Ottimizzazione dei Sistemi Complessi

G. Liuzzi[1]

Giovedì 23 Marzo 2017

[1] Istituto di Analisi dei Sistemi ed Informatica IASI - CNR

Definizione

Teoricamente è possibile definire

$$q_\infty(x) = \begin{cases} 0 & \text{se } x \in \mathcal{F} \\ +\infty & \text{se } x \notin \mathcal{F} \end{cases}$$

e quindi

$$P_\infty(x) = f(x) + q_\infty(x)$$

È evidente che:

- risolvere il problema $\min\ P_\infty(x)$ fornisce soluzioni del problema (P_0).
- minimizzare $P_\infty(x)$ è (o può essere) molto "complicato"

Definizione

Teoricamente è possibile definire

$$q_\infty(x) = \begin{cases} 0 & \text{se } x \in \mathcal{F} \\ +\infty & \text{se } x \notin \mathcal{F} \end{cases}$$

e quindi

$$P_\infty(x) = f(x) + q_\infty(x)$$

È evidente che:

- risolvere il problema min $P_\infty(x)$ fornisce soluzioni del problema (P_0).
- minimizzare $P_\infty(x)$ è (o può essere) molto "complicato"

Definizione

Anziché $q_\infty(x)$, definiamo

$$q(x) = \begin{cases} 0 & \text{se } x \in \mathcal{F} \\ > 0 & \text{se } x \notin \mathcal{F} \end{cases}$$

e quindi

$$P_\epsilon(x) = f(x) + \frac{1}{\epsilon} q(x)$$

N.B.:

- se $f(x)$ e $q(x)$ sono cont. differenziabili, allora anche $P_\epsilon(x)$ lo è
- per ogni $x \in \mathbb{R}^n$, $\lim_{\epsilon \to 0} P_\epsilon(x) = P_\infty(x)$

Espressione di q(x)

Il termine $q(x)$ deve "penalizzare con continuità" l'eventuale inammissibilità di x

P.es., con riferimento a (P_0),

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

- per nessun valore di $\epsilon > 0$, min $P_\epsilon(x)$ è equivalente a (P_0)
- per ogni valore di $\epsilon > 0$, una min. non vincolata di $P_\epsilon(x)$ produce (in genere) un punto $x^* \notin \mathcal{F}$, cioè $q(x^*) > 0$

Algoritmo di soluzione

Algoritmo `SEQPEN`

INPUT: $\texttt{maxit}$, $\tau > 0$, $k = 0$, $\epsilon_0 > 0$

for $k = 0, 1, \ldots, \texttt{maxit}$

Trova una soluzione $(x(\epsilon_k))$ del problema

$$\min_{x \in \mathbb{R}^n} P_{\epsilon_k}(x). \tag{1}$$

if $q(x(\epsilon_k)) \leq \tau$ **then** STOP

$\epsilon_{k+1} = \epsilon_k / 2$

endfor

OUTPUT: una sol. approssimata $x^* = x(\epsilon_k)$

Proprietà di convergenza

Sia $\mathcal{F}_\sigma = \{x \in \mathbb{R}^n : \ |h(x)| \leq \sigma, \ g(x) \leq \sigma\}$ insieme ammissibile "rilassato" $(\mathcal{F} \subset \mathcal{F}_\sigma)$

Teorema (Conv. a minimi globali)

Siano soddisfatte le ipotesi seguenti:

- *esiste un* $\sigma > 0$ *t.c.* $\mathcal{F}_\sigma$ *è compatto;*
- *per ogni* k, *la funzione* $P_{\epsilon_k}(x)$ *ammette minimo globale* x^k.

Allora, $\{x^k\}$ *ammette punti limite ed ogni punto limite è un minimo globale di* (P_0)

Esempio 1

Consideriamo il seguente problema vincolato (banale)

$$\min\ x^2$$
$$x \geq 1$$

Ammete $x^* = 1$ come **unico** punto di minimo globale

La funzione di penalità esterna è:

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

Esempio 1

Consideriamo il seguente problema vincolato (banale)

$$\min\ x^2$$
$$x \geq 1$$

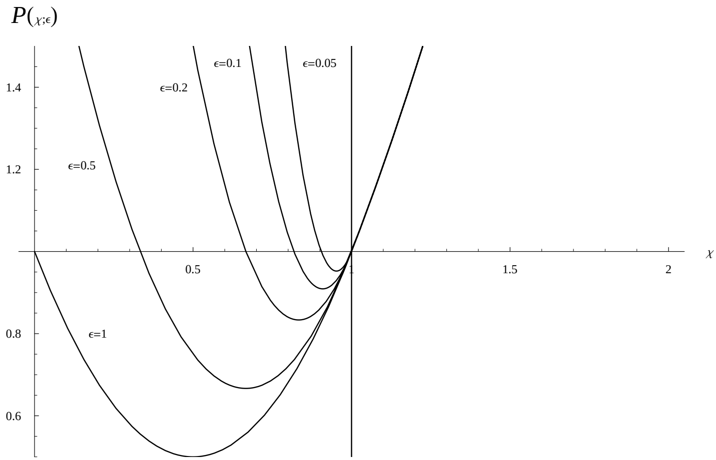
Ammete $x^* = 1$ come **unico** punto di minimo globale

La funzione di penalità esterna è:

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

Esempio 1

Grafici di $P_\epsilon(x)$ ($\epsilon = 1, 0.5, 0.2, 0.1, 0.05$)



Esempio 2

Consideriamo il seguente problema vincolato (banale)

$$\min \; -x^4$$
$$x = 1$$

Ammete (ovviamente) $x^* = 1$ come **unico** punto di minimo globale

La funzione di penalità esterna è:

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

Esempio 2

Consideriamo il seguente problema vincolato (banale)

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

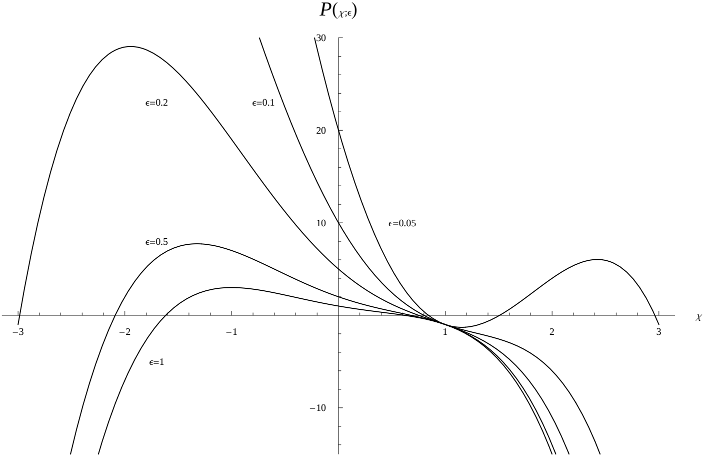
Ammete (ovviamente) $x^* = 1$ come **unico** punto di minimo globale

La funzione di penalità esterna è:

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

Esempio 2

Grafici di $P_\epsilon(x)$ ($\epsilon = 1, 0.5, 0.2, 0.1, 0.05$)



Algoritmo SEQPEN modificato

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

INPUT: $\texttt{maxit}$, $\{\tau_k\} \to 0$, $\rho > 0$, $k = 0$, $\epsilon_0 > 0$

for $k = 0, 1, \dots, \texttt{maxit}$

Calcola $x(\epsilon_k)$ t.c. $\|\nabla P_{\epsilon_k}(x(\epsilon_k))\| \le \tau_k$

if $\tau_k < \rho$ **and** $q(x(\epsilon_k)) \le \rho$ **then** STOP

$\epsilon_{k+1} = \epsilon_k / 2$

endfor

OUTPUT: una sol. approssimata $x^* = x(\epsilon_k)$

Proprietà di convergenza

Teorema

Sia $\{x_k\}$ *una successione t.c.* $\|\nabla P_{\epsilon_k}(x_k)\| \leq \tau_k$ *e* $\{x_k\}_K$ *una sottosucc. convergente ad* x^* *t.c. siano lin. indipendenti*

$$\nabla g_i(x^*),\ i \in I_0(x^*),\ \nabla h_j(x^*),\ \forall\, j.$$

Allora, x^* *è un punto di KKT di* (P_0) *con molt.* λ^* *e* μ^* *t.c.*

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

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

Esempio 3

Consideriamo il seguente problema vincolato (banale)

$$\begin{array}{c} \min \ -x - y \\ x^2 + y^2 = 1 \end{array}$$

Ammete $(x^*, y^*) = (1,\ 1)/\sqrt{2}$ come **unico** punto di minimo globale

La funzione di penalità esterna è:

$$P_\epsilon(x) = -x - y + \frac{1}{\epsilon}(1 - x^2 - y^2)^2$$

Esempio 3

Consideriamo il seguente problema vincolato (banale)

$$\min \; -x-y$$
$$x^2+y^2=1$$

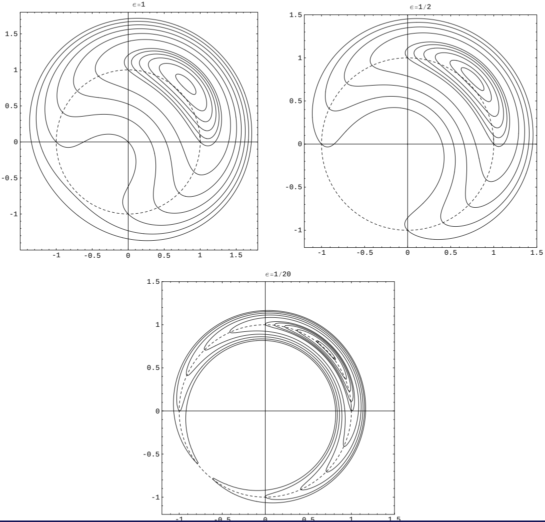
Ammete $(x^*, y^*) = (1,\ 1)/\sqrt{2}$ come **unico** punto di minimo globale

La funzione di penalità esterna è:

$$P_\epsilon(x) = -x - y + \frac{1}{\epsilon}(1 - x^2 - y^2)^2$$

Esempio 3

Grafici di $P_\epsilon(x)$ ($\epsilon = 1, 0.5, 0.05$)



Tanto per cominciare

Consideriamo il problema con soli vincoli di uguaglianza

$$\begin{array}{ll} \min & f(x) \\ c.v. & h(x) = 0 \end{array}$$

Funzione Lagrangiana: $L(x, \mu) = f(x) + \mu^\top h(x)$

con gradiente (vi ricordo):

$$\begin{array}{rcl} \nabla_x L(x, \mu) & = & \nabla f(x) + \nabla h(x)\mu \\ \nabla_\mu L(x, \mu) & = & h(x) \end{array}$$

Percui, le C.N. di ottimo si possono riassumere in: $\nabla L(x, \mu) = 0$

Tanto per cominciare

Consideriamo il problema con soli vincoli di uguaglianza

$$\begin{array}{ll} \min & f(x) \\ c.v. & h(x) = 0 \end{array}$$

Funzione Lagrangiana: $L(x, \mu) = f(x) + \mu^\top h(x)$

con gradiente (vi ricordo):

$$\begin{array}{rcl} \nabla_x L(x, \mu) & = & \nabla f(x) + \nabla h(x)\mu \\ \nabla_\mu L(x, \mu) & = & h(x) \end{array}$$

Percui, le C.N. di ottimo si possono riassumere in: $\nabla L(x, \mu) = 0$

Tanto per cominciare

Consideriamo il problema con soli vincoli di uguaglianza

$$\begin{array}{ll} \min & f(x) \\ c.v. & h(x) = 0 \end{array}$$

Funzione Lagrangiana: $L(x,\mu) = f(x) + \mu^\top h(x)$

con gradiente (vi ricordo):

$$\begin{array}{rcl} \nabla_x L(x,\mu) & = & \nabla f(x) + \nabla h(x)\mu \\ \nabla_\mu L(x,\mu) & = & h(x) \end{array}$$

Percui, le C.N. di ottimo si possono riassumere in: $\nabla L(x,\mu) = 0$

Tanto per cominciare

Consideriamo il problema con soli vincoli di uguaglianza

$$\begin{array}{ll} \min & f(x) \\ c.v. & h(x) = 0 \end{array}$$

Funzione Lagrangiana: $L(x, \mu) = f(x) + \mu^\top h(x)$

con gradiente (vi ricordo):

$$\begin{array}{rcl} \nabla_x L(x, \mu) & = & \nabla f(x) + \nabla h(x)\mu \\ \nabla_\mu L(x, \mu) & = & h(x) \end{array}$$

Percui, le C.N. di ottimo si possono riassumere in: $\nabla L(x, \mu) = 0$

Quindi ...

Punti che soddisfano le C.N. di ottimo sono punti staz. della funzione Lagrangiana $L(x, \mu)$!!

Purtroppo però, accade che le soluzione del problema originario corrispondono (spesso) con punti di "sella" della funzione Lagrangiana.

Quindi, non si può (banalmente) minimizzare $L(x, \mu)$ per risolvere il problema originario

Quindi ...

Punti che soddisfano le C.N. di ottimo sono punti staz. della funzione Lagrangiana $L(x, \mu)$!!

Purtroppo però, accade che le soluzione del problema originario corrispondono (spesso) con punti di "sella" della funzione Lagrangiana.

Quindi, non si può (banalmente) minimizzare $L(x, \mu)$ per risolvere il problema originario

Quindi ...

Punti che soddisfano le C.N. di ottimo sono punti staz. della funzione Lagrangiana $L(x, \mu)$!!

Purtroppo però, accade che le soluzione del problema originario corrispondono (spesso) con punti di "sella" della funzione Lagrangiana.

Quindi, non si può (banalmente) minimizzare $L(x, \mu)$ per risolvere il problema originario

Fortunatamente ...

Possiamo "convessificare" $L(x, \mu)$ nell'intorno delle soluzione del problema originario (semplicemente) aggiungendo un termine di penalità, con $\epsilon > 0$ "sufficientemente" piccolo

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

con gradiente

$$\begin{aligned} \nabla_x L_a(x, \mu; \epsilon) &= \nabla_x L(x, \mu) + \frac{2}{\epsilon}\nabla h(x) h(x) \\ \nabla_\mu L_a(x, \mu; \epsilon) &= \nabla_\mu L(x, \mu) = h(x) \end{aligned}$$

Fortunatamente ...

Possiamo "convessificare" $L(x, \mu)$ nell'intorno delle soluzione del problema originario (semplicemente) aggiungendo un termine di penalità, con $\epsilon > 0$ "sufficientemente" piccolo

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

con gradiente

$$\begin{aligned} \nabla_x L_a(x, \mu; \epsilon) &= \nabla_x L(x, \mu) + \frac{2}{\epsilon}\nabla h(x) h(x) \\ \nabla_\mu L_a(x, \mu; \epsilon) &= \nabla_\mu L(x, \mu) = h(x) \end{aligned}$$

Prima proprietà

Proposizione

Per ogni valore $\epsilon > 0$, ogni punto staz. di $L_a(x, \mu; \epsilon)$ è un punto di KKT del problema originario e viceversa

Dim.: Sia $(\bar{x}, \bar{\mu})$ un punto t.c. $\nabla L_a(\bar{x}, \bar{\mu}) = 0$. Allora, dalla espr. del gradiente risulta, banalmente: $h(\bar{x}) = 0$ e quindi $\nabla_x L(\bar{x}, \bar{\mu}) = 0$, ovvero $(\bar{x}, \bar{\mu})$ è un punto di KKT.

Supponiamo ora che $(\bar{x}, \bar{\mu})$ sia un punto di KKT e quindi, in particolare, $h(\bar{x}) = 0$ e $\nabla_x L(\bar{x}, \bar{\mu}) = 0$. Allora, dalla espr. del gradiente di L_a si ricava nuovamente $\nabla L_a(\bar{x}, \bar{\mu}) = 0$. $\square$

Prima proprietà

Proposizione

Per ogni valore $\epsilon > 0$, ogni punto staz. di $L_a(x, \mu; \epsilon)$ è un punto di KKT del problema originario e viceversa

Dim.: Sia $(\bar{x}, \bar{\mu})$ un punto t.c. $\nabla L_a(\bar{x}, \bar{\mu}) = 0$. Allora, dalla espr. del gradiente risulta, banalmente: $h(\bar{x}) = 0$ e quindi $\nabla_x L(\bar{x}, \bar{\mu}) = 0$, ovvero $(\bar{x}, \bar{\mu})$ è un punto di KKT.

Supponiamo ora che $(\bar{x}, \bar{\mu})$ sia un punto di KKT e quindi, in particolare, $h(\bar{x}) = 0$ e $\nabla_x L(\bar{x}, \bar{\mu}) = 0$. Allora, dalla espr. del gradiente di L_a si ricava nuovamente $\nabla L_a(\bar{x}, \bar{\mu}) = 0$. $\square$

Esempio

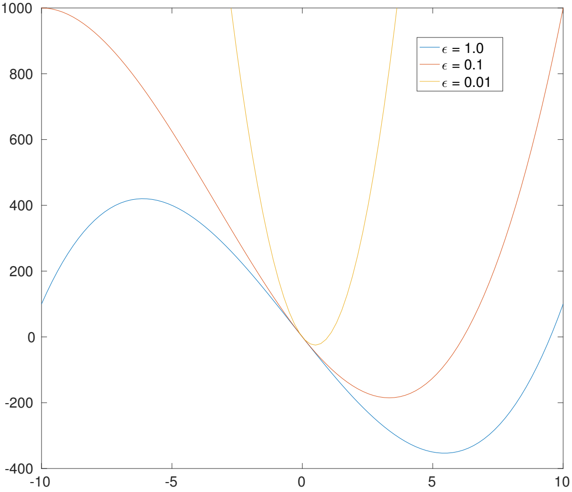
Consideriamo il problema

$$\begin{aligned} &\min\ x^3 \\ &s.t.\ x = 0 \end{aligned}$$

per cui risulta (banalmente) $x^* = 0$ e $\mu^* = 0$

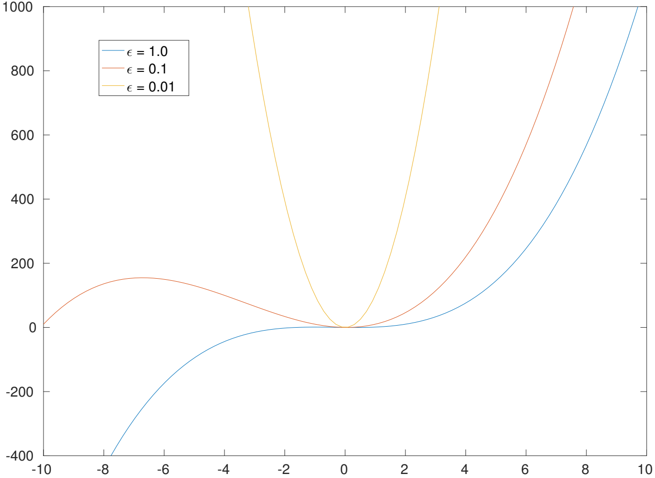
Esempio

Andamenti della funzione $L_a(x, -100; \epsilon)$ per $\epsilon = 1, 0.1, 0.01$



Esempio

Andamenti della funzione $L_a(x, -1; \epsilon)$ per $\epsilon = 1, 0.1, 0.01$



Seconda proprietà

È possibile "minimizzare $L_a(x, \mu)$" per determinare minimi locali del problema originario?

No, $L_a(x, \mu)$ è lineare in μ

Si può tentare di minimizzare $L_a(x, \bar{\mu})$ per μ fissato

Cosa manca?

Seconda proprietà

È possibile "minimizzare $L_a(x, \mu)$" per determinare minimi locali del problema originario?

No, $L_a(x, \mu)$ è lineare in μ

Si può tentare di minimizzare $L_a(x, \bar{\mu})$ per μ fissato

Cosa manca?

Seconda proprietà

È possibile "minimizzare $L_a(x, \mu)$" per determinare minimi locali del problema originario?

No, $L_a(x, \mu)$ è lineare in μ

Si può tentare di minimizzare $L_a(x, \bar{\mu})$ per μ fissato

Cosa manca?

Seconda proprietà

È possibile "minimizzare $L_a(x, \mu)$" per determinare minimi locali del problema originario?

No, $L_a(x, \mu)$ è lineare in μ

Si può tentare di minimizzare $L_a(x, \bar{\mu})$ per μ fissato

Cosa manca?

Seconda proprietà

Sia $\bar{x}$ una sol. del problema vincolato originale tale che:

- $\bar{x}$ è un punto regolare per i vincoli
- la coppia di KKT $(\bar{x}, \bar{\mu})$ soddisfa le C.S. del II ordine

Proposizione

Per **valori sufficientemente piccoli di** $\epsilon > 0$, $\bar{x}$ *è un minimo locale stretto di* $L_a(x, \bar{\mu}; \epsilon)$

N.B. è necessario conoscere $\bar{\mu}$!

Seconda proprietà

Sia $\bar{x}$ una sol. del problema vincolato originale tale che:

- $\bar{x}$ è un punto regolare per i vincoli
- la coppia di KKT $(\bar{x}, \bar{\mu})$ soddisfa le C.S. del II ordine

Proposizione

Per **valori sufficientemente piccoli di** $\epsilon > 0$, $\bar{x}$ *è un minimo locale stretto di* $L_a(x, \bar{\mu}; \epsilon)$

N.B. è necessario conoscere $\bar{\mu}$!

Seconda proprietà

Sia $\bar{x}$ una sol. del problema vincolato originale tale che:

- $\bar{x}$ è un punto regolare per i vincoli
- la coppia di KKT $(\bar{x}, \bar{\mu})$ soddisfa le C.S. del II ordine

Proposizione

Per **valori sufficientemente piccoli di** $\epsilon > 0$, $\bar{x}$ *è un minimo locale stretto di* $L_a(x, \bar{\mu}; \epsilon)$

N.B. è necessario conoscere $\bar{\mu}$!

Esempio

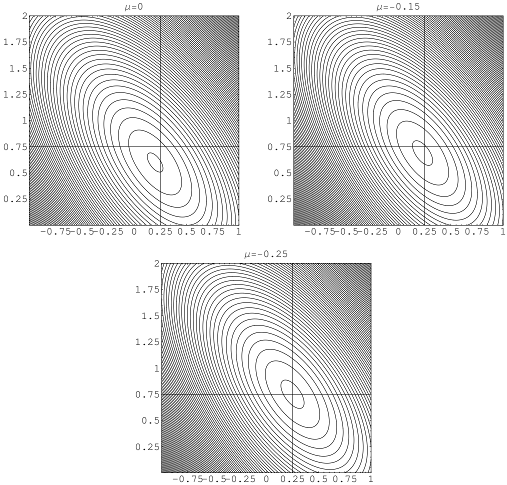
Consideriamo il problema

$$\min \ \frac{1}{2}x^2 + \frac{1}{6}y^2$$
$$s.t. \ x + y = 1$$

Calcolare la soluzione del problema (x^*, y^*, μ^*)

Esempio

Curve di livello di $L_a(x, y, \bar{\mu}; \epsilon)$ per $\epsilon = 1$ a $\bar{\mu} = 0, -0.15, -0.25$



Metodo di soluzione

Algoritmo `SEQLAGR`

Dati: $\epsilon_\circ > 0$, $\beta > 1$, $\{\tau_k\} \to 0$, $\{\eta_k\} \to 0$, $\rho > 0$, $\mu_\circ$, `maxit`

for $k = 0, 1, \ldots,$`maxit`

Calcola x_k t.c. $\|\nabla L_a(x_k, \mu_k; \epsilon_k)\| \leq \tau_k$

if $\|\nabla L_a(x_k, \mu_k; \epsilon_k)\| < \rho$ **then**

$x^* \leftarrow x_k$, $\mu^* \leftarrow \mu_k$, STOP

endif

endfor

Return: miglior coppia trovata (x^*, μ^*)

Metodo di soluzione

Algoritmo `SEQLAGR`

Dati: $\epsilon_\circ > 0$, $\beta > 1$, $\{\tau_k\} \to 0$, $\{\eta_k\} \to 0$, $\rho > 0$, $\mu_\circ$, `maxit`

for $k = 0, 1, \ldots,$`maxit`

Calcola x_k t.c. $\|\nabla L_a(x_k, \mu_k; \epsilon_k)\| \le \tau_k$

if $\|\nabla L_a(x_k, \mu_k; \epsilon_k)\| < \rho$ **then**

$x^* \leftarrow x_k$, $\mu^* \leftarrow \mu_k$, STOP

endif

if $\|h(x_{k+1})\| \le \eta_k$ **then** $\epsilon_{k+1} = \epsilon_k$

else $\epsilon_{k+1} = \epsilon_k / \beta$

Calcola μ_{k+1}

endfor

Return: miglior coppia trovata (x^*, μ^*)

Metodo di soluzione

Algoritmo `SEQLAGR`

Dati: $\epsilon_\circ > 0$, $\beta > 1$, $\{\tau_k\} \to 0$, $\{\eta_k\} \to 0$, $\rho > 0$, $\mu_\circ$, `maxit`

for $k = 0, 1, \ldots,$`maxit`

Calcola x_k t.c. $\|\nabla L_a(x_k, \mu_k; \epsilon_k)\| \leq \tau_k$

if $\|\nabla L_a(x_k, \mu_k; \epsilon_k)\| < \rho$ **then**

$x^* \leftarrow x_k$, $\mu^* \leftarrow \mu_k$, STOP

endif

if $\|h(x_{k+1})\| \leq \eta_k$ **then** $\epsilon_{k+1} = \epsilon_k$

else $\epsilon_{k+1} = \epsilon_k / \beta$

Calcola μ_{k+1}

endfor

Return: miglior coppia trovata (x^*, μ^*)

Metodo di soluzione

Algoritmo `SEQLAGR`

Dati: $\epsilon_\circ > 0$, $\beta > 1$, $\{\tau_k\} \to 0$, $\{\eta_k\} \to 0$, $\rho > 0$, $\mu_\circ$, `maxit`

for $k = 0, 1, \dots,$`maxit`

  Calcola x_k t.c. $\|\nabla L_a(x_k, \mu_k; \epsilon_k)\| \le \tau_k$

  if $\|\nabla L_a(x_k, \mu_k; \epsilon_k)\| < \rho$ **then**

    $x^* \leftarrow x_k$, $\mu^* \leftarrow \mu_k$, STOP

  endif

  if $\|h(x_{k+1})\| \le \eta_k$ **then** $\epsilon_{k+1} = \epsilon_k$

  else $\epsilon_{k+1} = \epsilon_k / \beta$

  Calcola μ_{k+1}

endfor

Return: miglior coppia trovata (x^*, μ^*)

Aggiornamento del moltiplicatore – 1

Come si passa da μ_k a μ_{k+1}?

Consideriamo la *funzione duale*

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

ove, per ϵ suff. piccoli e per ogni μ, $x(\mu)$ è **minimo locale** di L_a quindi (stazionario e) tale che

$$\nabla_x L_a(x(\mu), \mu; \epsilon) = 0.$$

Aggiornamento del moltiplicatore – 1

Come si passa da μ_k a μ_{k+1}?

Consideriamo la *funzione duale*

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

ove, per ϵ suff. piccoli e per ogni μ, $x(\mu)$ è **minimo locale** di L_a quindi (stazionario e) tale che

$$\nabla_x L_a(x(\mu), \mu; \epsilon) = 0.$$

Aggiornamento del moltiplicatore – 2

Supponiamo che x^* sia soluzione del problema originale (quindi punto di KKT) con moltiplicatore μ^*

Per le prop. di L_a risulta: $x(\mu^*) = x^*$ e $h(x^*) = 0$, quindi

$$\psi(\mu^*) = L_a(x^*, \mu^*; \epsilon) = f(x^*) = L_a(x^*, \mu; \epsilon) \geq L_a(x(\mu), \mu; \epsilon) = \psi(\mu)$$

Cioè, μ^* è un punto di **massimo locale** della funzione duale $\psi(\mu)$

Quindi ... per determinare μ^* è (teoricamente) possibile minimizzare la funzione $-\psi(\mu)$

Aggiornamento del moltiplicatore – 2

Supponiamo che x^* sia soluzione del problema originale (quindi punto di KKT) con moltiplicatore μ^*

Per le prop. di L_a risulta: $x(\mu^*) = x^*$ e $h(x^*) = 0$, quindi

$$\psi(\mu^*) = L_a(x^*, \mu^*; \epsilon) = f(x^*) = L_a(x^*, \mu; \epsilon) \geq L_a(x(\mu), \mu; \epsilon) = \psi(\mu)$$

Cioè, μ^* è un punto di **massimo locale** della funzione duale $\psi(\mu)$

Quindi ... per determinare μ^* è (teoricamente) possibile minimizzare la funzione $-\psi(\mu)$

Aggiornamento del moltiplicatore – 2

Supponiamo che x^* sia soluzione del problema originale (quindi punto di KKT) con moltiplicatore μ^*

Per le prop. di L_a risulta: $x(\mu^*) = x^*$ e $h(x^*) = 0$, quindi

$$\psi(\mu^*) = L_a(x^*, \mu^*; \epsilon) = f(x^*) = L_a(x^*, \mu; \epsilon) \geq L_a(x(\mu), \mu; \epsilon) = \psi(\mu)$$

Cioè, μ^* è un punto di **massimo locale** della funzione duale $\psi(\mu)$

Quindi ... per determinare μ^* è (teoricamente) possibile minimizzare la funzione $-\psi(\mu)$

Aggiornamento del moltiplicatore – 2

Supponiamo che x^* sia soluzione del problema originale (quindi punto di KKT) con moltiplicatore μ^*

Per le prop. di L_a risulta: $x(\mu^*) = x^*$ e $h(x^*) = 0$, quindi

$$\psi(\mu^*) = L_a(x^*, \mu^*; \epsilon) = f(x^*) = L_a(x^*, \mu; \epsilon) \geq L_a(x(\mu), \mu; \epsilon) = \psi(\mu)$$

Cioè, μ^* è un punto di **massimo locale** della funzione duale $\psi(\mu)$

Quindi ... per determinare μ^* è (teoricamente) possibile minimizzare la funzione $-\psi(\mu)$

Aggiornamento del moltiplicatore – 2

Supponiamo che x^* sia soluzione del problema originale (quindi punto di KKT) con moltiplicatore μ^*

Per le prop. di L_a risulta: $x(\mu^*) = x^*$ e $h(x^*) = 0$, quindi

$$\psi(\mu^*) = L_a(x^*, \mu^*; \epsilon) = f(x^*) = L_a(x^*, \mu; \epsilon) \geq L_a(x(\mu), \mu; \epsilon) = \psi(\mu)$$

Cioè, μ^* è un punto di **massimo locale** della funzione duale $\psi(\mu)$

Quindi ... per determinare μ^* è (teoricamente) possibile minimizzare la funzione $-\psi(\mu)$

Aggiornamento del moltiplicatore – 3

Vediamo come è fatto il gradiente della funzione duale

$$
\begin{aligned}
\nabla\psi(\mu) &= \nabla_\mu x(\mu)\nabla_x L_a(x(\mu),\mu;\epsilon) + \nabla_\mu L_a(x(\mu),\mu;\epsilon) \\
& \quad (\text{N.B. } \nabla_x L_a(x(\mu),\mu;\epsilon) = 0) \\
&= \nabla_\mu L_a(x(\mu),\mu;\epsilon) = h(x(\mu))
\end{aligned}
$$

Aggiornamento del moltiplicatore – 3

Vediamo come è fatto il gradiente della funzione duale

$$
\begin{aligned}
\nabla\psi(\mu) &= \nabla_\mu x(\mu)\nabla_x L_a(x(\mu),\mu;\epsilon) + \nabla_\mu L_a(x(\mu),\mu;\epsilon) \\
&\quad (\text{N.B. } \nabla_x L_a(x(\mu),\mu;\epsilon) = 0) \\
&= \nabla_\mu L_a(x(\mu),\mu;\epsilon) = h(x(\mu))
\end{aligned}
$$

Una iterazione del metodo del gradiente è quindi:

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

N.B. Se (x_k, μ_k) è una coppia di KKT ($h(x_k) = 0$), allora $\mu_{k+1} = \mu_k$

Aggiornamento del moltiplicatore – 3

Vediamo come è fatto il gradiente della funzione duale

$$\begin{aligned} \nabla\psi(\mu) &= \nabla_\mu x(\mu)\nabla_x L_a(x(\mu),\mu;\epsilon) + \nabla_\mu L_a(x(\mu),\mu;\epsilon) \\ &\quad (\text{N.B. } \nabla_x L_a(x(\mu),\mu;\epsilon) = 0) \\ &= \nabla_\mu L_a(x(\mu),\mu;\epsilon) = h(x(\mu)) \end{aligned}$$

Una iterazione del metodo del gradiente è quindi:

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

N.B. Se (x_k, μ_k) è una coppia di KKT ($h(x_k) = 0$), allora $\mu_{k+1} = \mu_k$

Aggiornamento del moltiplicatore – 3

Vediamo come è fatto il gradiente della funzione duale

$$\begin{aligned}\nabla\psi(\mu) &= \nabla_\mu x(\mu)\nabla_x L_a(x(\mu),\mu;\epsilon) + \nabla_\mu L_a(x(\mu),\mu;\epsilon)\\ &\quad \text{(N.B. } \nabla_x L_a(x(\mu),\mu;\epsilon) = 0\text{)}\\ &= \nabla_\mu L_a(x(\mu),\mu;\epsilon) = h(x(\mu))\end{aligned}$$

Una iterazione del metodo del gradiente è quindi:

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

N.B. Se (x_k, μ_k) è una coppia di KKT ($h(x_k) = 0$), allora $\mu_{k+1} = \mu_k$

Aggiornamento del moltiplicatore – 3

Vediamo come è fatto il gradiente della funzione duale

$$
\begin{aligned}
\nabla\psi(\mu) &= \nabla_\mu x(\mu)\nabla_x L_a(x(\mu),\mu;\epsilon) + \nabla_\mu L_a(x(\mu),\mu;\epsilon) \\
&\quad (\text{N.B. } \nabla_x L_a(x(\mu),\mu;\epsilon) = 0) \\
&= \nabla_\mu L_a(x(\mu),\mu;\epsilon) = h(x(\mu))
\end{aligned}
$$

Una iterazione del metodo del gradiente è quindi:

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

N.B. Se (x_k, μ_k) è una coppia di KKT ($h(x_k) = 0$),
allora $\mu_{k+1} = \mu_k$

Aggiornamento del moltiplicatore – 3

Vediamo come è fatto il gradiente della funzione duale

$$\begin{aligned}\nabla\psi(\mu) &= \nabla_\mu x(\mu)\nabla_x L_a(x(\mu),\mu;\epsilon) + \nabla_\mu L_a(x(\mu),\mu;\epsilon)\\ &\quad (\text{N.B. } \nabla_x L_a(x(\mu),\mu;\epsilon) = 0)\\ &= \nabla_\mu L_a(x(\mu),\mu;\epsilon) = h(x(\mu))\end{aligned}$$

Una iterazione del metodo del gradiente è quindi:

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

N.B. Se (x_k, μ_k) è una coppia di KKT ($h(x_k) = 0$), allora $\mu_{k+1} = \mu_k$