{x}, \bar{\mu}, \bar{\lambda})$ soddisfi la (SOSC). Allora esiste un valore $\bar{\epsilon} > 0$ tale che, per ogni $\epsilon \in (0, \bar{\epsilon}]$, \bar{x} è un minimo locale stretto della funzione $L_a(x, \bar{\mu}, \bar{\lambda}; \epsilon)$.*

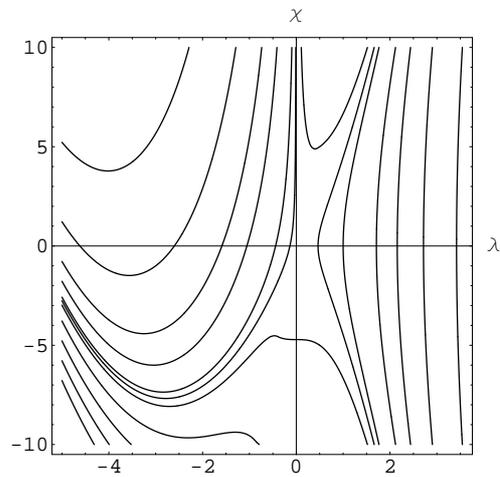


Figure 5: Curve di livello di $L_a(x, \lambda; \bar{\epsilon})$ ($\bar{\epsilon} = 0.09$) per il problema dell'Esempio 5.

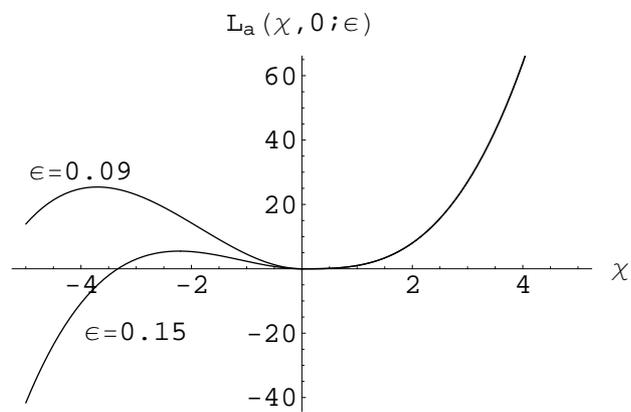


Figure 6: Andamento della funzione $L_a(x, \lambda^*; \bar{\epsilon})$ del problema dell'Esempio 5 con $\lambda^* = 0$ e $\epsilon = 0.09, 0.15$.

Dim. vedi [11] pag. 519. □

In base ai punti (a) e (b) della Proposizione 5, ogni punto stazionario del problema non vincolato

$$\min_{x, \mu, \lambda} L_a(x, \mu, \lambda; \epsilon)$$

per ϵ fissato è un punto di KT del problema (2). Pertanto, una tecnica di soluzione del problem (2) potrebbe essere quella di fissare ϵ ad un valore strettamente positivo e quindi minimizzare la funzione Lagrangiana aumentata nel tentativo di determinare un punto stazionario di L_a e, di conseguenza, una tripla di KT del problema (2).

Purtroppo, a meno di casi estremamente favorevoli, questa tecnica di soluzione è altamente sconsigliabile. Infatti, anche nel caso in cui il problema (2) ammetta una soluzione globale, è facile notare che la funzione $L_a(x, \mu, \lambda; \epsilon)$ risulta illimitata inferiormente rispetto ai moltiplicatori (μ, λ) . Pertanto, una sua minimizzazione non vincolata potrebbe produrre una successione di punti divergente. Il punto (c) di Proposizione 5 e la Proposizione 6 ci aiutano solo parzialmente. Questi due risultati teorici si basano infatti pesantemente sulla conoscenza dei moltiplicatori associati ad un punto di minimo globale (locale) \bar{x} . Solo in questo caso, è assicurata l'esistenza di un valore $\bar{\epsilon} > 0$ tale che, per ogni $\epsilon \in (0, \bar{\epsilon}]$, \bar{x} è un minimo globale (locale) di $L_a(x, \bar{\mu}, \bar{\lambda}; \epsilon)$.

Notiamo però che, se (x^*, μ^*, λ^*) è una tripla di KT per il problema (2) che soddisfa la (SOSC), allora (dal punto (a) di Proposizione 5) $\nabla_x L_a(x^*, \mu^*, \lambda^*; \epsilon) = 0$. Applicando quindi il teorema delle funzioni implicite al sistema di equazioni

$$\nabla_a L_a(x, \mu, \lambda; \epsilon) = 0,$$

risulta che esistono un intorno aperto di (μ^*, λ^*) ($S(\mu^*, \lambda^*)$), un intorno aperto di x^* ($S(x^*)$) ed una funzione $x(\mu, \lambda) : S(\mu^*, \lambda^*) \rightarrow S(x^*)$, continuamente differenziabile e tale che

$$\nabla_a L_a(x(\mu, \lambda), \mu, \lambda; \epsilon) = 0,$$

per ogni $(\mu, \lambda) \in S(\mu^*, \lambda^*)$. Inoltre, la Proposizione 6 ci assicura che esiste un valore $\bar{\epsilon} > 0$ tale che, per ogni $\epsilon \in (0, \bar{\epsilon}]$, la matrice $\nabla_{xx}^2 L_a(x^*, \mu^*, \lambda^*; \epsilon)$ è definita positiva cosicché è possibile assumere che per ogni $(x, \mu, \lambda) \in S(x^*) \times S(\mu^*, \lambda^*)$ anche $\nabla_{xx}^2 L_a(x, \mu, \lambda; \epsilon)$ è definita positiva e quindi $x(\mu, \lambda)$ è l'unico punto in $S(x^*)$ che, fissato $(\mu, \lambda) \in S(\mu^*, \lambda^*)$, minimizza rispetto ad x la funzione $L_a(x, \mu, \lambda; \epsilon)$. Quindi, possiamo immaginare che se (μ_k, λ_k) sono una buona stima dei moltiplicatori (μ^*, λ^*) , allora una minimizzazione della funzione $L_a(x, \mu_k, \lambda_k; \epsilon)$, per un valore di ϵ sufficientemente piccolo, fornisce una buona stima x_k di x^* .

Esempio 6 Si consideri il seguente problema non lineare

$$\min \begin{aligned} & \frac{1}{2}x^2 + \frac{1}{6}y^2 \\ & x + y = 1. \end{aligned}$$

È possibile verificare che questo problema ammette $(x^*, y^*) = (0.25, 0.75)$ come soluzione con moltiplicatore associato $\mu^* = -0.25$.

In Figura 7 sono riportate le curve di livello della funzione $L_a(x, y, \bar{\mu}; \epsilon)$ per $\epsilon = 1$ al variare del valore del moltiplicatore $\bar{\mu} = 0, -0.15, -0.25$. Si vede come il punto di minimo globale della funzione $L_a(x, y, \bar{\mu}; \epsilon)$ si avvicina alla soluzione fino a quando, per $\bar{\mu} = \mu^* = -0.25$, il minimo globale della funzione L_a coincide esattamente con la soluzione del problema.

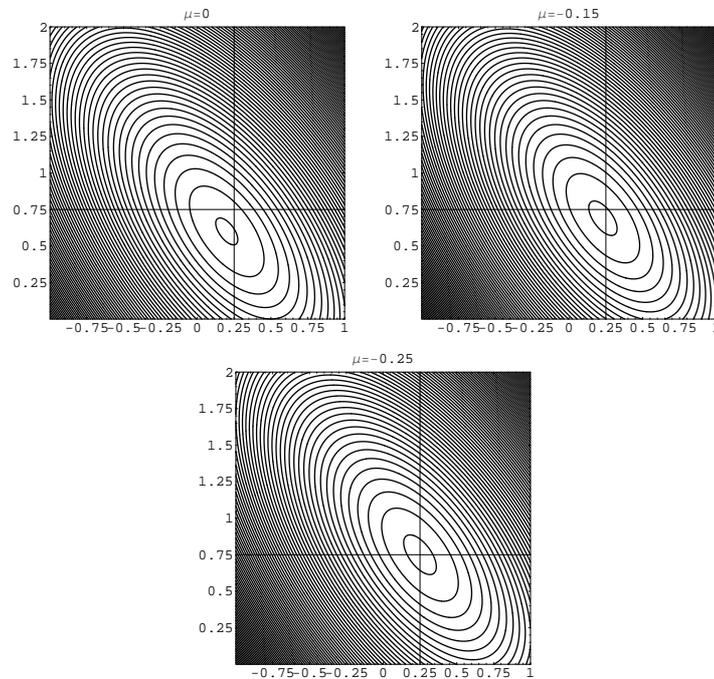


Figure 7: Grafici di $L_a(x, y, \bar{\mu}; \epsilon)$ per il problema dell'Esempio 6 con $\epsilon = 1$ e $\bar{\mu} = 0, -0.15, -0.25$. L'intersezione degli assi indica la soluzione del problema $(x^*, y^*) = (0.25, 0.75)$.

Per questi motivi, un metodo di soluzione basato sulla funzione Lagrangiana aumentata $L_a(x, \mu, \lambda; \epsilon)$ dovrà sempre includere una adeguata procedura per la stima dei moltiplicatori μ^* e λ^* unitamente ad una regola per la riduzione del parametro di penalità ϵ .

Dati un punto iniziale \bar{x} , un valore del parametro di penalità ϵ , una stima dei moltiplicatori $\hat{\mu}, \hat{\lambda}$ ed una tolleranza τ , definiamo l'operatore $\mathcal{M}_{L_a}[\bar{x}, \hat{\mu}, \hat{\lambda}, \epsilon, \tau]$ come quell'operatore che dato il punto iniziale \bar{x} fornisce un nuovo punto \hat{x} tale che

$$\|\nabla_x L_a(\hat{x}, \hat{\mu}, \hat{\lambda}; \epsilon)\| \leq \tau.$$

Uno schema di algoritmo per la soluzione di un problema vincolato mediante l'uso di funzioni Lagrangiane aumentate è il seguente.

Algoritmo SEQLAGR

Dati: $\epsilon_0 > 0, \beta > 1, \{\tau_k\} \rightarrow 0, \rho > 0, \tilde{x}, \mu_0$ e λ_0

Step 0: Poni $k = 0$ e $\bar{x} = \tilde{x}$.

Step 1: Calcola $x_k = \mathcal{M}_{L_a}[\bar{x}, \mu_k, \lambda_k, \epsilon_k, \tau_k]$

Step 2: Se $\|\nabla L_a(x_k, \mu_k, \lambda_k; \epsilon_k)\| < \rho$ allora STOP, altrimenti

2.1 poni $\bar{x} = x_k$,

2.2 calcola le nuove stime μ_{k+1} e λ_{k+1} dei moltiplicatori,

2.3 poni $\epsilon_{k+1} = \epsilon_k/\beta, k = k + 1$ e torna allo Step 1.

Facciamo nuovamente notare che, nel calcolare il nuovo punto x_k allo Step 1, è necessario controllare il comportamento del minimizzatore locale e, eventualmente, interromperne le iterazioni qualora ci si dovesse accorgere che la successione dei punti prodotti sta divergendo.

Per quanto riguarda l'aggiornamento dei moltiplicatori allo Step 2.2, i metodi più comunemente usati possono essere interpretati come riportato di seguito. Consideriamo per semplicità il problema con soli vincoli di uguaglianza (5) e la *funzione duale* data da

$$\psi(\mu) = L_a(x(\mu), \mu; \epsilon),$$

ove $\mu \in S(\mu^*), \epsilon \in (0, \bar{\epsilon}]$ ed $x(\mu) : S(\mu^*) \rightarrow S(x^*)$ tale che $\nabla_x L_a(x(\mu), \mu; \epsilon) = 0$. Se x^* è una soluzione locale di (5) e μ^* il corrispondente moltiplicatore di KT, risulta evidentemente $x(\mu^*) = x^*$ e $h(x^*) = 0$; pertanto abbiamo che

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

il che prova che μ^* è un punto di massimo locale per la funzione $\psi(\mu)$. Per determinare μ^* possiamo quindi applicare un metodo di minimizzazione nonvincolata alla funzione $-\psi(\mu)$. Dato che risulta

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

e, per definizione di $x(\mu)$, si ha $\nabla_x L_a(x(\mu), \mu; \epsilon) = 0$, otteniamo

$$\nabla\psi(\mu) = h(x(\mu)).$$

Perciù, minimizzare $-\psi(\mu)$ con il metodo del gradiente corrisponde all'iterazione

$$\mu_{k+1} = \mu_k + \alpha h(x_k).$$

In base ad ulteriori considerazioni, si pone normalmente il passo $\alpha = 1/\epsilon_k$ per ottenere la formula di aggiornamento

$$\mu_{k+1} = \mu_k + \frac{1}{\epsilon_k} h(x_k).$$

(N.B. Se (x_k, μ_k) è una coppia di KT, allora, in base all'iterazione precedente, $\mu_{k+1} = \mu_k$). Nel caso in cui siano presenti anche vincoli di disuguaglianza, la regola di aggiornamento dei moltiplicatori associati è la seguente:

$$\lambda_{k+1} = \max \left\{ 0, \lambda_k + \frac{1}{\epsilon_k} g(x_k) \right\} = \lambda_k + \frac{1}{\epsilon_k} \max\{g(x_k), -\epsilon_k \lambda_k\}.$$

(N.B. Se (x_k, λ_k) è una coppia di KT, allora, in base all'iterazione precedente, $\lambda_{k+1} = \lambda_k$).

3.1 Un caso pratico: LANCELOT

In questa sezione presenteremo lo schema generale di un algoritmo noto in letteratura con l'acronimo LANCELOT (Large And Nonlinear Constrained Extended Lagrangian Optimization Techniques) di Conn, Gould e Toint [12].

LANCELOT si applica alla soluzione di problemi vincolati del seguente tipo:

$$\begin{aligned} \min f(x) \\ h(x) &= 0 \\ l \leq x &\leq u, \end{aligned} \tag{42}$$

ove $x, l, u \in \mathfrak{R}^n$, $h(x) : \mathfrak{R}^n \rightarrow \mathfrak{R}^q$ e ove si assume che se l' i -esima variabile non è limitata inferiormente (superiormente) allora $l_i = -\infty$ ($u_i = +\infty$). Notiamo che il problema (2) può sempre essere ricondotto alla forma del problema (42) mediante l'aggiunta di una *variabile di slack* s_i per ogni vincolo di disuguaglianza $g_i(x) \leq 0$, $i = 1, \dots, m$ ed esplicitando i vincoli di bound sulle variabili originali come fatto nel problema che segue.

$$\begin{aligned} \min f(x) \\ h(x) &= 0 \\ g(x) + s &= 0 \\ s &\geq 0 \\ l \leq x &\leq u, \end{aligned}$$

con $s \in \mathfrak{R}^m$.

La funzione Lagrangiana aumentata usata da LANCELOT [12] è quella che si otterrebbe per il problema (42) trascurando i bound sulle variabili ovvero:

$$L_a(x, \mu; \epsilon) = f(x) + \mu^\top h(x) + \frac{1}{2\epsilon} \|h(x)\|^2. \quad (43)$$

I vincoli di bound sulle variabili sono infatti trattati esplicitamente, nel senso che ad ogni iterazione e fissati i moltiplicatori $\mu \in \mathfrak{R}^q$ alla stima corrente $\bar{\mu}$ viene risolto un sottoproblema della forma

$$\min_x L_a(x, \bar{\mu}; \epsilon) \\ l \leq x \leq u. \quad (44)$$

Mediante una semplice riscrittura delle condizioni di KT per il problema (44), è immediato verificare che condizione necessaria affinché un punto \bar{x} sia una soluzione del problema (44) è che

$$\mathcal{P}_{[l,u]} \nabla_x L_a(\bar{x}, \mu; \epsilon) = 0,$$

dove $\mathcal{P}_{[l,u]}g$ indica la proiezione di del vettore g sul box $[l, u]$ ed è definito come

$$(\mathcal{P}_{[l,u]}g)_i = \begin{cases} \min\{0, g_i\} & \text{se } x_i = l_i \\ g_i & \text{se } l_i < x_i < u_i \\ \max\{0, g_i\} & \text{se } x_i = u_i \end{cases} \quad \forall i = 1, \dots, n.$$

Sia $\mathcal{M}_{L_a, l, u}[\bar{x}, \hat{\mu}, \epsilon, \omega]$ quell'operatore che dato il punto iniziale \bar{x} , una stima dei moltiplicatori $\hat{\mu}$, un valore del parametro di penalità ϵ ed una tolleranza ω , fornisce un nuovo punto \hat{x} tale che

$$\|\mathcal{P}_{[l,u]} \nabla_x L_a(\hat{x}, \hat{\mu}; \epsilon)\| \leq \omega,$$

ovvero una soluzione approssimata del problema (44). Con queste premesse è possibile definire schematicamente il funzionamento di LANCELOT [12] come segue.

Algoritmo LANCELOT

Dati: $0 < \epsilon_o \leq 1, \bar{\eta}, \bar{\omega}, \tau < 1, \bar{\gamma} < 1, \alpha_\omega, \beta_\omega,$

$$\alpha_\eta < \min\{1, \alpha_\omega\}, \beta_\eta < \min\{1, \beta_\omega\}, \eta_\star, \omega_\star, \tilde{x}, \mu_o.$$

Step 0: Poni $\alpha_o = \min\{\epsilon_o, \bar{\gamma}\}, \eta_o = \bar{\eta}(\alpha_o)^{\alpha_\eta}, \omega_o = \bar{\omega}(\alpha_o)^{\alpha_\omega},$
 $k = 0$ e $\bar{x} = \tilde{x}.$

Step 1: Calcola $x_k = \mathcal{M}_{L_a, l, u}[\bar{x}, \mu_k, \epsilon_k, \omega_k]$

Step 2: Se $\|c(x_k)\| \leq \eta_k$ allora vai allo Step 3
altrimenti vai allo Step 4.

Step 3: (*Test di convergenza*)

Se $\|c(x_k)\| \leq \eta_\star$ e $\|\mathcal{P}_{[l, u]} \nabla_x L_a(x_k, \mu_k; \epsilon_k)\| \leq \omega_\star$ allora STOP
altrimenti (*aggiorna i moltiplicatori e riduci le tolleranze*)

$$\mu_{k+1} = \mu_k + (1/\epsilon_k)c(x_k);$$

$$\epsilon_{k+1} = \epsilon_k;$$

$$\alpha_{k+1} = \epsilon_k;$$

$$\eta_{k+1} = \eta_k(\alpha_{k+1})^{\beta_\eta};$$

$$\omega_{k+1} = \omega_k(\alpha_{k+1})^{\beta_\omega};$$

poni $k = k + 1$ e torna allo Step 1.

Step 4: (*Riduci il parametro di penalità e resetta le tolleranze*)

$$\mu_{k+1} = \mu_k;$$

$$\epsilon_{k+1} = \tau \epsilon_k;$$

$$\alpha_{k+1} = \bar{\gamma} \epsilon_{k+1};$$

$$\eta_{k+1} = \bar{\eta}(\alpha_{k+1})^{\alpha_\eta};$$

$$\omega_{k+1} = \bar{\omega}(\alpha_{k+1})^{\alpha_\omega};$$

poni $k = k + 1$ e torna allo Step 1.

Notiamo che allo Step 2 dell'algoritmo viene controllata l'ammissibilità ($\|c(x_k)\| \leq \eta_k$) del punto soluzione del sottoproblema (44). Nel caso in cui il punto x_k risulta essere troppo inammissibile ($\|c(x_k)\| > \eta_k$) allora, evidentemente, l'attuale valore del parametro di penalità ϵ_k è troppo alto nel senso che non è sufficiente a forzare l'ammissibilità dei punti prodotti. In questo caso quindi si riduce ϵ_k e si stringono tutte le tolleranze. Al contrario, se il punto corrente x_k è giudicato sufficientemente ammissibile, allora si controlla il criterio di arresto complessivo e nel caso questo non sia soddisfatto si procede con l'aggiornamento

dei moltiplicatori e la riinizializzazione delle tolleranze senza però modificare il parametro di penalità che sembra essere sufficientemente piccolo da forzare l'ammissibilità dei punti prodotti.

4 Funzioni di Penalità esatte continuamente differenziabili

Come abbiamo visto nella Sezione 3, il punto (c) di Proposizione 5 e la Proposizione 6, nell'ipotesi di conoscere i moltiplicatori associati ad un punto di minimo globale (rispettivamente, locale) del problema (2), assicurano l'esistenza di un valore di soglia $\bar{\epsilon}$ per il parametro di penalità tale che per ogni $\epsilon \in (0, \bar{\epsilon}]$ è possibile stabilire una corrispondenza tra punti di minimo della funzione Lagrangiana aumentata e punti di minimo del problema (2). Tuttavia, questa corrispondenza non è soddisfacente, da un punto di vista teorico, se il nostro scopo è quello di risolvere il problema vincolato (2) mediante una singola minimizzazione non vincolata di una opportuna funzione continuamente differenziabile. Al contrario, una *funzione di Penalità esatta* continuamente differenziabile, per sua stessa definizione, è una funzione $P_E(x; \epsilon)$ con la proprietà che, per “opportuni” valori del parametro di penalità ϵ , una sua singola minimizzazione non vincolata fornisce una soluzione del problema vincolato (2). Affinchè queste funzioni abbiano un qualche interesse applicativo è usuale imporre che le proprietà di esattezza permangano inalterate per tutti i valori di ϵ al di sotto di una certa soglia, in modo che sia facile determinare un valore di ϵ che realizzi la sopracitata corrispondenza.

In pratica, come vedremo, l'esistenza di un valore di soglia per il parametro di penalità ϵ e quindi le proprietà di esattezza della funzione $P_E(x; \epsilon)$ possono essere stabilite solo con riferimento ad un opportuno insieme compatto \mathcal{C} . Quindi, invece di considerare il problema (2), bisognerà considerare il problema

$$\min_{x \in \mathcal{F} \cap \mathcal{C}} f(x), \quad (45)$$

dove \mathcal{C} è un opportuno insieme che soddisfa la seguente assunzione:

Assunzione 1 \mathcal{C} è un insieme compatto tale che $\mathcal{G}(\mathcal{C}) \subset \overset{\circ}{\mathcal{C}}$.

Come vedremo, sarà possibile stabilire un legame, almeno per valori sufficientemente piccoli di ϵ , tra il problema (45) ed il seguente problema non vincolato:

$$\min_{x \in \overset{\circ}{\mathcal{C}}} P_E(x; \epsilon). \quad (46)$$

Dal momento che $\overset{\circ}{\mathcal{C}}$ è un insieme aperto, ogni soluzione locale del problema (46), se esiste, è non vincolata.

Analogamente a quanto fatto nell'introduzione, indichiamo con $\mathcal{G}_{P_E}(X; \epsilon)$ e con $\mathcal{L}_{P_E}(X; \epsilon)$ rispettivamente, l'insieme dei punti di minimo globale e l'insieme dei punti di minimo locale della funzione $P_E(x; \epsilon)$ sull'insieme $\overset{\circ}{X}$, ovvero

$$\mathcal{G}_{P_E}(X; \epsilon) = \{x \in \overset{\circ}{X} : P_E(x; \epsilon) \leq P_E(y; \epsilon), \forall y \in \overset{\circ}{X}\},$$

$$\mathcal{L}_{P_E}(X; \epsilon) = \{x \in \overset{\circ}{X} : \exists \rho > 0, P_E(x; \epsilon) \leq P_E(y; \epsilon), \forall y \in \overset{\circ}{X} \cap \mathcal{B}(x; \rho)\}.$$

Ovviamente, quindi, $\mathcal{G}_{P_E}(\mathcal{C}; \epsilon)$ ($\mathcal{L}_{P_E}(\mathcal{C}; \epsilon)$) è l'insieme delle soluzioni globali (locali) del problema (46).

Nel seguito di questa sezione adotteremo la notazione

$$\begin{aligned} \nabla_x L(x, \mu(x), \lambda(x)) &\triangleq [\nabla_x L(x, \mu, \lambda)]_{\substack{\mu = \mu(x) \\ \lambda = \lambda(x)}} \\ \nabla_x^2 L(x, \mu(x), \lambda(x)) &\triangleq [\nabla_x^2 L(x, \mu, \lambda)]_{\substack{\mu = \mu(x) \\ \lambda = \lambda(x)}} \end{aligned}$$

e supporremo sempre verificata la seguente assunzione:

Assunzione 2 Ogni punto $x \in \mathcal{F}$ soddisfa la (LICQ)

L'idea su cui si basano le funzioni di penalità esatte continuamente differenziabili è quella di approssimare in una funzione Lagrangiana aumentata i moltiplicatori di KT $(\bar{\mu}, \bar{\lambda})$ mediante delle funzioni della sola variabile x . Più precisamente la struttura base di una funzione di penalità di questo tipo è la seguente [13], [14], [15], [16], [17]:

$$P_E(x; \epsilon) = f(x) + \frac{1}{\epsilon} \phi_E(x, \mu(x), \lambda(x); \epsilon), \quad (47)$$

ove il termine di penalità $\phi_E(x, \mu(x), \lambda(x); \epsilon)$ è lo stesso di quello usato per definire funzioni Lagrangiane aumentate mentre $\mu(x)$ e $\lambda(x)$ sono delle particolari funzioni che godono delle seguenti proprietà:

- (i) $\mu(\cdot) : \mathfrak{R}^n \rightarrow \mathfrak{R}^q$ e $\lambda(\cdot) : \mathfrak{R}^n \rightarrow \mathfrak{R}^m$;
- (ii) se (x^*, μ^*, λ^*) è una tripla di KT allora $\mu(x^*) = \mu^*$ e $\lambda(x^*) = \lambda^*$;
- (iii) $\mu(\cdot)$ e $\lambda(\cdot)$ sono funzioni continuamente differenziabili.

In letteratura, tutte le funzioni che godono delle proprietà (i)–(iii) sono dette *funzioni moltiplicatrici*.

Un modo per definire delle funzioni moltiplicatrici $\mu(x)$, $\lambda(x)$ è quello di considerare la funzione

$$\Psi(\mu, \lambda; x) = \|\nabla_x L(x, \mu, \lambda)\|^2 + \|G(x)\lambda\|^2 + r(x) [\|\mu\|^2 + \|\lambda\|^2],$$

ove

$$r(x) = \sum_{i=1}^m \max\{0, g_i(x)\}^3 + \sum_{j=1}^q h_j(x)^2.$$

Per ogni fissato x , la funzione $\Psi(\mu, \lambda; x)$, può essere vista come una misura della violazione delle condizioni necessarie

$$\nabla_x L(x, \mu, \lambda) = 0, \quad h(x) = 0, \quad g(x) \leq 0, \quad G(x)\lambda = 0.$$

Sia $M(x)$ una matrice $(m+q) \times (m+q)$ definita come segue

$$M(x) = \begin{bmatrix} \nabla h(x)^\top \nabla h(x) + r(x)I_q & \nabla h(x)^\top \nabla g(x) \\ \nabla g(x)^\top \nabla h(x) & \nabla g(x)^\top \nabla g(x) + G^2(x) + r(x)I_m \end{bmatrix}. \quad (48)$$

Sotto l'Assunzione 2, è possibile ottenere le espressioni analitiche delle funzioni moltiplicatrici calcolando le espressioni delle funzioni $\mu(x)$ e $\lambda(x)$ che minimizzano la funzione $\Psi(\mu, \lambda; x)$ come evidenziato dalla proposizione che segue [18], [17].

Proposizione 7

- (a) la matrice $M(x)$ definita in (48) è definita positiva;
 (b) esiste un unico punto di minimo $(\mu(x), \lambda(x))$ della funzione $\Psi(\mu, \lambda; x)$ dato da:

$$\begin{bmatrix} \mu(x) \\ \lambda(x) \end{bmatrix} = -M^{-1}(x) \begin{bmatrix} \nabla h(x)^\top \\ \nabla g(x)^\top \end{bmatrix} \nabla f(x); \quad (49)$$

- (c) se (x^*, μ^*, λ^*) è una tripla di KT per il problema (2), allora si ha $\mu(x^*) = \mu^*$ e $\lambda(x^*) = \lambda^*$;
 (d) le funzioni $\mu(x)$ e $\lambda(x)$ definite dalla (49) sono continuamente differenziabili e risulta:

$$\begin{bmatrix} \nabla \mu(x)^\top \\ \nabla \lambda(x)^\top \end{bmatrix} = -M(x)^{-1} \begin{bmatrix} Q(x) \\ R(x) \end{bmatrix}, \quad (50)$$

dove

$$\begin{aligned} Q(x) &= \sum_{j=1}^q e_j \nabla_x L(x, \mu(x), \lambda(x))^\top \nabla^2 h_j(x) \\ &\quad + \nabla h(x)^\top \nabla_x^2 L(x, \mu(x), \lambda(x)) + \mu(x) \nabla r(x)^\top, \\ R(x) &= \sum_{i=1}^m e_i \nabla_x L(x, \mu(x), \lambda(x))^\top \nabla^2 g_i(x) + 2\Lambda(x) G(x) \nabla g(x)^\top \\ &\quad + \nabla g(x)^\top \nabla_x^2 L(x, \mu(x), \lambda(x)) + \lambda(x) \nabla r(x)^\top \end{aligned}$$

e

$$\nabla r(x)^\top = 3 \sum_{i=1}^m \max\{0, g_i(x)\}^2 \nabla g_i(x)^\top + 2 \sum_{j=1}^q h_j(x) \nabla h_j(x)^\top.$$

Dim. Punto (a). Notiamo che la matrice $M(x)$ può essere equivalentemente riscritta come

$$M(x) = \begin{bmatrix} \nabla h(x)^\top & 0 & r(x)^{1/2} I_q & 0 \\ \nabla g(x)^\top & G(x) & 0 & r(x)^{1/2} I_m \end{bmatrix} \begin{bmatrix} \nabla h(x) & \nabla g(x) \\ 0 & G(x) \\ r(x)^{1/2} I_q & 0 \\ 0 & r(x)^{1/2} I_m \end{bmatrix}.$$

In base all'Assunzione 2, abbiamo che

$$\text{rank} \begin{bmatrix} \nabla h(x) & \nabla g(x) \\ 0 & G(x) \\ r(x)^{1/2} I_q & 0 \\ 0 & r(x)^{1/2} I_m \end{bmatrix} = m + q,$$

perciù la matrice $M(x)$ è non singolare e definita positiva, che prova il punto (a). Il punto (a) della proposizione implica che il vettore $(\mu(x), \lambda(x))$ dato dalla (49) è l'unico punto di minimo globale della funzione quadratica $\Psi(\mu, \lambda; x)$, che prova il punto (b). Per quanto riguarda il punto (c), supponiamo ora che (x^*, μ^*, λ^*) sia una tripla di KT per il problema (2), allora si ha $\Psi(\mu^*, \lambda^*; x^*) = 0$ e quindi (visto che $\Psi(\mu, \lambda; x) \geq 0$) $\mu(x^*) = \mu^*$ e $\lambda(x^*) = \lambda^*$.

Passiamo ora al punto (d). Notiamo anzitutto che dalla (49) si ha

$$\begin{bmatrix} \nabla h(x)^\top \\ \nabla g(x)^\top \end{bmatrix} \nabla_x L(x, \mu(x), \lambda(x)) + \begin{bmatrix} r(x) & 0 \\ 0 & G^2(x) + r(x) \end{bmatrix} \begin{bmatrix} \mu(x) \\ \lambda(x) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \quad (51)$$

Usando l'espansione in diadi, possiamo riscrivere la (51) come segue

$$\begin{aligned} \sum_{j=1}^q e_j \nabla h_j(x)^\top \nabla_x L(x, \mu(x), \lambda(x)) + r(x) \mu(x) &= 0, \\ \sum_{i=1}^m e_i \nabla g_i(x)^\top \nabla_x L(x, \mu(x), \lambda(x)) + (G(x)^2 + r(x)) \lambda(x) &= 0, \end{aligned}$$

da cui, derivando rispetto ad x otteniamo

$$\begin{aligned} &\nabla h(x)^\top \left[\nabla_{\mu, x}^2 L(x, \mu(x), \lambda(x)) \nabla \mu(x)^\top + \nabla_{\lambda, x}^2 L(x, \mu(x), \lambda(x)) \nabla \lambda(x)^\top \right] \\ &+ r(x) \nabla \mu(x)^\top = \\ &\quad - \left[\sum_{j=1}^q e_j \nabla_x L(x, \mu(x), \lambda(x))^\top \nabla^2 h_j(x) + \nabla h(x)^\top \nabla_x^2 L(x, \mu(x), \lambda(x)) + \mu(x) \nabla r(x)^\top \right], \\ &\nabla g(x)^\top \left[\nabla_{\mu, x}^2 L(x, \mu(x), \lambda(x)) \nabla \mu(x)^\top + \nabla_{\lambda, x}^2 L(x, \mu(x), \lambda(x)) \nabla \lambda(x)^\top \right] \\ &+ (G(x)^2 + r(x)) \nabla \lambda(x)^\top = \\ &\quad - \left[\sum_{i=1}^m e_i \nabla_x L(x, \mu(x), \lambda(x))^\top \nabla^2 g_i(x) + \nabla g(x)^\top \nabla_x^2 L(x, \mu(x), \lambda(x)) \right. \\ &\quad \left. + 2\Lambda(x) G(x) \nabla g(x)^\top + \lambda(x) \nabla r(x)^\top \right]. \end{aligned}$$

A questo punto, è sufficiente ricordare che $\nabla_{\mu, x}^2 L(x, \mu(x), \lambda(x)) = \nabla h(x)$ e $\nabla_{\lambda, x}^2 L(x, \mu(x), \lambda(x)) = \nabla g(x)$, per ottenere la (50). \square

Grazie al punto (d) di Proposizione 7, è possibile dimostrare che la funzione $P_E(x; \epsilon)$ ottenuta usando come funzioni moltiplicatrici le funzioni $\mu(x)$ e $\lambda(x)$ date da (49), è continuamente differenziabile ([18]).

Proposizione 8 La funzione $P_E(x; \epsilon)$ definita da (47) con $\mu(x)$ e $\lambda(x)$ date da (49) è continuamente differenziabile per ogni $x \in \mathbb{R}^n$ ed il suo gradiente è dato da:

$$\begin{aligned} \nabla P_E(x; \epsilon) &= \nabla_x L(x, \mu(x), \lambda(x)) + \frac{1}{\epsilon} (\nabla h(x) + \epsilon \nabla \mu(x)) h(x) \\ &\quad + \frac{1}{\epsilon} (\nabla g(x) + \epsilon \nabla \lambda(x)) \max\{g(x), -\epsilon \lambda(x)\}, \end{aligned} \quad (52)$$

dove $\nabla \mu(x)$ e $\nabla \lambda(x)$ sono dati dalla (50).

Dim. Ricordando che

$$P_E(x; \epsilon) = [L_a(x, \mu, \lambda; \epsilon)]_{\substack{\mu = \mu(x) \\ \lambda = \lambda(x)}}$$

abbiamo, per l'espressione del gradiente rispetto ad x ,

$$\begin{aligned} \nabla P_E(x; \epsilon) &= [\nabla_x L_a(x, \mu, \lambda; \epsilon)]_{\substack{\mu = \mu(x) \\ \lambda = \lambda(x)}} + \nabla \mu(x) [\nabla_\mu L_a(x, \mu, \lambda; \epsilon)]_{\substack{\mu = \mu(x) \\ \lambda = \lambda(x)}} \\ &\quad + \nabla \lambda(x) [\nabla_\lambda L_a(x, \mu, \lambda; \epsilon)]_{\substack{\mu = \mu(x) \\ \lambda = \lambda(x)}} \end{aligned}$$

che, richiamando le (39)–(41) da esattamente la (52). \square

Un primo risultato teorico che consente di stabilire un legame tra triple di KT per il problema vincolato (2) e punti stazionari della funzione $P_E(x; \epsilon)$ è il seguente:

Proposizione 9 Sia (x^*, μ^*, λ^*) una tripla di KT per il problema (2). Allora, per ogni $\epsilon > 0$:

- (a) x^* è un punto stazionario di $P_E(x; \epsilon)$;
- (b) $P_E(x^*; \epsilon) = f(x^*)$.

Dim. Dal momento che (x^*, μ^*, λ^*) è una tripla di KT per il problema (2), il punto (c) di Proposizione 7 ci assicura che $\mu(x^*) = \mu^*$ e $\lambda(x^*) = \lambda^*$. Inoltre, dalle condizioni di KT abbiamo che

$$\begin{aligned} \nabla_x L(x^*, \mu(x^*), \lambda(x^*)) &= 0, \\ h(x) = 0, \quad \max\{g(x^*), -\epsilon \lambda(x^*)\} &= 0, \end{aligned}$$

da cui segue, ricordando l'espressione (52), che $\nabla P_E(x^*; \epsilon) = 0$.

Per dimostrare il punto (b) è sufficiente ricordare la proprietà (iii) del termine di penalità $\phi_E(x, \mu, \lambda; \epsilon)$ e la (47). \square

In Figure 8 sono riportati gli andamenti delle curve di livello per la funzione $P_E(x, y; \epsilon)$ relativa al problema dell'Esempio 6. Come si può notare, la funzione $P_E(x, y; \epsilon)$ ha in $(x^*, y^*) = (0.25, 0.75)$ un punto stazionario per qualunque

valore del parametro di penalità ϵ . Notiamo anche che (x^*, y^*) è un minimo di $P_E(x, y; \epsilon)$ solo per valori di ϵ sufficientemente piccoli anche se, abbassando troppo il valore del parametro di penalità, comincia a diventare evidente il mal condizionamento della funzione $P_E(x, y; \epsilon)$.

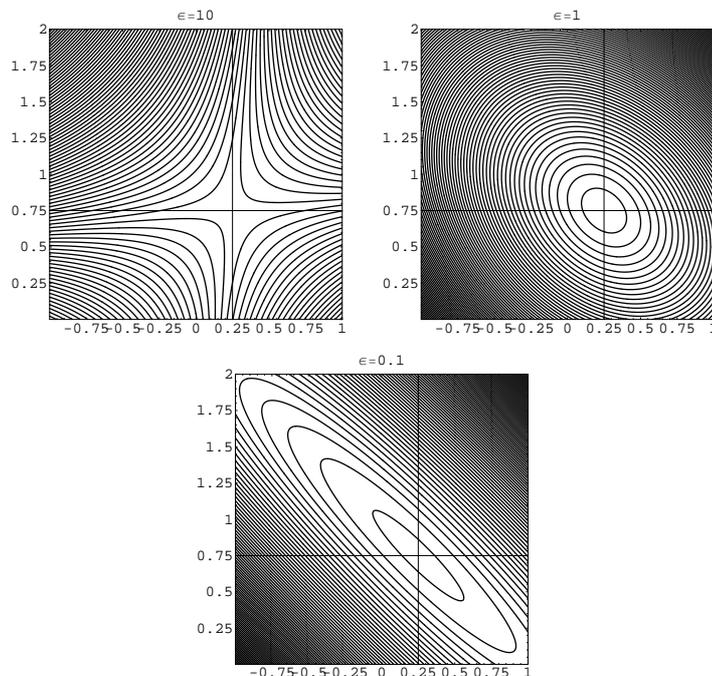


Figure 8: Grafici di $P_E(x, y; \epsilon)$ ($\epsilon = 10, 1, 0.1$) per il problema dell'Esempio 6.

Esempio 7 Si consideri il seguente problema non lineare

$$\begin{aligned} \max \quad & (x - 1)^2 + (y - 1)^2 \\ & x^2 + y^2 \leq 1 \\ & x^2 - 2x - y \leq -1. \end{aligned}$$

È possibile verificare che questo problema ammette $(\bar{x}, \bar{y}) = (0, 1)$ e $(\hat{x}, \hat{y}) = (1, 0)$ come soluzioni globali con moltiplicatori associati $\bar{\lambda} = (1/2, 1)$ e $\hat{\lambda} = (0, 2)$ rispettivamente.

In Figura 7 sono riportate le curve di livello della funzione $P_E(x, y; \epsilon)$ per $\epsilon = 10, 1, 0.1$. Si vede come in (\bar{x}, \bar{y}) e (\hat{x}, \hat{y}) la funzione $P_E(x, y; \epsilon)$ ha sempre dei punti stazionari che diventano minimi stretti solo per ϵ sufficientemente piccoli.

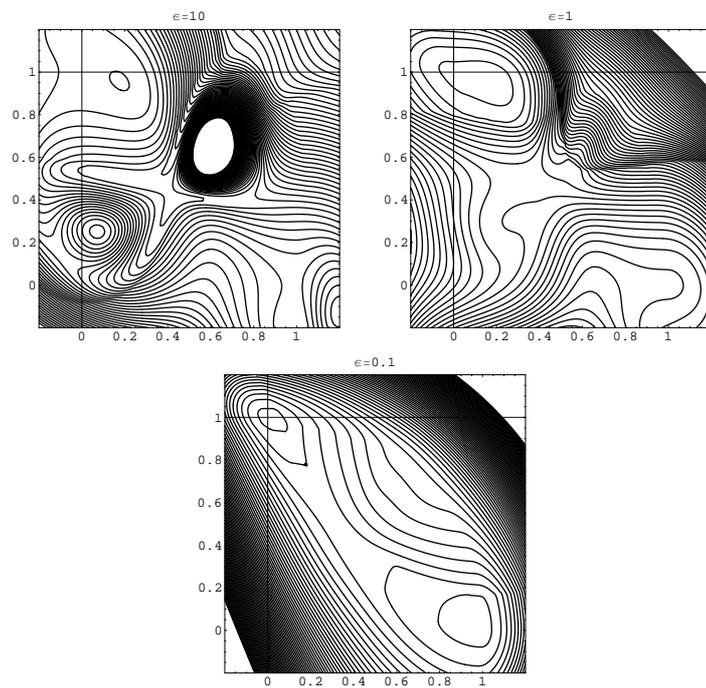


Figure 9: Grafici di $P_E(x, y; \epsilon)$ ($\epsilon = 10, 1, 0.1$) per il problema dell'Esempio 7.

Proposizione 10 Sia $x^* \in \mathfrak{R}^n$ un punto tale che

$$\nabla P_E(x^*; \epsilon) = 0, \quad h(x^*) = 0, \quad \max\{g(x^*), -\epsilon\lambda(x^*)\} = 0.$$

Allora, per ogni $\epsilon > 0$, $(x^*, \mu(x^*), \lambda(x^*))$ è una tripla di KT per il problema (2).

Dim. La dimostrazione segue banalmente dalle ipotesi e considerando l'espressione (52) del gradiente della funzione $P_E(x; \epsilon)$. \square

Nel seguito avremo bisogno della seguente assunzione sul problema vincolato (2).

(EMFCQ)

Extended Mangasarian-Fromovitz Constraint Qualification

Un punto $x \in \mathfrak{R}^n$ soddisfa la (EMFCQ) se non è possibile trovare dei numeri $v_j, j = 1, \dots, q$ e $u_i, i \in I_+(x)$ tali che

$$\begin{aligned} \sum_{i \in I_+(x)} u_i \nabla g_i(x) + \sum_{j=1}^q v_j \nabla h_j(x) &= 0, \\ u_i &\geq 0, \quad i \in I_+(x), \\ (u_i, i \in I_+(x), v_j, j = 1, \dots, q) &\neq 0. \end{aligned}$$

Dato un generico sottoinsieme X di \mathfrak{R}^n , introduciamo la seguente assunzione:

Assunzione 3 L'insieme X è tale che ogni punto $x \in X$ soddisfa la (EMFCQ).

Nel seguito vedremo come mediante l'Assunzione 3 sia possibile dimostrare l'implicazione inversa di quella contenuta nella Proposizione 9 e cioè che per valori sufficientemente piccoli di ϵ , ma comunque strettamente positivi, ogni punto stazionario della funzione di Penalità $P_E(x; \epsilon)$ è un punto di KT per il problema (2) ([18], [15]).

Proposizione 11 Sia \mathcal{C} un sottoinsieme di \mathfrak{R}^n che soddisfa l'Assunzione 3.

Allora, esiste un $\bar{\epsilon} > 0$ tale che, per ogni $\epsilon \in (0, \bar{\epsilon}]$, se $x_\epsilon \in \overset{\circ}{\mathcal{C}}$ è un punto stazionario di $P_E(x; \epsilon)$, la tripla $(x_\epsilon, \mu(x_\epsilon), \lambda(x_\epsilon))$ è una tripla di KT per il problema (2).

Occorre qui sottolineare l'importanza teorica del risultato di Proposizione 11. Infatti, grazie a questo risultato e al punto (a) di Proposizione 9 è possibile stabilire una corrispondenza biunivoca tra punti stazionari della funzione di Penalità e triple di KT del problema vincolato (2), almeno con riferimento ad un insieme compatto \mathcal{C} e per valori di ϵ sufficientemente piccoli ma comunque strettamente positivi. Con riferimento alla funzione $P_E(x; \epsilon)$ è possibile stabilire

un legame anche tra minimi globali (locali) del problema (46) e minimi globali (locali) del problema (45).

Valgono le seguenti due proposizioni.

Proposizione 12 *Sia \mathcal{C} un sottoinsieme di \mathbb{R}^n che soddisfa l'Assunzione 1 e supponiamo che il problema (2) sia ben posto ovvero che $\mathcal{F} \neq \emptyset$. Allora, esiste un $\bar{\epsilon} > 0$ tale che, per ogni $\epsilon \in (0, \bar{\epsilon}]$, ogni punto di minimo globale del problema (45) è un punto di minimo globale di $P_E(x; \epsilon)$ su $\overset{\circ}{\mathcal{C}}$ e viceversa.*

Proposizione 13 *Sia \mathcal{C} un sottoinsieme di \mathbb{R}^n che soddisfa l'Assunzione 1 e sia \bar{x} un punto di minimo locale isolato per il problema (45). Allora, esiste un $\bar{\epsilon} > 0$ tale che, per ogni $\epsilon \in (0, \bar{\epsilon}]$, il punto \bar{x} è un punto di minimo locale isolato della funzione $P_E(x; \epsilon)$.*

Proposizione 14 *Sia \mathcal{C} un sottoinsieme di \mathbb{R}^n che soddisfa le Assunzioni 1 e 3. Allora, esiste un $\bar{\epsilon} > 0$ tale che, per ogni $\epsilon \in (0, \bar{\epsilon}]$, se $\bar{x} \in \overset{\circ}{\mathcal{C}}$ è un punto di minimo locale di $P_E(x; \epsilon)$, allora \bar{x} è un punto di minimo locale del problema (45) e $\mu(\bar{x})$, $\lambda(\bar{x})$ sono i rispettivi moltiplicatori di KT.*

I precedenti risultati teorici ed in particolare le Proposizioni 9, 11, 12, 14 meritano alcune parole di commento. Da un punto di vista teorico, questi risultati garantiscono (per valori sufficientemente piccoli del parametro di penalità ϵ e sotto l'Assunzione 2) l'esistenza di una corrispondenza biunivoca tra punti stazionari (punti di minimo globale/locale) della funzione $P_E(x; \epsilon)$ e triple di KT (punti di minimo globale/locale) del problema (45) almeno in relazione ad un sottoinsieme compatto $\mathcal{C} \subset \mathbb{R}^n$. Più in particolare, queste corrispondenze consentono (sempre per valori sufficientemente piccoli di ϵ) di risolvere il problema vincolato (45) ricercando delle soluzioni del problema non vincolato (46).

Tuttavia, questa equivalenza non è sufficiente, da un punto di vista algoritmico, a garantire che una minimizzazione non vincolata di $P_E(x; \epsilon)$ produca punti stazionari per il problema (45) e tanto meno per il problema (2). Infatti, un qualsiasi algoritmo di minimizzazione non vincolata applicato alla funzione $P_E(x; \epsilon)$ a partire da punti $x^\circ \in \overset{\circ}{\mathcal{C}}$ potrebbe venire attratto da un punto stazionario della funzione di penalità non appartenente all'insieme $\overset{\circ}{\mathcal{C}}$ e che quindi potrebbe non avere alcuna relazione con il problema vincolato (2). Analogamente, se si modifica l'algoritmo di minimizzazione non vincolata forzando il fatto che tutti i punti prodotti devono appartenere all'insieme $\overset{\circ}{\mathcal{C}}$ potremmo ottenere una successione di punti che può avere dei punti di accumulazione sulla frontiera di \mathcal{C} . Tutto questo accade perché non è stato sufficientemente indagato il comportamento della funzione $P_E(x; \epsilon)$ sulla frontiera dell'insieme \mathcal{C} .

4.1 Penalità esatte con termini di barriera ([1])

La difficoltà pratica evidenziata al termine della precedente discussione può essere superata definendo delle funzioni di penalità che abbiano la proprietà che,

per ogni ϵ , l'intersezione tra i loro insiemi di livello ed un opportuno insieme aperto \mathcal{A} siano degli insiemi compatti. In questo caso, infatti, un qualsiasi algoritmo di minimizzazione non vincolata modificato in modo da generare tutti i punti appartenenti all'insieme \mathcal{A} , produrrà una sequenza di punti i cui punti di accumulazione sono contenuti nell'insieme \mathcal{A} .

Funzioni di penalità che godano della precedente proprietà di compattezza possono essere definite sotto opportune ipotesi e facendo comparire nella loro espressione dei termini di barriera che assicurano, tra l'altro, che il loro valore tenda all'infinito all'avvicinarsi della frontiera dell'insieme \mathcal{A} . Prima di continuare ricordiamo che anche in questa sottosezione supporremo sempre valida l'Assunzione 2. Nel seguito di questa sezione, allo scopo di non appesantire eccessivamente la notazione, ci riferiremo al problema con *solli vincoli di disuguaglianza* (6).

Una funzione di penalità esatta che gode delle precedenti proprietà ha la seguente struttura

$$P_G(x; \epsilon) = f(x) + \frac{1}{\epsilon} \phi_G(x, \lambda(x); \epsilon), \quad (53)$$

dove $\lambda(x)$ sono le funzioni moltiplicatrici date da (49) e ϕ_G è un termine di penalità che presenta le stesse proprietà (i)–(iv) del termine ϕ_E con l'ulteriore proprietà che

$$\lim_{x \rightarrow \partial \mathcal{A}} \phi_G(x, \lambda; \epsilon) = +\infty. \quad (54)$$

Dato un qualsiasi punto $\tilde{x} \in \mathfrak{R}^n$, consideriamo il seguente insieme aperto

$$\mathcal{A} = \{x \in \mathfrak{R}^n : a(x) > 0\}, \quad (55)$$

dove

$$a(x) = \alpha - \sum_{i=1}^m \max\{g_i(x), 0\}^3, \quad (56)$$

ed $\alpha > 0$ è uno scalare scelto in modo tale che $\tilde{x} \in \mathcal{A}$. Notiamo che, per come è definito, l'insieme aperto \mathcal{A} è un rilassamento dell'insieme ammissibile cioè $\mathcal{F} \subset \mathcal{A}$ e che la funzione $a(x)$ si annulla per ogni x appartenente alla frontiera di \mathcal{A} .

Con queste premesse, una possibile espressione per il termine ϕ_G (particolarizzata al caso di soli vincoli di disuguaglianza) è riportata di seguito [17].

$$\begin{aligned} \phi_G(x, \lambda; \epsilon) &= \sum_{i=1}^m \frac{1}{p_i(x, \lambda_i; \epsilon)} \varphi(g_i(x), \epsilon p_i(x, \lambda_i; \epsilon) \lambda_i) \\ &= \sum_{i=1}^m \left(\epsilon \lambda_i + \frac{\max\{g_i(x), -\epsilon p_i(x, \lambda_i; \epsilon) \lambda_i\}}{2p_i(x, \lambda_i; \epsilon)} \right) \\ &\quad \max\{g_i(x), -\epsilon p_i(x, \lambda_i; \epsilon) \lambda_i\}, \end{aligned} \quad (57)$$

dove i termini $1/p_i(x, \lambda_i; \epsilon)$, $i = 1, \dots, m$, detti *termini di barriera*, devono essere positivi e tali da assicurare che, per ogni $\epsilon > 0$, sia verificata la proprietà (54). La scelta più semplice che assicura la proprietà (54) è sicuramente

$$p_i(x) = p_i(x, \lambda_i; \epsilon) = a(x), \quad i = 1, \dots, m,$$

dove la funzione $a(x)$ è data dalla (56).

Tuttavia, utilizzando delle scelte leggermente più complesse per i termini di barriera p_i , $i = 1, \dots, m$, si possono assicurare proprietà teoriche più forti. In particolare, la scelta che sembra essere più indicata (vedi [17]) è la seguente:

$$p_i(x, \lambda_i; \epsilon) = \frac{a(x)}{1 + \epsilon a(x) \lambda_i^2}, \quad i = 1, \dots, m. \quad (58)$$

La proposizione che segue riassume alcune proprietà preliminari della funzione $P_G(x; \epsilon)$.

Proposizione 15 Per ogni $\epsilon > 0$,

(a) $P_G(x; \epsilon)$ è continuamente differenziabile per ogni $x \in \mathcal{A}$ ed il suo gradiente è dato da:

$$\begin{aligned} \nabla P_G(x; \epsilon) &= \nabla_x L(x, \lambda(x)) + \sum_{i=1}^m \max\{g_i(x), -\epsilon p_i(x; \epsilon) \lambda_i(x)\} \\ &\cdot \left[\frac{1}{\epsilon p_i(x; \epsilon)} \nabla g_i(x) + \nabla \lambda_i(x) \right. \\ &\quad \left. - \frac{\max\{g_i(x), -\epsilon p_i(x; \epsilon) \lambda_i(x)\}}{2\epsilon p_i(x; \epsilon)^2} \nabla p_i(x; \epsilon) \right], \end{aligned} \quad (59)$$

dove, per $i = 1, \dots, m$, $p_i(x; \epsilon) = p_i(x, \lambda_i(x); \epsilon)$ e

$$\begin{aligned} \nabla p_i(x; \epsilon) &= -p_i(x; \epsilon)^2 \left[4\epsilon \lambda_i(x) \nabla \lambda_i(x) \right. \\ &\quad \left. + \frac{3}{a(x)^2} \sum_{j=1}^m \max\{0, g_j(x)\}^2 \nabla g_j(x) \right], \end{aligned} \quad (60)$$

e $\nabla \lambda(x)^\top$ è dato da (50);

(b) $P_G(x; \epsilon) \leq f(x)$ per ogni $x \in \mathcal{F}$;

(c) $f(x) - m/2 \leq P_G(x; \epsilon)$ per ogni $x \in \mathcal{A}$.

Dim. Il punto (a) segue direttamente dall'espressione (53) di $P_G(x; \epsilon)$ mediante ragionamenti analoghi a quelli fatti nella dimostrazione di Proposizione 8. Il punto (b)

può essere dimostrato ragionando come segue. Dalla (53) abbiamo che

$$P_G(x; \epsilon) - f(x) = \sum_{i=1}^m \left(\lambda_i(x) + \frac{\max\{g_i(x), -\epsilon p_i(x; \epsilon)\lambda_i(x)\}}{2\epsilon p_i(x; \epsilon)} \right) \quad (61)$$

$$\max\{g_i(x), -\epsilon p_i(x; \epsilon)\lambda_i(x)\}.$$

Ora si consideri l' i -esimo termine della sommatoria in (61). Sono possibili due situazioni:

- (i) $\max\{g_i(x), -\epsilon p_i(x; \epsilon)\lambda_i(x)\} = g_i(x)$. Questo implica, considerato che $x \in \mathcal{F}$, che

$$-\epsilon p_i(x; \epsilon)\lambda_i(x) \leq \max\{g_i(x), -\epsilon p_i(x; \epsilon)\lambda_i(x)\} = g_i(x) \leq 0,$$

ovvero che

$$\frac{g_i(x)}{\epsilon p_i(x; \epsilon)} + \lambda_i(x) \geq 0.$$

Quindi, dato che $g_i(x) \leq 0$, a maggior ragione risulta

$$\frac{g_i(x)}{2\epsilon p_i(x; \epsilon)} + \lambda_i(x) \geq 0.$$

Pertanto, in questo caso, l' i -esimo termine della sommatoria in (61) è negativo o nullo.

- (ii) $\max\{g_i(x), -\epsilon p_i(x; \epsilon)\lambda_i(x)\} = -\epsilon p_i(x; \epsilon)\lambda_i(x)$. In questo caso l' i -esimo termine della sommatoria in (61) risulta essere uguale a $-\epsilon p_i(x; \epsilon)\lambda_i(x)^2/2 \leq 0$.

Dai precedenti due punti rimane provato il punto (b) della proposizione. Passiamo ora a considerare il punto (c) e supponiamo quindi che $x \in \mathcal{A}$. Notiamo che, dalla espressione (58) dei termini $p_i(x; \epsilon)$, possiamo riscrivere la (61) nel seguente modo:

$$P_G(x; \epsilon) = f(x) + \sum_{i=1}^m \left(\lambda_i(x) \max\{g_i(x), -\epsilon p_i(x; \epsilon)\lambda_i(x)\} \right. \\ \left. + \max\{g_i(x), -\epsilon p_i(x; \epsilon)\lambda_i(x)\}^2 \left(\frac{1}{2\epsilon a(x)} + \frac{\lambda_i(x)^2}{2} \right) \right).$$

Ora, considerando che il minimo della forma quadratica $u + u^2/2$ è $-1/2$, abbiamo che $P_G(x; \epsilon) \geq f(x) - m/2$. \square

Nel seguito faremo uso della seguente assunzione:

Assunzione 4 *Almeno una delle due seguenti condizioni è verificata:*

- (a) *l'insieme $\bar{\mathcal{A}}$ è compatto;*
- (b) *il punto noto \tilde{x} è tale che $\tilde{x} \in \mathcal{A} \cap \mathcal{F}$ e la funzione obiettivo $f(x)$ è radialmente illimitata su $\bar{\mathcal{A}}$ (cioè, per ogni successione $\{x_k\} \subseteq \mathcal{A}$, tale che $\|x_k\| \rightarrow \infty$, si ha $f(x_k) \rightarrow +\infty$).*

Dato un punto $\tilde{x} \in \mathcal{A}$, consideriamo il seguente insieme di livello della funzione $P_G(x; \epsilon)$ relativo all'insieme \mathcal{A} :

$$\Omega(\tilde{x}; \epsilon) = \{x \in \mathcal{A} : P_G(x; \epsilon) \leq P_G(\tilde{x}; \epsilon)\}, \quad (62)$$

ove assumiamo che $\tilde{x} \in \mathcal{F}$ se il punto noto \tilde{x} è ammissibile.

Proposizione 16 *Sia verificata l'Assunzione 4. Allora, per ogni $\epsilon > 0$, l'insieme $\Omega(\tilde{x}; \epsilon)$ è un insieme chiuso e esiste un insieme compatto \mathcal{C} tale che*

$$\Omega(\tilde{x}; \epsilon) \subseteq \mathcal{C}.$$

Dim. Dimostriamo prima di tutto che $\Omega(\tilde{x}; \epsilon)$ è un insieme chiuso, cioè facciamo vedere che ogni punto limite \hat{x} di ogni successione $\{x_k\}$, tale che $x_k \in \Omega(\tilde{x}; \epsilon)$ per ogni k , appartiene a $\Omega(\tilde{x}; \epsilon)$. Supponiamo per contraddizione che $\hat{x} \notin \Omega(\tilde{x}; \epsilon)$. In questo caso, per la stessa definizione di $\Omega(\tilde{x}; \epsilon)$ e per la continuità di $P_G(x; \epsilon)$, abbiamo che $\hat{x} \in \partial\mathcal{A}$ e quindi che $a(\hat{x}) = 0$. Allora, ricordando che $x_k \in \Omega(\tilde{x}; \epsilon)$, per ogni k , otteniamo che

$$a(x_k)P_G(x_k; \epsilon) \leq a(x_k)P_G(\tilde{x}; \epsilon), \quad (63)$$

$$\lim_{k \rightarrow \infty} a(x_k)P_G(x_k; \epsilon) \leq \lim_{k \rightarrow \infty} a(x_k)P_G(\tilde{x}; \epsilon) = 0. \quad (64)$$

Dalla (64) e ricordando l'espressione (53) di $P_G(x; \epsilon)$, abbiamo che

$$\lim_{k \rightarrow \infty} a(x_k)P_G(x_k; \epsilon) = \|\max\{g_i(\hat{x}), -\epsilon p_i(\hat{x}; \epsilon)\lambda_i(\hat{x})\}\|^2 = 0$$

da cui segue che $\hat{x} \in \mathcal{F}$, che contraddice l'assunzione $\hat{x} \in \partial\mathcal{A}$.

Ora dobbiamo provare che esiste un insieme compatto \mathcal{C} tale che $\Omega(\tilde{x}; \epsilon) \subseteq \mathcal{C}$ per ogni $\epsilon > 0$. Supponiamo che sia verificato il punto (a) dell'Assunzione 4. In questo caso è sufficiente scegliere $\mathcal{C} = \bar{\mathcal{A}}$. Supponiamo ora che sia verificato il punto (b) di Assunzione 4. In tal caso, $\tilde{x} \in \mathcal{F}$. Pertanto, per ogni $x \in \Omega(\tilde{x}; \epsilon)$, i punti (b) e (c) di Proposizione 15 implicano che

$$f(x) - m/2 \leq P_G(\tilde{x}; \epsilon) \leq f(\tilde{x}).$$

Perciò, possiamo scrivere che

$$\Omega(\tilde{x}; \epsilon) \subseteq \{x \in \mathcal{A} : f(x) \leq f(\tilde{x}) + m/2\} = \mathcal{C}$$

con \mathcal{C} insieme compatto. Infatti, \mathcal{C} è chiuso per continuità di $f(x)$ ed è anche limitato perchè se esistesse una successione $\{x_k\} \subset \mathcal{C}$ tale che $\|x_k\| \rightarrow \infty$, dal punto (b) di Assunzione 4 avremmo che $f(x_k) \rightarrow +\infty$ il che contraddice il fatto che $f(x_k) \leq f(\tilde{x}) + m/2$. \square

Le proposizioni che seguono estendono alla funzione $P_G(x; \epsilon)$ le proprietà già viste per la funzione $P_E(x; \epsilon)$ (vedi Proposizioni 9 e 10).

Proposizione 17 Sia (x^*, λ^*) una tripla di KT per il problema (6). Allora, per ogni $\epsilon > 0$:

- (a) x^* è un punto stazionario di $P_G(x; \epsilon)$;
- (b) $P_G(x^*; \epsilon) = f(x^*)$.

Dim. La dimostrazione segue banalmente dalla dimostrazione di Proposizione 9. \square

Proposizione 18 Sia $x^* \in \mathfrak{R}^n$ un punto tale che

$$\nabla P_G(x^*; \epsilon) = 0, \quad \max\{g(x^*), -\epsilon\lambda(x^*)\} = 0.$$

Allora, per ogni $\epsilon > 0$, $(x^*, \lambda(x^*))$ è una tripla di KT per il problema (2).

Dim. La dimostrazione segue banalmente dalla dimostrazione di Proposizione 10. \square
Per poter dimostrare le proprietà di esattezza di cui gode la funzione $P_G(x; \epsilon)$ abbiamo bisogno della seguente assunzione:

(WMFCQ)

Weak Mangasarian-Fromovitz Constraint Qualification

Un punto $x \in \mathfrak{R}^n$ soddisfa la (WMFCQ) se non è possibile trovare dei numeri $u_i, i \in 1, \dots, m$ tali che

$$\begin{aligned} \sum_{i \in I_+} u_i \nabla g_i(x) &= 0, \\ u_i &> 0, \quad i \in I_+(x). \end{aligned}$$

Assunzione 5 Almeno una delle seguenti due condizioni è soddisfatta:

- (a) ogni punto $x \in \mathcal{A} \setminus \mathcal{F}$ soddisfa la (WMFCQ);
- (b) esiste ed è noto un punto $\tilde{x} \in \mathcal{F}$.

Proposizione 19 Siano verificate le Assunzioni 4 e 5. Allora, esiste un valore $\epsilon^* > 0$ tale che, per ogni $\epsilon \in (0, \epsilon^*]$, se $x_\epsilon \in \Omega(\tilde{x}; \epsilon)$ è un punto stazionario di $P_G(x; \epsilon)$, la coppia $(x_\epsilon, \lambda(x_\epsilon))$ è una coppia di KT per il problema (6).

Proposizione 20 Supponiamo che il problema (6) sia ben posto, ovvero che $\mathcal{F} \neq \emptyset$ e che sia soddisfatta l'Assunzione 4. Allora, esiste un ϵ^* tale che, per ogni $\epsilon \in (0, \epsilon^*]$, ogni punto di minimo globale del problema (6) è un punto di minimo globale di $P_G(x; \epsilon)$ su \mathcal{A} e vice versa.

Proposizione 21 Sia x^* un punto di minimo locale isolato per il problema (6). Allora esiste un ϵ^* tale che, per ogni $\epsilon \in (0, \epsilon^*]$, x^* è un punto di minimo locale non vincolato di $P_G(x; \epsilon)$.

Proposizione 22 *Sia soddisfatta l'Assunzione 5. Allora esiste un ϵ^* tale che, per ogni $\epsilon \in (0, \epsilon^*]$, se $x^* \in \mathcal{A}$ è un punto di minimo locale non vincolato di $P_G(x; \epsilon)$, allora x^* è un minimo locale del problema (6) con moltiplicatore di KT $\lambda(x^*)$.*

Con queste proprietà è possibile trasformare il problema vincolato (6) nel problema non vincolato

$$\min_{x \in \mathcal{A}} P_G(x; \epsilon), \quad (65)$$

almeno per valori sufficientemente piccoli del parametro di penalità ϵ , e quindi risolvendo quest'ultimo, determinare delle soluzioni (coppie di KT) per il problema vincolato di partenza (6).

Esempio 8 Si consideri il seguente problema non lineare

$$\begin{aligned} \max \quad & x^3 \\ & x \geq 1 \end{aligned}$$

È possibile verificare che questo problema ammette $\bar{x} = 1$ come unico punto di minimo globale con moltiplicatore associato $\bar{\lambda} = 3$. Notiamo inoltre che la regione ammissibile è non limitata e $f(x) = x^3$ è radialmente illimitata su \mathcal{F} . In Figura 10 sono riportati gli andamenti della funzione $P_G(x; \epsilon)$ (ottenuta scegliendo $\alpha = 1$ nella definizione dell'insieme \mathcal{A}) per $\epsilon = 10, 1, 0.1$ per $x \in \mathcal{A} = \{x \in \mathbb{R}^n : x \geq 0\}$.

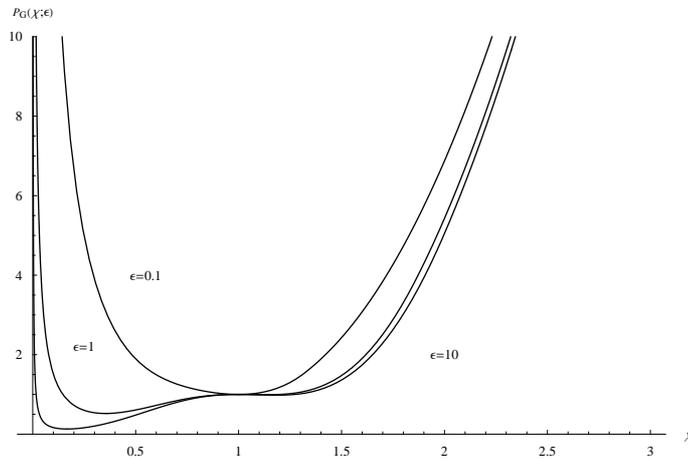


Figure 10: Grafici di $P_G(x; \epsilon)$ ($\epsilon = 10, 1, 0.1$) per il problema dell'Esempio 8.

La via più immediata da seguire è quella di definire metodi che si basino su una minimizzazione non vincolata della funzione $P_G(x; \epsilon)$. Nel perseguire questa

strada il primo problema che bisogna necessariamente affrontare è quello di definire delle regole che consentano di determinare dei valori del parametro di penalità ϵ per cui sono verificate le proprietà di esattezza viste precedentemente, cioè dei valori di ϵ che siano più piccoli del valore di soglia $\bar{\epsilon}$. Questo può essere fatto utilizzando degli algoritmi che incorporano al loro interno delle regole per l'aggiornamento automatico del parametro di penalità. Nel seguito descriveremo brevemente un modello di algoritmo basato sull'utilizzo della funzione di penalità esatta $P_G(x; \epsilon)$. Questo modello fa uso di una applicazione punto insieme $\mathcal{M} : \mathcal{A} \rightarrow 2^{\mathcal{A}}$ che schematizza una iterazione di un generico metodo di minimizzazione non vincolata. L'unica richiesta che deve soddisfare l'applicazione \mathcal{M} è la seguente:

Assunzione 6 Per ogni $\epsilon > 0$ e per qualunque punto $x_o \in \mathcal{A}$ ogni punto x_k prodotto da \mathcal{M} deve appartenere all'insieme di livello $\Omega(x_o; \epsilon) = \{x \in \mathcal{A} : P_G(x; \epsilon) \leq P_G(x_o; \epsilon)\}$ ed ogni punto di accumulazione della sequenza di punti $\{x_k\}$ prodotta da \mathcal{M} deve essere un punto stazionario di $P_G(x; \epsilon)$.

Algoritmo EXTPEN

Dati: $\epsilon > 0, \beta > 1, \delta > 0, \tilde{x} \in \mathbb{R}^n$.

Step 0: Scegli $\alpha > 0$ tale che $\tilde{x} \in \mathcal{A}$ e si pone $z = \tilde{x}$.

Step 1: Poni $k = 0$.

Se $P_G(\tilde{x}; \epsilon) \leq P_G(z; \epsilon)$ poni $x_o = \tilde{x}$; altrimenti $x_o = z$.

Step 2: (Test sul parametro di penalità) Se

$$\begin{aligned} & \|\nabla P_G(x_k; \epsilon)\|^2 + \|\nabla g(x_k)^\top \nabla P_G(x_k; \epsilon)\|^2 \\ & < \delta \sum_{i=1}^m \max\{g_i(x_k), -\epsilon p_i(x_k; \epsilon) \lambda_i(x_k)/2\}^2, \end{aligned}$$

allora $\epsilon := \epsilon/\beta$, poni $z = x_k$ e vai allo Step 1.

Step 3: Calcola $x_{k+1} \in \mathcal{M}[x_k]$, poni $k = k + 1$.

Step 4: Se $(x_k, \lambda(x_k))$ è una coppia di KT per il problema (6) allora STOP; altrimenti vai allo Step 2.

Proposizione 23 Siano verificate le Assunzioni 5, 4 e 6. Allora l'algoritmo EXTPEN termina in qualche punto $x_\nu \in \mathcal{A}$ e $(x_\nu, \lambda(x_\nu))$ è una coppia di KT per il problema (6), oppure l'algoritmo produce una successione infinita $\{x_k\} \subseteq \mathcal{A}$ tale che ogni punto di accumulazione x^* fornisce una coppia di KT $(x^*, \lambda(x^*))$ del problema (6).

Per quanto concerne invece il soddisfacimento della Assunzione 6 da parte dell'applicazione \mathcal{M} , essa può essere facilmente soddisfatta modificando leggermente un qualunque algoritmo di minimizzazione non vincolata globalmente convergente. Infatti, poichè, come già detto, si ha che $P_G(x; \epsilon) \rightarrow +\infty$ per $x \rightarrow \partial\mathcal{A}$, l'unica cosa da assicurare è che i punti di prova prodotti da \mathcal{M} rimangano nell'insieme di livello $\{x \in \mathcal{A} : P_G(x; \epsilon) \leq P_G(x_0; \epsilon)\}$. Questo può essere effettuato semplicemente limitando, quando necessario, lo spostamento lungo la direzione di ricerca d_k . In questo modo tutti i punti generati apparterranno all'insieme di livello e non potrà mai succedere che una sottosuccessione di punti converga verso un punto $\bar{x} \in \partial\mathcal{A}$.

Un approccio alternativo per sfruttare dal punto di vista algoritmico le funzioni di penalità esatte continuamente differenziabili consiste nell'utilizzarle per definire delle modifiche globalmente convergenti di algoritmi di minimizzazione vincolata che presentano delle buone proprietà in termini di rapidità di convergenza (cioè sono superlinearmente o quadraticamente convergenti) ma sono solo localmente convergenti (cioè convergono solo se il punto iniziale è "sufficientemente" vicino ad una coppia di KT del problema vincolato).

5 Funzioni Lagrangiane aumentate esatte

Le funzioni di Penalità esatte viste nella precedente sezione si basano pesantemente sull'uso di funzioni moltiplicatrici $\mu(x), \lambda(x)$. Purtroppo, il calcolo di queste funzioni, come è evidenziato dalla Proposizione 7, richiede l'inversione di una matrice $(M(x))$ quadrata di ordine pari alla somma $m + q$ cioè pari al numero dei vincoli che definiscono la regione ammissibile. Inoltre, dalla espressione (48), si nota che spesso la matrice $M(x)$ è densa. Quindi, quando il numero dei vincoli è elevato potrebbe risultare estremamente oneroso calcolare il valore delle funzioni moltiplicatrici. Per ovviare a questo inconveniente pur continuando ad avere una funzione di merito esatta, si può fare ricorso all'uso di funzioni Lagrangiane aumentate esatte che non necessitano dell'uso di funzioni moltiplicatrici. Il prezzo da pagare per questo è quello che queste funzioni sono definite sullo spazio esteso delle variabili primali-duali cioè \mathfrak{R}^{n+m+q} .

Analogamente a quanto visto nel caso delle funzioni di penalità esatte, anche per definire funzioni Lagrangiane esatte si trae ispirazione dall'espressione (38) di una funzione Lagrangiana aumentata. Infatti, l'espressione originaria di una funzione di merito di questa classe è la seguente ([19], [20], [21], [22]):

$$L_E(x, \mu, \lambda; \epsilon) = f(x) + \frac{1}{\epsilon} \phi_E(x, \mu, \lambda; \epsilon) + \phi_L(x, \mu, \lambda), \quad (66)$$

dove, nuovamente, il termine di penalità ϕ_E è quello dato dalla (37) mentre il nuovo termine $\phi_L(x, \mu, \lambda)$ ha le seguenti caratteristiche:

- (i) $\phi_L(x, \mu, \lambda)$ è una funzione continuamente differenziabile nella tripla (x, μ, λ) ;
- (iii) se (x^*, μ^*, λ^*) è una tripla di KT per il problema (2), allora

$$\phi_L(x^*, \mu, \lambda) = 0$$

se e solo se $\mu = \mu^*$ e $\lambda = \lambda^*$.

Da un punto di vista strettamente teorico, per la funzione $L_E(x, \mu, \lambda; \epsilon)$ si possono stabilire le stesse proprietà di esattezza già viste per la funzione $P_E(x; \epsilon)$. Sotto l'aspetto più prettamente algoritmico, questa classe di funzioni Lagrangiane aumentate esatte soffre degli stessi problemi della funzione $P_E(x; \epsilon)$. Anche in questo caso, quindi, lo sforzo teorico è quello di modificare leggermente l'espressione della funzione $L_E(x, \mu, \lambda; \epsilon)$ per ottenere una nuova funzione di merito tale che l'intersezione tra i suoi insiemi di livello ed un certo aperto siano degli insiemi compatti.

5.1 Funzioni Lagrangiane aumentate esatte con termini di barriera ([1])

In questa ultima sezione descriveremo brevemente una funzione Lagrangiana aumentata esatta con termini di barriera L_G . Nell'ambito della classe delle funzioni Lagrangiane aumentate esatte, la funzione L_G riveste lo stesso ruolo che la funzione P_G aveva nell'ambito delle funzioni di penalità esatte continuamente differenziabili.

Anche in questa ultima sezione ci limiteremo a considerare problemi della forma (6) ovvero problemi con *soli vincoli di disuguaglianza*. Sia \mathcal{A} (definito come in (55)) un rilassamento aperto dell'insieme ammissibile \mathcal{F} , tale che $\tilde{x} \in \mathcal{A}$.

Su questo insieme aperto si può definire la funzione L_G nella seguente maniera [22]:

$$L_G(x, \lambda; \epsilon) = f(x) + \frac{1}{\epsilon} \phi_G(x, \lambda; \epsilon) + \phi_L(x, \lambda), \quad (67)$$

dove la funzione $\phi_G(x, \lambda; \epsilon)$ è data da (57) con

$$p_i(x, \lambda_i; \epsilon) = p(x, \lambda) = \frac{a(x)}{1 + \|\lambda\|^2},$$

ed il termine $\phi_L(x, \lambda)$ è dato dalla

$$\phi_L(x, \lambda) = \|\nabla g(x)^\top \nabla_x L(x, \lambda) + G(x)^2 \lambda\|^2. \quad (68)$$

Notiamo anzitutto che il termine ϕ_L definito in (68) gode delle proprietà (i) e (ii) viste poco sopra. infatti ϕ_L è una funzione continuamente differenziabile ed in più, se vale l'Assunzione 2, si ha che, se (x^*, λ^*) è una coppia di KT del problema (2), $\phi_L(x^*, \lambda) = 0$ se e solo se $\lambda = \lambda^*$.

Dati $\tilde{x} \in \mathcal{A}$ e $\tilde{\lambda} \in \mathfrak{R}^m$, si definisce il seguente insieme di livello relativo ad \mathcal{A} :

$$\mathcal{L}(\tilde{x}, \tilde{\lambda}; \epsilon) = \{(x, \lambda) \in \mathcal{A} \times \mathfrak{R}^m : L_G(x, \lambda; \epsilon) \leq L_G(\tilde{x}, \tilde{\lambda}; \epsilon)\}. \quad (69)$$

Le proposizione che segue ([22]) riassume le proprietà di compattezza dell'insieme di livello $\mathcal{L}(\tilde{x}, \tilde{\lambda}; \epsilon)$.

Proposizione 24 *Sia verificata l'Assunzione 4. Allora,*

- (a) *per ogni $\epsilon_M > 0$, esiste un insieme compatto $\mathcal{C} \subset \mathfrak{R}^n$ tale che $\mathcal{L}(\tilde{x}, \tilde{\lambda}; \epsilon) \subseteq \mathcal{C} \times \mathfrak{R}^m$ per ogni $\epsilon \in (0, \epsilon_M]$;*
- (b) *se è verificata l'Assunzione 2. Allora, per ogni $\epsilon > 0$, l'insieme di livello $\mathcal{L}(\tilde{x}, \tilde{\lambda}; \epsilon)$ è compatto.*

Nel seguito descriveremo brevemente un modello di algoritmo basato sull'utilizzo della funzione Lagrangiana aumentata esatta $L_G(x, \lambda; \epsilon)$. Questo modello fa uso di una applicazione punto insieme $\mathcal{T} : \mathcal{A} \times \mathfrak{R}^m \rightarrow 2^{\mathcal{A} \times \mathfrak{R}^m}$ che schematizza una iterazione di un generico metodo di minimizzazione non vincolata. L'unica richiesta che deve soddisfare l'applicazione \mathcal{T} è la seguente:

Assunzione 7 *Per ogni $\epsilon > 0$ e per qualunque coppia $(x_o, \lambda_o) \in \mathcal{A} \times \mathfrak{R}^m$ ogni coppia (x_k, λ_k) prodotta da \mathcal{T} deve appartenere all'insieme di livello $\mathcal{L}(x_o, \lambda_o; \epsilon)$ ed ogni punto di accumulazione della sequenza $\{(x_k, \lambda_k)\}$ prodotta da \mathcal{T} deve essere un punto stazionario di $L_G(x, \lambda; \epsilon)$.*

Algoritmo EXTLAGR

Dati: $\epsilon > 0$, $\beta > 1$, $\delta > 0$, $\tilde{x} \in \mathfrak{R}^n$, $\tilde{\lambda} \in \mathfrak{R}^m$.

Step 0: Scegli $\alpha > 0$ tale che $\tilde{x} \in \mathcal{A}$ e si pone $z = (\tilde{x}, \tilde{\lambda})$.

Step 1: Poni $k = 0$.

Se $L_G(\tilde{x}, \tilde{\lambda}; \epsilon) \leq L_G(z; \epsilon)$ poni $(x_o, \lambda_o) = (\tilde{x}, \tilde{\lambda})$;
altrimenti $(x_o, \lambda_o) = z$.

Step 2: (*Test sul parametro di penalità*) Se

$$\|\nabla L_G(x_k, \lambda_k; \epsilon)\|^2 < \epsilon \|\max\{g(x_k), -\epsilon p(x_k, \lambda_k)\lambda_k\}\|^2,$$

allora $\epsilon := \epsilon/\beta$, poni $z = (x_k, \lambda_k)$ e vai allo Step 1.

Step 3: Calcola $(x_{k+1}, \lambda_{k+1}) \in \mathcal{T}[x_k, \lambda_k]$, poni $k = k + 1$.

Step 4: Se (x_k, λ_k) è una coppia di KT per il problema (6)
allora STOP; altrimenti vai allo Step 2.

Proposizione 25 *Siano verificate le Assunzioni 5, 4, 2 e 7. Allora l'algoritmo EXTLAGR termina in qualche coppia $(x_\nu, \lambda_\nu) \in \mathcal{A} \times \mathfrak{R}^m$ e (x_ν, λ_ν) è una coppia di KT per il problema (6), oppure l'algoritmo produce una successione infinita $\{(x_k, \lambda_k)\} \subseteq \mathcal{A} \times \mathfrak{R}^m$ tale che ogni punto di accumulazione (x^*, λ^*) è una coppia di KT del problema (6).*

6 Ringraziamenti

Un ringraziamento particolare va a Gianni Di Pillo e Stefano Lucidi che con i loro manoscritti ([2], [1] e [3]) e con numerosi e preziosi suggerimenti e indicazioni hanno notevolmente contribuito alla stesura di queste note. Ringrazio inoltre gli studenti del corso di Ottimizzazione Continua (a.a. 2003-04) ed in particolare gli studenti dei dottorati di ricerca in Ingegneria dei Sistemi ed in Ricerca Operativa per l'attenzione e la passione con la quale hanno seguito le lezioni e per l'avermi segnalato errori e sviste, alcune delle quali, certamente ancora presenti in questa versione delle note.

References

- [1] S. Lucidi. Algoritmi di programmazione non lineare. In G. Di Pillo, editor, *Metodi di Ottimizzazione per le Decisioni*, pages 117–190. Mason, editoriale ESA, 1993.
- [2] G. Di Pillo. Metodi per la soluzione di problemi di programmazione non lineare. Dispense del corso di Ottimizzazione.
- [3] G. Di Pillo and S. Lucidi. On exact augmented lagrangian functions in nonlinear programming. In G. Di Pillo and F. Giannessi, editors, *Nonlinear Optimization and Applications*, pages 85–100. Plenum Press, New York, 1996.
- [4] A. V. Fiacco and G. P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. John Wiley and Sons, New York, 1969.
- [5] M. Avriel. *Nonlinear Programming: Analysis and Methods*. Prentice-Hall, Englewood Cliffs, 1976.
- [6] R. Fletcher. *Practical Methods of Optimization*. John Wiley and Sons, New York, 1987.
- [7] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, Massachusetts, 1995.
- [8] M. R. Hestenes. Multiplier and gradient methods. *Journal of Optimization Methods and Applications*, 4:303–320, 1969.
- [9] M. J. D. Powell. A method for nonlinear constraints in minimization problems. In R. Fletcher, editor, *Optimization*, pages 283–298. Academic Press, New York, 1969.
- [10] R.T. Rockafellar. The multiplier method of hestenes and powell applied to convex programming. *J. Optim. Theory and Appl.*, 12:555–562, 1973.

- [11] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, Berlin, 1999.
- [12] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. *LANCELOT: a Fortran package for large-scale nonlinear optimization*, volume 17 of *Springer Series in Computational Mathematics*. Springer Verlag, Heidelberg, New York, 1992.
- [13] T. Glad and E. Polak. A multiplier method with automatic limitation of penalty growth. *Math Programming*, 17:140–155, 1979.
- [14] G. Di Pillo and L. Grippo. An exact penalty method with global convergence properties. *Math Programming*, 36:1–18, 1986.
- [15] G. Di Pillo and L. Grippo. Exact penalty functions in constrained optimization. *SIAM J. Control and Optimization*, 27:1333–1360, 1989.
- [16] S. Lucidi. New results on a continuously differentiable exact penalty function. *SIAM J. Optimization*, 2:558–574, 1992.
- [17] G. Contaldi, G. Di Pillo, and S. Lucidi. A continuously differentiable exact penalty function for nonlinear programming problems with unbounded feasible set. *Operations Research Letters*, 14:153–161, 1993.
- [18] G. Di Pillo and L. Grippo. A continuously differentiable exact penalty function for nonlinear programming problems with inequality constraints. *SIAM J. Control and Optimization*, 23:72–84, 1985.
- [19] G. Di Pillo and L. Grippo. A new class of augmented lagrangians in nonlinear programming. *SIAM J. Control and Optimization*, 17:495–519, 1979.
- [20] G. Di Pillo and L. Grippo. An augmented lagrangian for inequality constraints in nonlinear programming problems. *J. Optim. Theory and Appl.*, 36:495–519, 1982.
- [21] S. Lucidi. New results on a class of exact augmented lagrangians. *J. Optim. Theory and Appl.*, 58:259–282, 1988.
- [22] G. Di Pillo and S. Lucidi. An augmented lagrangian function with improved exactness properties. *SIAM J. Optimization*, 12(2):376–406, 2001.