

Ottimizzazione Non Lineare

G. Liuzzi¹

Giovedì 20 Dicembre 2018

¹Istituto di Analisi dei Sistemi ed Informatica IASI - CNR

Algoritmo SEQPEN

Algoritmo SEQPEN

INPUT: maxit , $\tau > 0$, $\epsilon_0 > 0$

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

Trova una soluzione globale (x^k) del problema

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

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

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

endfor

OUTPUT: una sol. approssimata $x^* = x^k$

Algoritmo SEQPEN

L'algoritmo che abbiamo visto:

- ad ogni passo richiede di risolvere all'ottimo **globale**

$$\min_{x \in \mathbb{R}^n} P_\epsilon(x);$$

- è (o può essere) **molto costoso**;
- non sempre è possibile;
- gli algoritmi di min. non vincolata determinano \hat{x} t.c.

$$\|\nabla P_\epsilon(\hat{x})\| \leq \text{tol.}$$

Algoritmo SEQPEN modificato (1)

Algoritmo SEQPEN_{mod}

INPUT: maxit, $\rho > 0$, $k = 0$, $\epsilon_0 > 0$

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

Calcola x^k t.c. $\|\nabla P_{\epsilon_k}(x^k)\| \leq \rho$

if $q(x^k) \leq \rho$ **then** STOP

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

endfor

OUTPUT: una sol. approssimata $x^* = x^k$

Algoritmo SEQPEN modificato (1)

Algoritmo SEQPEN_{mod}

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

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

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

if $\tau_k < \rho$ **and** $q(x^k) \leq \rho$ **then** STOP

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

endfor

OUTPUT: una sol. approssimata $x^* = x^k$

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.

$$\left\{ \frac{2}{\epsilon_k} \max\{g(x_k), 0\} \right\}_K \rightarrow \lambda^*$$
$$\left\{ \frac{2}{\epsilon_k} h(x_k) \right\}_K \rightarrow \mu^*$$

Un po' di esempi di risoluzione

esempio3:

$$\begin{aligned} \min \quad & -x_1 - x_2 \\ \text{s.t.} \quad & x_1^2 + x_2^2 = 1 \end{aligned}$$

$$x^0 = (-1, -1)^\top$$

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

La funzione di penalità esterna è:

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

Un po' di esempi di risoluzione

esempio3:

$$\begin{aligned} \min \quad & -x_1 - x_2 \\ \text{s.t.} \quad & x_1^2 + x_2^2 = 1 \end{aligned}$$

$$x^0 = (-1, -1)^\top$$

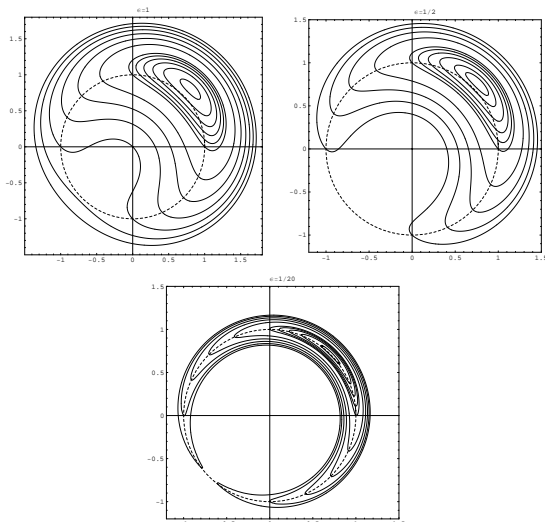
Ammette $(x^*, y^*) = (1, 1)/\sqrt{2}$ come **unico** punto di minimo globale

La funzione di penalità esterna è:

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

Un po' di esempi di risoluzione

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



Un po' di esempi di risoluzione

esempio4:

$$\begin{aligned} \min \quad & -x_1 - x_2 \\ \text{s.t.} \quad & x_1^2 + x_2^2 \leq 1 \end{aligned}$$

$$x^0 = (-1, -1)^\top$$

Un po' di esempi di risoluzione

maratos:

$$\begin{aligned} \min \quad & -x_1 + 2(x_1^2 + x_2^2 - 1) \\ \text{s.t.} \quad & x_1^2 + x_2^2 = 1 \end{aligned}$$

$$x^0 = (-1, -1)^\top$$

Un po' di esempi di risoluzione

hs14:

$$\begin{aligned} \min & (x_1 - 2)^2 + (x_2 - 1)^2 \\ \text{s.t.} & x_1^2/4 + x_2^2 \leq 1 \\ & x_1 - 2x_2 = -1 \end{aligned}$$

$$x^0 = (2, 2)^\top$$

Un po' di esempi di risoluzione

hs24:

$$\begin{aligned} \min \quad & x_2^3((x_1 - 3)^2 - 9)/(27\sqrt{3}) \\ \text{s.t.} \quad & x_1/\sqrt{3} - x_2 \geq 0 \\ & x_1 + \sqrt{3}x_2 \geq 0 \\ & -x_1 - \sqrt{3}x_2 + 6 \geq 0 \\ & x_1 \geq 0, \quad x_2 \geq 0 \end{aligned}$$

$$x^0 = (1, 1/2)^\top$$

Un po' di esempi di risoluzione

hs32:

$$\begin{aligned} \min & (x_1 + 3 * x_2 + x_3)^2 + 4(x_1 - x_2)^2 \\ \text{s.t.} & x_1 + x_2 + x_3 = 1 \\ & 6x_2 + 4x_3 - x_1^3 \geq 3 \\ & x_1 \geq 0, x_2 \geq 0, x_3 \geq 0 \end{aligned}$$

$$x^0 = (0.1, 0.7, 0.2)^T$$

Un po' di esempi di risoluzione

hs41:

$$\begin{aligned} \min \quad & 2 - x_1 x_2 x_3 \\ \text{s.t.} \quad & x_1 + 2x_2 + 2x_3 - x_4 = 0 \\ & 0 \leq x_i \leq 1, \quad i = 1, 2, 3, 4 \end{aligned}$$

$$x^0 = (2, 2, 2, 2)^\top$$

Un po' di esempi di risoluzione

hs55:

$$\begin{aligned} \min \quad & x_1 + 2x_2 + 4x_5 + e^{x_1x_4} \\ \text{s.t.} \quad & x_1 + 2x_2 + 5x_5 - 6 = 0 \\ & x_1 + x_2 + x_3 - 3 = 0 \\ & x_4 + x_5 + x_6 - 2 = 0 \\ & x_1 + x_4 - 1 = 0 \\ & x_2 + x_5 - 2 = 0 \\ & x_3 + x_6 - 2 = 0 \\ & x_i \geq 0, \quad i = 1, \dots, 6 \\ & x_1 \leq 1, \quad x_4 \leq 1 \end{aligned}$$

$$x^0 = (1, 2, 0, 0, 0, 2)^\top$$

Un po' di esempi di risoluzione

hs60:

$$\begin{aligned} \min & (x_1 - 1)^2 + (x_1 - x_2)^2 + (x_2 - x_3)^4 \\ \text{s.t.} & x_1(1 + x_2^2) + x_3^4 - 4 - 3\sqrt{2} = 0 \\ & -10 \leq x_i \leq 10, \quad i = 1, 2, 3 \end{aligned}$$

$$x^0 = (2, 2, 2)^\top$$

Tanto per cominciare

Consideriamo il problema con soli vincoli di uguaglianza

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

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

con gradiente (vi ricordo):

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

Perciò, 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

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

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

con gradiente (vi ricordo):

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

Perciò, 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

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

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

con gradiente (vi ricordo):

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

Perciò, 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

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

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

con gradiente (vi ricordo):

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

Perciò, 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:

- la funzione $L(x, \mu)$ è lineare rispetto a μ e non è quindi limitata inferiormente su \mathbb{R}^{n+p} ;
- anche ammettendo di conoscere i moltiplicatori $\bar{\mu}$ associati ad un punto di KKT, una minimizzazione non vincolata di $L(x, \bar{\mu})$ produce un punto $\hat{x} \notin \mathcal{F}$!

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:

- la funzione $L(x, \mu)$ è lineare rispetto a μ e non è quindi limitata inferiormente su \mathbb{R}^{n+p} ;
- anche ammettendo di conoscere i moltiplicatori $\bar{\mu}$ associati ad un punto di KKT, una minimizzazione non vincolata di $L(x, \bar{\mu})$ produce un punto $\hat{x} \notin \mathcal{F}$!

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:

- la funzione $L(x, \mu)$ è lineare rispetto a μ e non è quindi limitata inferiormente su \mathbb{R}^{n+p} ;
- anche ammettendo di conoscere i moltiplicatori $\bar{\mu}$ associati ad un punto di KKT, una minimizzazione non vincolata di $L(x, \bar{\mu})$ produce un punto $\hat{x} \notin \mathcal{F}$!

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

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

$$\nabla_\mu L_a(x, \mu; \epsilon) = \nabla_\mu L(x, \mu) = h(x)$$

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

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

$$\nabla_\mu L_a(x, \mu; \epsilon) = \nabla_\mu L(x, \mu) = h(x)$$

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 $\nabla L_a(\bar{x}, \bar{\mu}) = 0$. □

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 $\nabla L_a(\bar{x}, \bar{\mu}) = 0$. □

Esempio

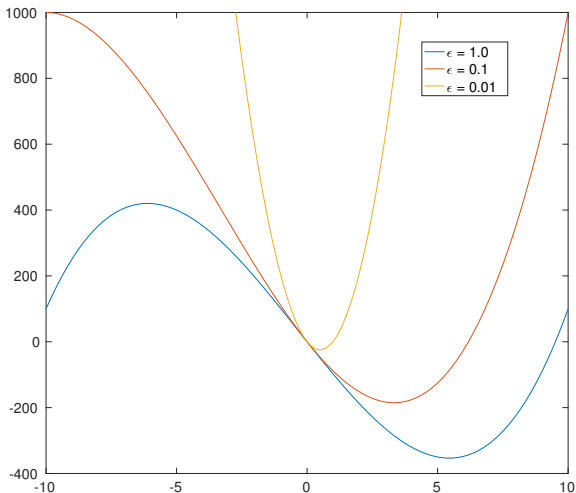
Consideriamo il problema

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

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

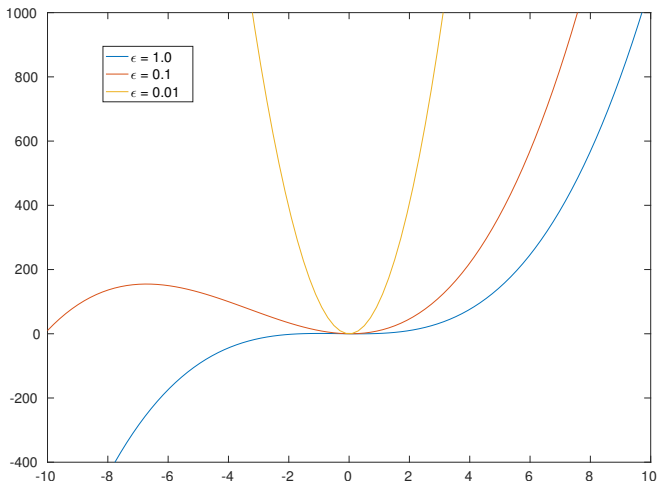
Esempio

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

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

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

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

Esempio

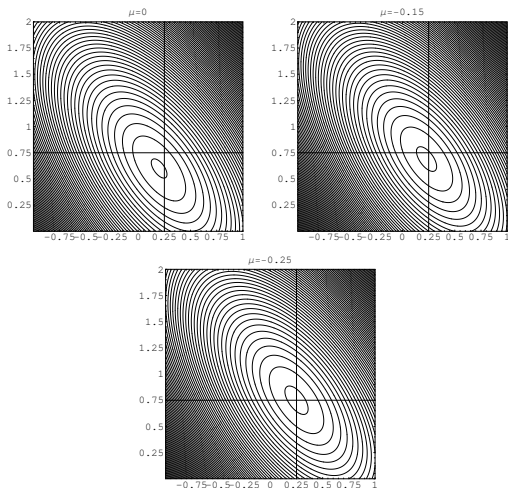
Consideriamo il problema

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

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

Esempio

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



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
- Sia $(\bar{x}, \bar{\mu})$ una sol. del problema originale. Per **valori sufficientemente piccoli di** $\epsilon > 0$, \bar{x} è un minimo locale di $L_a(x, \bar{\mu}; \epsilon)$

Metodo di soluzione

Algoritmo SEQLAGR**Dati:** $\epsilon_0 > 0$, $\beta > 1$, $\{\tau_k\} \rightarrow 0$, $\rho > 0$, μ_0 , maxit**for** $k = 0, 1, \dots, \text{maxit}$ Calcola x_k t.c. $\|\nabla_x L_a(x_k, \mu_k; \epsilon_k)\| \leq \tau_k$ **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:** $\epsilon_0 > 0$, $\beta > 1$, $\{\tau_k\} \rightarrow 0$, $\rho > 0$, μ_0 , maxit**for** $k = 0, 1, \dots, \text{maxit}$ Calcola x_k t.c. $\|\nabla_x L_a(x_k, \mu_k; \epsilon_k)\| \leq \tau_k$ **if** $\|\nabla L_a(x_k, \mu_k; \epsilon_k)\| < \rho$ **then** $x^* \leftarrow x_k$, $\mu^* \leftarrow \mu_k$, STOP**endif** $\epsilon_{k+1} = \epsilon_k / \beta$ Calcola μ_{k+1} **endfor****Return:** miglior coppia trovata (x^*, μ^*)

Metodo di soluzione

Algoritmo SEQLAGR**Dati:** $\epsilon_0 > 0$, $\beta > 1$, $\{\tau_k\} \rightarrow 0$, $\rho > 0$, μ_0 , maxit**for** $k = 0, 1, \dots, \text{maxit}$ Calcola x_k t.c. $\|\nabla_x L_a(x_k, \mu_k; \epsilon_k)\| \leq \tau_k$ **if** $\|\nabla L_a(x_k, \mu_k; \epsilon_k)\| < \rho$ **then** $x^* \leftarrow x_k$, $\mu^* \leftarrow \mu_k$, STOP**endif** $\epsilon_{k+1} = \epsilon_k / \beta$ Calcola μ_{k+1} **endfor****Return:** miglior coppia trovata (x^*, μ^*)

Metodo di soluzione

Algoritmo SEQLAGR

Dati: $\epsilon_0 > 0$, $\beta > 1$, $\{\tau_k\} \rightarrow 0$, $\rho > 0$, μ_0 , maxit

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

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

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

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

endif

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

Calcola μ_{k+1}

endfor

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

Aggiornamento del moltiplicatore

Come si passa da μ_k a μ_{k+1} ?

Supponiamo di aver calcolato x_k e supponiamo anche che x_k sia stazionario per la funzione $L_a(x, \mu_k; \epsilon_k)$ cioè

$$\nabla_x L_a(x_k, \mu_k; \epsilon_k) = 0$$

Ovvero

$$\nabla_x L_a(x_k, \mu_k; \epsilon_k) = \nabla f(x_k) + \nabla h(x_k) \left(\mu_k + \frac{2}{\epsilon_k} h(x_k) \right) = 0$$

Se definiamo $\mu_{k+1} = \mu_k + \frac{2}{\epsilon_k} h(x_k)$ risulterà

$$\nabla f(x_k) + \nabla h(x_k) \mu_{k+1} = 0$$

per cui x_k soddisfa la condizione di stazionarietà a meno della condizione $h(x_k) = 0$

Aggiornamento del moltiplicatore

Come si passa da μ_k a μ_{k+1} ?

Supponiamo di aver calcolato x_k e supponiamo anche che x_k sia stazionario per la funzione $L_a(x, \mu_k; \epsilon_k)$ cioè

$$\nabla_x L_a(x_k, \mu_k; \epsilon_k) = 0$$

Ovvero

$$\nabla_x L_a(x_k, \mu_k; \epsilon_k) = \nabla f(x_k) + \nabla h(x_k) \left(\mu_k + \frac{2}{\epsilon_k} h(x_k) \right) = 0$$

Se definiamo $\mu_{k+1} = \mu_k + \frac{2}{\epsilon_k} h(x_k)$ risulterà

$$\nabla f(x_k) + \nabla h(x_k) \mu_{k+1} = 0$$

per cui x_k soddisfa la condizione di stazionarietà a meno della condizione $h(x_k) = 0$

Aggiornamento del moltiplicatore

Come si passa da μ_k a μ_{k+1} ?

Supponiamo di aver calcolato x_k e supponiamo anche che x_k sia stazionario per la funzione $L_a(x, \mu_k; \epsilon_k)$ cioè

$$\nabla_x L_a(x_k, \mu_k; \epsilon_k) = 0$$

Ovvero

$$\nabla_x L_a(x_k, \mu_k; \epsilon_k) = \nabla f(x_k) + \nabla h(x_k) \left(\mu_k + \frac{2}{\epsilon_k} h(x_k) \right) = 0$$

Se definiamo $\mu_{k+1} = \mu_k + \frac{2}{\epsilon_k} h(x_k)$ risulterà

$$\nabla f(x_k) + \nabla h(x_k) \mu_{k+1} = 0$$

per cui x_k soddisfa la condizione di stazionarietà a meno della condizione $h(x_k) = 0$

Metodo di soluzione - Metodo dei Moltiplicatori

Algoritmo SEQLAGR**Dati:** $\epsilon_0 > 0$, $\beta > 1$, $\{\tau_k\} \rightarrow 0$, $\rho > 0$, μ_0 , maxit**for** $k = 0, 1, \dots, \text{maxit}$ Calcola x_k t.c. $\|\nabla_x L_a(x_k, \mu_k; \epsilon_k)\| \leq \tau_k$ **if** $\|\nabla L_a(x_k, \mu_k; \epsilon_k)\| < \rho$ **then** $x^* \leftarrow x_k$, $\mu^* \leftarrow \mu_k$, STOP**endif** $\epsilon_{k+1} = \epsilon_k / \beta$

$$\mu_{k+1} = \mu_k + \frac{2h(x_k)}{\epsilon_k}$$

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