Ottimizzazione dei Sistemi Complessi

G. Liuzzi[1]

Mercoledì 15 Aprile 2020

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

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 $L(x, \mu)$!!

Idea: minimizzare $L(x, \mu)$ sullo spazio $\mathbb{R}^{n+p}$ per determinare $(\bar{x}, \bar{\mu})$ tale che $\nabla L(\bar{x}, \bar{\mu}) = \mathbf{0}$

Purtroppo: $L(x, \mu)$ è lineare rispetto alle variabili duali μ!!

Quindi ...

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

Idea: minimizzare $L(x, \mu)$ sullo spazio $\mathbb{R}^{n+p}$ per determinare $(\bar{x}, \bar{\mu})$ tale che $\nabla L(\bar{x}, \bar{\mu}) = \mathbf{0}$

Purtroppo: $L(x, \mu)$ è lineare rispetto alle variabili duali μ!!

Quindi ...

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

Idea: minimizzare $L(x, \mu)$ sullo spazio $\mathbb{R}^{n+p}$ per determinare $(\bar{x}, \bar{\mu})$ tale che $\nabla L(\bar{x}, \bar{\mu}) = \mathbf{0}$

Purtroppo: $L(x, \mu)$ è lineare rispetto alle variabili duali μ!!

Però ...

Siano: x^* un minimo locale (regolare) del problema e μ^* i moltiplicatori di KKT associati.

Supponiamo di conoscere i moltiplicatori μ^* ottimi e di voler determinare x^*.
Cosa succede se tentiamo di minimizzare la funzione $L(x, \mu^*)$ su $\mathbb{R}^n$?

Purtroppo: spesso si ottiene un punto $\bar{x} \neq x^*$ e tale che $h(\bar{x}) \neq \mathbf{0}$!!

Però ...

Siano: x^* un minimo locale (regolare) del problema e μ^* i moltiplicatori di KKT associati.

Supponiamo di conoscere i moltiplicatori μ^* ottimi e di voler determinare x^*.
Cosa succede se tentiamo di minimizzare la funzione $L(x, \mu^*)$ su $\mathbb{R}^n$?

Purtroppo: spesso si ottiene un punto $\bar{x} \neq x^*$ e tale che $h(\bar{x}) \neq \mathbf{0}$!!

Esempio

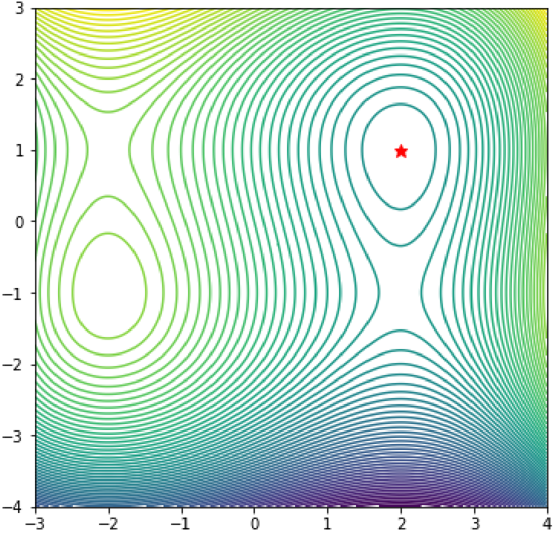
Consideriamo per un momento il problema

$$\begin{aligned} \min\ & x_1^3 + x_2^3 \\ s.t.\ & x_1 - 2 = 0 \\ & x_2 - 1 = 0 \end{aligned}$$

Il punto di minimo globale del problema è (ovviamente) il punto $x^* = (2,\ 1)^\top$ con moltiplicatori $\mu_1^* = -12$, $\mu_2^* = -3$

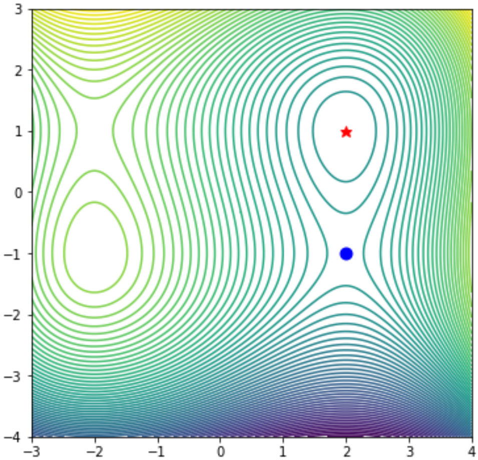
$$L(x, \mu^*) = x_1^3 + x_2^3 - 12(x_1 - 2) - 3(x_2 - 1)$$

Esempio



Esempio

Una minimizzazione locale della funzione $L(x, \mu^*)$ a partire dal punto $x_\circ = (0,\ -1)^\top$ produce il punto $\bar{x} = (2,\ -1)^\top \notin \mathcal{F}$!!



Allora?

Bisogna modificare $L(x, \mu^*)$ "aggiungendo" un termine che consenta di penalizzare la violazione dei vincoli ...

$$\begin{aligned} L_a(x, \mu; \epsilon) &= L(x, \mu) + \frac{1}{\epsilon}\|h(x)\|^2 \\ &= f(x) + \mu^\top h(x) + \frac{1}{\epsilon}\|h(x)\|^2 \end{aligned}$$

Esempio:

$$L_a(x, \mu^*, \epsilon) = x_1^3 + x_2^3 - 12(x_1 - 2) - 3(x_2 - 1) + \frac{1}{\epsilon}(x_1 - 2)^2 + \frac{1}{\epsilon}(x_2 - 1)^2$$

Una minimizzazione di $L_a(x, \mu^*, \epsilon)$ con $\epsilon = 0.5$ a partire da $x_\circ = (0,\ -1)^\top$ produce la soluzione globale x^*

Allora?

Bisogna modificare $L(x, \mu^*)$ "aggiungendo" un termine che consenta di penalizzare la violazione dei vincoli ...

$$\begin{aligned} L_a(x, \mu; \epsilon) &= L(x, \mu) + \frac{1}{\epsilon}\|h(x)\|^2 \\ &= f(x) + \mu^\top h(x) + \frac{1}{\epsilon}\|h(x)\|^2 \end{aligned}$$

Esempio:

$$L_a(x, \mu^*, \epsilon) = x_1^3 + x_2^3 - 12(x_1 - 2) - 3(x_2 - 1) + \frac{1}{\epsilon}(x_1 - 2)^2 + \frac{1}{\epsilon}(x_2 - 1)^2$$

Una minimizzazione di $L_a(x, \mu^*, \epsilon)$ con $\epsilon = 0.5$ a partire da $x_\circ = (0,\ -1)^\top$ produce la soluzione globale x^*

Allora?

Bisogna modificare $L(x,\mu^*)$ "aggiungendo" un termine che consenta di penalizzare la violazione dei vincoli ...

$$\begin{aligned} L_a(x,\mu;\epsilon) &= L(x,\mu) + \frac{1}{\epsilon}\|h(x)\|^2 \\ &= f(x) + \mu^\top h(x) + \frac{1}{\epsilon}\|h(x)\|^2 \end{aligned}$$

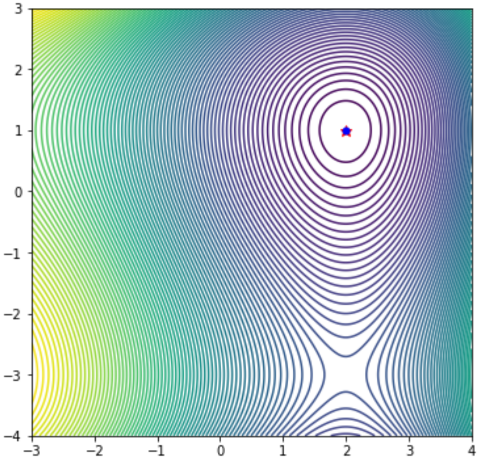
Esempio:

$$L_a(x,\mu^*,\epsilon) = x_1^3 + x_2^3 - 12(x_1-2) - 3(x_2-1) + \frac{1}{\epsilon}(x_1-2)^2 + \frac{1}{\epsilon}(x_2-1)^2$$

Una minimizzazione di $L_a(x,\mu^*,\epsilon)$ con $\epsilon = 0.5$ a partire da $x_\circ = (0,\ -1)^\top$ produce la soluzione globale x^*

Esempio

Una minimizzazione di $L_a(x, \mu^*, \epsilon)$ con $\epsilon = 0.5$ a partire da $x_\circ = (0,\ -1)^\top$ produce la soluzione globale x^*



Lagrangiano aumentato

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

con gradiente

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

Lagrangiano aumentato

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

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

Seconda proprietà

Indichiamo con $\mathcal{G}(\mathcal{C})$ l'insieme dei punti di minimo globale del problema

$$\begin{array}{l} \min \ f(x) \\ s.t. \ x \in \mathcal{F} \cap \mathcal{C} \end{array}$$

Supponiamo:

- che esista $\mathcal{C} \subseteq \mathbb{R}^n$ compatto tale che $\mathcal{G}(\mathcal{C}) \subseteq \mathring{\mathcal{C}}$
- $\bar{x} \in \mathcal{G}(\mathcal{C})$ e $\bar{\mu}$ moltiplicatori di KKT associati

Proposizione

Allora **per valori sufficiente piccoli** *di ϵ, $\bar{x}$ è un minimo globale non vincolato di $L_a(x, \bar{\mu}; \epsilon)$ su $\mathcal{C}$*

Dim. Per contraddizione, supponiamo che per ogni $k \in \mathbb{N}$ esiste un $\epsilon_k \leq 1/k$ ed un punto x_k di minimo globale di $L_a(x, \bar{\mu}; \epsilon_k)$ su $\mathcal{C}$ tale che

$$L_a(x_k, \bar{\mu}; \epsilon_k) < L_a(\bar{x}, \bar{\mu}; \epsilon_k) = f(\bar{x})$$

Seconda proprietà

Indichiamo con $\mathcal{G}(\mathcal{C})$ l'insieme dei punti di minimo globale del problema

$$\begin{array}{l} \min \ f(x) \\ s.t. \ x \in \mathcal{F} \cap \mathcal{C} \end{array}$$

Supponiamo:

- che esista $\mathcal{C} \subseteq \mathbb{R}^n$ compatto tale che $\mathcal{G}(\mathcal{C}) \subseteq \mathring{\mathcal{C}}$
- $\bar{x} \in \mathcal{G}(\mathcal{C})$ e $\bar{\mu}$ moltiplicatori di KKT associati

Proposizione

Allora **per valori sufficiente piccoli** *di ϵ, $\bar{x}$ è un minimo globale non vincolato di $L_a(x, \bar{\mu}; \epsilon)$ su $\mathcal{C}$*

Dim. Per contraddizione, supponiamo che per ogni $k \in \mathbb{N}$ esiste un $\epsilon_k \leq 1/k$ ed un punto x_k di minimo globale di $L_a(x, \bar{\mu}; \epsilon_k)$ su $\mathcal{C}$ tale che

$$L_a(x_k, \bar{\mu}; \epsilon_k) < L_a(\bar{x}, \bar{\mu}; \epsilon_k) = f(\bar{x})$$

Seconda proprietà – segue

Quindi

$$f(x_k) + \bar{\mu}^\top h(x_k) + \frac{1}{\epsilon_k}\|h(x_k)\| < f(\bar{x}).$$

Ora $\{x_k\}$ è tutta contenuta in un compatto, quindi ammette punti di accumulazione. Esiste quindi certamente una sottosuccessione convergente, che rinominiamo $\{x_k\}$, tale che

$$\lim_{k\to\infty} x_k = \hat{x}.$$

Deve pertanto risultare

$$\lim_{k\to\infty} L_a(x_k, \bar{\mu}; \epsilon_k) \leq f(\bar{x})$$

ovvero $h(\hat{x}) = 0$ e $f(\hat{x}) \leq f(\bar{x})$, ovvero anche il punto $\hat{x}$ è minimo globale del problema su $\mathcal{C}$. Quindi $\hat{x} \in \overset{\circ}{\mathcal{C}}$. Ma allora, per k suff. elevato $x_k \in \overset{\circ}{\mathcal{C}}$. E allora $\nabla_x L_a(x_k, \bar{\mu}; \epsilon) = 0$. Di conseguenza, $(x_k, \bar{\mu})$ è una coppia di KKT per il problema, tale che percui

$$L_a(x_k, \bar{\mu}; \epsilon_k) = f(x_k) < f(\bar{x})$$

che contraddice il fatto che $\bar{x}$ è minimo globale del problema. $\square$

Terza 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 $\bar{\mu}$!

Terza 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 $\bar{\mu}$!

Terza 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

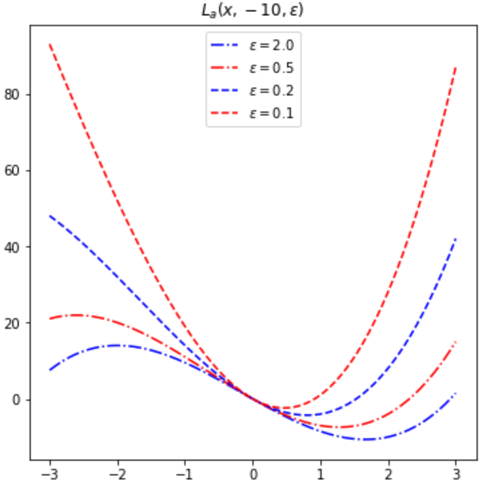
Consideriamo il problema

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

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

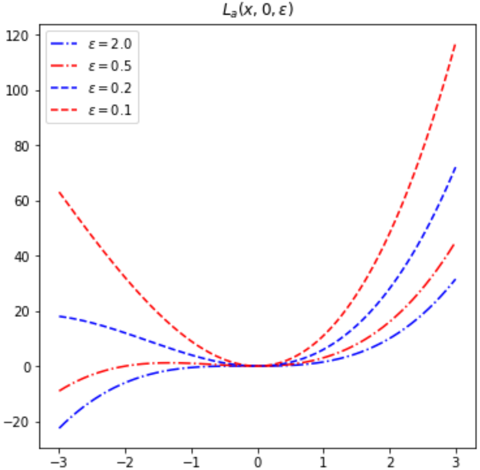
Esempio

Andamenti della funzione $L_a(x, -10; \epsilon)$ per $\epsilon = 2, 0.5, 0.2, 0.1$



Esempio

Andamenti della funzione $L_a(x, 0; \epsilon)$ per $\epsilon = 2, 0.5, 0.2, 0.1$



Esempio

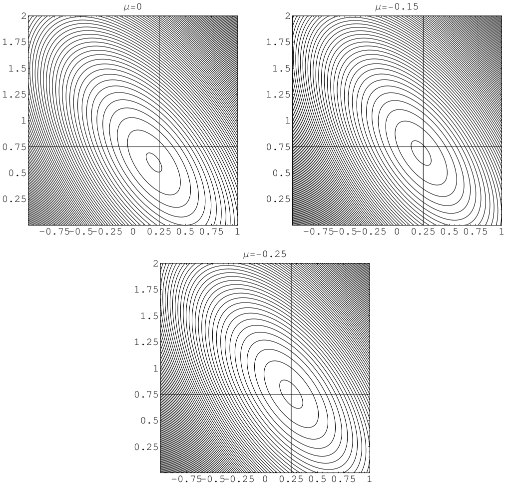
Consideriamo il problema

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

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

Esempio

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



Metodo di soluzione

Algoritmo `SEQLAGR`

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

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

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

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

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

endif

endfor

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

Metodo di soluzione

Algoritmo `SEQLAGR`

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

for $k = 0, 1, \ldots,$`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: $\epsilon_\circ > 0$, $\beta > 1$, $\{\tau_k\} \to 0$, $\rho > 0$, $\mu_\circ$, `maxit`

for $k = 0, 1, \ldots,$`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: $\epsilon_\circ > 0$, $\beta > 1$, $\{\tau_k\} \to 0$, $\rho > 0$, $\mu_\circ$, $\texttt{maxit}$

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

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

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

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

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

endfor

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

Proprietà di convergenza

Teorema

Assumiamo che:

- $\{\epsilon_k\} \to 0$, $\{\tau_k\} \to 0$
- *per ogni* k, x_k *t.c* $\|\nabla_x L_a(x_k, \mu_k; \epsilon_k)\| \leq \tau_k$
- *per ogni* k, $x_k \in D$ *compatto*, $\boxed{\mu_k \in A \text{ compatto}}$
- *per ogni* $x \in D$, *i gradienti* $\nabla h_j(x)$, $\forall\, j$, *sono lin. indipendenti.*

Allora, se x^* *è un punto limite di* $\{x_k\}$, *i.e.* $\exists K$ *s.t.*

$$\lim_{k \to \infty, k \in K} x_k = x^*$$

x^* *è un punto di KKT con molt.* μ^* *t.c.*

$$\left\{\mu_k + \frac{2}{\epsilon_k} h(x_k)\right\}_K \to \mu^*$$

Proprietà di convergenza

Teorema

Assumiamo che:

- $\{\epsilon_k\} \to 0$, $\{\tau_k\} \to 0$
- *per ogni* k, x_k *t.c* $\|\nabla_x L_a(x_k, \mu_k; \epsilon_k)\| \le \tau_k$
- *per ogni* k, $x_k \in D$ *compatto,* $\boxed{\mu_k \in A \text{ compatto}}$
- *per ogni* $x \in D$, *i gradienti* $\nabla h_j(x)$, $\forall\, j$, *sono lin. indipendenti.*

Allora, se x^* *è un punto limite di* $\{x_k\}$, *i.e.* $\exists K$ *s.t.*

$$\lim_{k\to\infty, k\in K} x_k = x^*$$

x^* *è un punto di KKT con molt.* μ^* *t.c.*

$$\left\{\mu_k + \frac{2}{\epsilon_k} h(x_k)\right\}_K \to \mu^*$$

Proprietà di convergenza

Teorema

Assumiamo che:

- $\{\epsilon_k\} \to 0$, $\{\tau_k\} \to 0$
- *per ogni* k, x_k *t.c* $\|\nabla_x L_a(x_k, \mu_k; \epsilon_k)\| \leq \tau_k$
- *per ogni* k, $x_k \in D$ *compatto*, $\boxed{\mu_k \in A \text{ compatto}}$
- *per ogni* $x \in D$, *i gradienti* $\nabla h_j(x)$, $\forall\, j$, *sono lin. indipendenti.*

Allora, se x^* *è un punto limite di* $\{x_k\}$, *i.e.* $\exists K$ *s.t.*

$$\lim_{k \to \infty, k \in K} x_k = x^*$$

x^* *è un punto di KKT con molt.* μ^* *t.c.*

$$\left\{\mu_k + \frac{2}{\epsilon_k} h(x_k)\right\}_K \to \mu^*$$

Proprietà di convergenza

Teorema

Assumiamo che:

- $\{\epsilon_k\} \to 0$, $\{\tau_k\} \to 0$
- *per ogni* k, x_k *t.c* $\|\nabla_x L_a(x_k, \mu_k; \epsilon_k)\| \leq \tau_k$
- *per ogni* k, $x_k \in D$ *compatto*, $\boxed{\mu_k \in A \text{ compatto}}$
- *per ogni* $x \in D$, *i gradienti* $\nabla h_j(x)$, $\forall\, j$, *sono lin. indipendenti.*

Allora, se x^* *è un punto limite di* $\{x_k\}$, *i.e.* $\exists K$ *s.t.*

$$\lim_{k\to\infty, k\in K} x_k = x^*$$

x^* *è un punto di KKT con molt.* μ^* *t.c.*

$$\left\{\mu_k + \frac{2}{\epsilon_k} h(x_k)\right\}_K \to \mu^*$$

Proprietà di convergenza

Teorema

Assumiamo che:

- $\{\epsilon_k\} \to 0$, $\{\tau_k\} \to 0$
- *per ogni* k, x_k *t.c* $\|\nabla_x L_a(x_k, \mu_k; \epsilon_k)\| \leq \tau_k$
- *per ogni* k, $x_k \in D$ *compatto*, $\boxed{\mu_k \in A \text{ compatto}}$
- *per ogni* $x \in D$, *i gradienti* $\nabla h_j(x)$, $\forall\, j$, *sono lin. indipendenti.*

Allora, se x^* *è un punto limite di* $\{x_k\}$, *i.e.* $\exists K$ *s.t.*

$$\lim_{k\to\infty, k\in K} x_k = x^*$$

x^* *è un punto di KKT con molt.* μ^* *t.c.*

$$\left\{\mu_k + \frac{2}{\epsilon_k} h(x_k)\right\}_K \to \mu^*$$

Proprietà di convergenza

Teorema

Assumiamo che:

- $\{\epsilon_k\} \to 0$, $\{\tau_k\} \to 0$
- *per ogni* k, x_k *t.c* $\|\nabla_x L_a(x_k, \mu_k; \epsilon_k)\| \leq \tau_k$
- *per ogni* k, $x_k \in D$ *compatto,* $\boxed{\mu_k \in A \text{ compatto}}$
- *per ogni* $x \in D$, *i gradienti* $\nabla h_j(x)$, $\forall\, j$, *sono lin. indipendenti.*

Allora, se x^* *è un punto limite di* $\{x_k\}$, *i.e.* $\exists K$ *s.t.*

$$\lim_{k\to\infty, k\in K} x_k = x^*$$

x^* *è un punto di KKT con molt.* μ^* *t.c.*

$$\left\{\mu_k + \frac{2}{\epsilon_k} h(x_k)\right\}_K \to \mu^*$$

Dimostrazione

Dim. Rinominiamo $\{x_k\}$ la sottosucc. convergente a x^*.
Dalla espressione del gradiente risp.x di $L_a(x, \mu_k; \epsilon_k)$ possiamo scrivere

$$\nabla_x L_a(x_k, \mu_k; \epsilon_k) = \nabla f(x_k) + \nabla h(x_k)\tilde{\mu}_k$$

dove si è posto

$$\mu_k + \frac{2}{\epsilon_k} h(x_k) = \tilde{\mu}_k$$

Per ipotesi, per k suff. elevato, la matrice $\nabla h(x_k)$ ha colonne lin. indipendenti.
Quindi $\nabla h(x_k)^\top \nabla h(x_k)$ è invertibile

Quindi, possiamo ottenere $\tilde{\mu}_k$

$$\tilde{\mu}_k = (\nabla h(x_k)^\top \nabla h(x_k))^{-1} \nabla h(x_k)^\top (\nabla_x L_a(x_k, \mu_k; \epsilon_k) - \nabla f(x_k))$$

Dimostrazione

Dim. Rinominiamo $\{x_k\}$ la sottosucc. convergente a x^*.
Dalla espressione del gradiente risp.x di $L_a(x, \mu_k; \epsilon_k)$ possiamo scrivere

$$\nabla_x L_a(x_k, \mu_k; \epsilon_k) = \nabla f(x_k) + \nabla h(x_k)\tilde{\mu}_k$$

dove si è posto

$$\mu_k + \frac{2}{\epsilon_k} h(x_k) = \tilde{\mu}_k$$

Per ipotesi, per k suff. elevato, la matrice $\nabla h(x_k)$ ha colonne lin. indipendenti.
Quindi $\nabla h(x_k)^\top \nabla h(x_k)$ è invertibile

Quindi, possiamo ottenere $\tilde{\mu}_k$

$$\tilde{\mu}_k = (\nabla h(x_k)^\top \nabla h(x_k))^{-1} \nabla h(x_k)^\top (\nabla_x L_a(x_k, \mu_k; \epsilon_k) - \nabla f(x_k))$$

Dimostrazione

Dim. Rinominiamo $\{x_k\}$ la sottosucc. convergente a x^*.
Dalla espressione del gradiente risp.x di $L_a(x, \mu_k; \epsilon_k)$ possiamo scrivere

$$\nabla_x L_a(x_k, \mu_k; \epsilon_k) = \nabla f(x_k) + \nabla h(x_k)\tilde{\mu}_k$$

dove si è posto

$$\mu_k + \frac{2}{\epsilon_k} h(x_k) = \tilde{\mu}_k$$

Per ipotesi, per k suff. elevato, la matrice $\nabla h(x_k)$ ha colonne lin. indipendenti.
Quindi $\nabla h(x_k)^\top \nabla h(x_k)$ è invertibile

Quindi, possiamo ottenere $\tilde{\mu}_k$

$$\tilde{\mu}_k = (\nabla h(x_k)^\top \nabla h(x_k))^{-1} \nabla h(x_k)^\top (\nabla_x L_a(x_k, \mu_k; \epsilon_k) - \nabla f(x_k))$$

Dimostrazione – continua

Ora, siccome per ipotesi $\nabla_x L_a(x_k, \mu_k; \epsilon_k) \to \mathbf{0}$, otteniamo

$$\lim_{k\to\infty} \tilde{\mu}_k = \mu^* = -(M(x^*)^\top M(x^*))^{-1} M(x^*)^\top \nabla f(x^*)$$

e quindi da $\nabla_x L_a(x_k, \mu_k; \epsilon_k) = \nabla f(x_k) + \nabla h(x_k)\tilde{\mu}_k$ otteniamo

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

Dimostrazione – continua

Ora, siccome per ipotesi $\nabla_x L_a(x_k, \mu_k; \epsilon_k) \to \mathbf{0}$, otteniamo

$$\lim_{k\to\infty} \tilde{\mu}_k = \mu^* = -(M(x^*)^\top M(x^*))^{-1} M(x^*)^\top \nabla f(x^*)$$

e quindi da $\nabla_x L_a(x_k, \mu_k; \epsilon_k) = \nabla f(x_k) + \nabla h(x_k)\tilde{\mu}_k$ otteniamo

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

Dimostrazione – continua

Dal fatto che $\tilde{\mu}_k \to \mu^*$, ricordando le esperssioni di $\tilde{\mu}_k$

$$\mu_k + \frac{2}{\epsilon_k} h(x_k) = \tilde{\mu}_k$$

e per il fatto che $\epsilon_k \to 0$, otteniamo che $h(x^*) = 0$. $\square$