Ottimizzazione dei Sistemi Complessi

G. Liuzzi[1]

Venerdì 7 Aprile 2017

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

Un (altro) bel passo indietro ... o avanti!

Consideriamo il problema *NON vincolato*

$$\begin{array}{ll} \min & f(x) \\ c.v. & x \in \mathbb{R}^n \end{array}$$

Supponiamo di conoscere una soluzione x^* del problema e siano verificate le **C.S. del II ordine** in x^*:

- $\nabla f(x^*) = 0$
- $\nabla^2 f(x^*)$ definita positiva

N.B. Attenzione! Se x^* è minimo locale stretto, non è detto che $\nabla^2 f(x^*)$ sia definita positiva!

Il metodo di Newton (1)

Vogliamo costruire una approsimazione del problema in un intorno di x^*

Sviluppo in serie di Taylor troncato al secondo ordine di $f(x)$:

$$\begin{aligned} q(x;x^*) &= f(x^*) + \nabla f(x^*)^\top (x - x^*) + \frac{1}{2}(x - x^*)^\top \nabla^2 f(x^*)(x - x^*) \\ f(x) &\simeq q(x;x^*) \\ q(x^*;x^*) &= f(x^*) \\ \bar{q}(d) &= q(x^* + d; x^*) = f(x^*) + \frac{1}{2} d^\top \nabla^2 f(x^*) d \\ \nabla \bar{q}(d) &= \nabla^2 f(x^*) d \\ \nabla^2 \bar{q}(d) &= \nabla^2 f(x^*) \end{aligned}$$

Il metodo di Newton (1)

Vogliamo costruire una approsimazione del problema in un intorno di x^*

Sviluppo in serie di Taylor troncato al secondo ordine di $f(x)$:

$$
\begin{aligned}
q(x; x^*) &= f(x^*) + \nabla f(x^*)^\top (x - x^*) + \frac{1}{2}(x - x^*)^\top \nabla^2 f(x^*)(x - x^*) \\
f(x) &\simeq q(x; x^*) \\
q(x^*; x^*) &= f(x^*) \\
\bar{q}(d) &= q(x^* + d; x^*) = f(x^*) + \frac{1}{2} d^\top \nabla^2 f(x^*) d \\
\nabla \bar{q}(d) &= \nabla^2 f(x^*) d \\
\nabla^2 \bar{q}(d) &= \nabla^2 f(x^*)
\end{aligned}
$$

Il metodo di Newton (1)

Vogliamo costruire una approsimazione del problema in un intorno di x^*

Sviluppo in serie di Taylor troncato al secondo ordine di $f(x)$:

$$
\begin{aligned}
q(x;x^*) &= f(x^*) + \nabla f(x^*)^\top (x-x^*) + \frac{1}{2}(x-x^*)^\top \nabla^2 f(x^*)(x-x^*) \\
f(x) &\simeq q(x;x^*) \\
q(x^*;x^*) &= f(x^*) \\
\bar{q}(d) &= q(x^*+d;x^*) = f(x^*) + \frac{1}{2} d^\top \nabla^2 f(x^*) d \\
\nabla \bar{q}(d) &= \nabla^2 f(x^*) d \\
\nabla^2 \bar{q}(d) &= \nabla^2 f(x^*)
\end{aligned}
$$

Il metodo di Newton (2)

$q(x; x^*)$ e quindi $\bar{q}(d)$ sono una “buona” approssimazione di f in un intorno di x^* quando:

- x^* è minimo locale (oltre che di f) anche per $q(x; x^*)$ e
- $d^* = 0$ è minimo locale di $\bar{q}(d)$.

Infatti, da

$$
\begin{aligned}
\nabla \bar{q}(d) &= \nabla^2 f(x^*) d = 0 \\
\nabla^2 \bar{q}(d) &= \nabla^2 f(x^*) \ d.p.
\end{aligned}
$$

segue che $d^* = 0$ è proprio l'**unico** minimo locale di $\bar{q}(d)$

Il metodo di Newton (2)

$q(x; x^*)$ e quindi $\bar{q}(d)$ sono una "buona" approssimazione di f in un intorno di x^* quando:

- x^* è minimo locale (oltre che di f) anche per $q(x; x^*)$ e
- $d^* = 0$ è minimo locale di $\bar{q}(d)$.

Infatti, da

$$\begin{aligned} \nabla \bar{q}(d) &= \nabla^2 f(x^*) d = 0 \\ \nabla^2 \bar{q}(d) &= \nabla^2 f(x^*) \ d.p. \end{aligned}$$

segue che $d^* = 0$ è proprio l'**unico** minimo locale di $\bar{q}(d)$

Il metodo di Newton (3)

Supponiamo ora di **non** conoscere x^* ma
di disporre di una **stima** x_k di x^*

Idea

- definisco $q(x; x_k)$ e $\bar{q}_k(d)$
- minimizzo $q(x; x_k)$
- aggiorno la stima di x^* definendo x_{k+1}

Il metodo di Newton (3)

Supponiamo ora di **non** conoscere x^* ma

di disporre di una **stima** x_k di x^*

Idea

- definisco $q(x; x_k)$ e $\bar{q}_k(d)$
- minimizzo $q(x; x_k)$
- aggiorno la stima di x^* definendo x_{k+1}

Il metodo di Newton (4)

Naturalmente, risulta:

$$
\begin{aligned}
\bar{q}_k(d) &= f(x_k) + \nabla f(x_k)^\top d + \frac{1}{2} d^\top \nabla^2 f(x_k) d \\
\nabla \bar{q}_k(d) &= \nabla f(x_k) + \nabla^2 f(x_k) d
\end{aligned}
$$

$\nabla \bar{q}_k(d) = 0$ quando $\nabla f(x_k) + \nabla^2 f(x_k) d = 0$.

Se $\nabla^2 f(x_k)$ è non singolare, allora possiamo definire:

$$
\begin{aligned}
d_k &= -\nabla^2 f(x_k)^{-1} \nabla f(x_k) && \textit{direzione di Newton} \\
x_{k+1} &= x_k + d_k && \textit{iterazione di Newton}
\end{aligned}
$$

Il metodo di Newton (4)

Naturalmente, risulta:

$$\begin{aligned}\bar{q}_k(d) &= f(x_k) + \nabla f(x_k)^\top d + \frac{1}{2} d^\top \nabla^2 f(x_k) d \\ \nabla \bar{q}_k(d) &= \nabla f(x_k) + \nabla^2 f(x_k) d\end{aligned}$$

$\nabla \bar{q}_k(d) = 0$ quando $\nabla f(x_k) + \nabla^2 f(x_k) d = 0$.
Se $\nabla^2 f(x_k)$ è non singolare, allora possiamo definire:

$$\begin{aligned}d_k &= -\nabla^2 f(x_k)^{-1} \nabla f(x_k) && \textit{direzione di Newton} \\ x_{k+1} &= x_k + d_k && \textit{iterazione di Newton}\end{aligned}$$

Introduzione

Consideriamo il problema

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

Le condizioni di KKT per il problema danno luogo al sistema nonlineare

$$F(x,\mu) = \left[\begin{array}{c} \nabla f(x) + \nabla h(x)\mu \\ h(x) \end{array} \right] = 0$$

Idea: Risolvere il sistema $F(x,\mu)$ usando il **metodo di Newton**

Introduzione

Consideriamo il problema

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

Le condizioni di KKT per il problema danno luogo al sistema nonlineare

$$F(x,\mu) = \left[\begin{array}{c} \nabla f(x) + \nabla h(x)\mu \\ h(x) \end{array} \right] = 0$$

Idea: Risolvere il sistema $F(x,\mu)$ usando il **metodo di Newton**

Metodo di Newton-Lagrange

Se vogliamo usare il metodo di Newton, la prima cosa da fare è calcolare lo Jacobiano di F

$$\begin{bmatrix} \nabla^2_{xx} L(x,\mu) & \nabla h(x) \\ \nabla h(x)^\top & 0 \end{bmatrix}$$

Il metodo di Newton è definito dalla iterazione

$$\begin{pmatrix} x_{k+1} \\ \mu_{k+1} \end{pmatrix} = \begin{pmatrix} x_k \\ \mu_k \end{pmatrix} + \begin{pmatrix} \delta_k^x \\ \delta_k^\mu \end{pmatrix}$$

dove $(\delta_k^x, \delta_k^\mu)$ tali che

$$\begin{bmatrix} \nabla^2_{xx} L(x_k,\mu_k) & \nabla h(x_k) \\ \nabla h(x_k)^\top & 0 \end{bmatrix} \begin{pmatrix} \delta_k^x \\ \delta_k^\mu \end{pmatrix} = -F(x_k,\mu_k)$$

Metodo di Newton-Lagrange

Se vogliamo usare il metodo di Newton, la prima cosa da fare è calcolare lo Jacobiano di F

$$\begin{bmatrix} \nabla^2_{xx} L(x,\mu) & \nabla h(x) \\ \nabla h(x)^\top & 0 \end{bmatrix}$$

Il metodo di Newton è definito dalla iterazione

$$\begin{pmatrix} x_{k+1} \\ \mu_{k+1} \end{pmatrix} = \begin{pmatrix} x_k \\ \mu_k \end{pmatrix} + \begin{pmatrix} \delta_k^x \\ \delta_k^\mu \end{pmatrix}$$

dove $(\delta_k^x, \delta_k^\mu)$ tali che

$$\begin{bmatrix} \nabla^2_{xx} L(x_k,\mu_k) & \nabla h(x_k) \\ \nabla h(x_k)^\top & 0 \end{bmatrix} \begin{pmatrix} \delta_k^x \\ \delta_k^\mu \end{pmatrix} = -F(x_k,\mu_k)$$

Metodo di Newton-Lagrange

Se vogliamo usare il metodo di Newton, la prima cosa da fare è calcolare lo Jacobiano di F

$$\begin{bmatrix} \nabla^2_{xx} L(x,\mu) & \nabla h(x) \\ \nabla h(x)^\top & 0 \end{bmatrix}$$

Il metodo di Newton è definito dalla iterazione

$$\begin{pmatrix} x_{k+1} \\ \mu_{k+1} \end{pmatrix} = \begin{pmatrix} x_k \\ \mu_k \end{pmatrix} + \begin{pmatrix} \delta_k^x \\ \delta_k^\mu \end{pmatrix}$$

dove $(\delta_k^x, \delta_k^\mu)$ tali che

$$\begin{bmatrix} \nabla^2_{xx} L(x_k,\mu_k) & \nabla h(x_k) \\ \nabla h(x_k)^\top & 0 \end{bmatrix} \begin{pmatrix} \delta_k^x \\ \delta_k^\mu \end{pmatrix} = -F(x_k,\mu_k)$$

Metodo di Newton-Lagrange

Il passo k è ben definito quando la matrice

$$\begin{bmatrix} \nabla^2_{xx} L(x_k, \mu_k) & \nabla h(x_k) \\ \nabla h(x_k)^\top & 0 \end{bmatrix}$$

è **non singolare**. Questo accade sotto la seguente ipotesi:

Assunzione

- $rg(\nabla h(x_k)) = p$, *i.e. i gradienti dei vincoli in* x_k *sono lin. indipendenti*
- $d^\top \nabla^2_{xx} L(x_k, \mu_k) d > 0$ *per ogni* $d \neq 0$ *e tale che* $\nabla h(x_k)^\top d = 0$

Metodo di Newton-Lagrange

Il passo k è ben definito quando la matrice

$$\begin{bmatrix} \nabla^2_{xx} L(x_k, \mu_k) & \nabla h(x_k) \\ \nabla h(x_k)^\top & 0 \end{bmatrix}$$

è **non singolare**. Questo accade sotto la seguente ipotesi:

Assunzione

- $rg(\nabla h(x_k)) = p$, *i.e. i gradienti dei vincoli in* x_k *sono lin. indipendenti*
- $d^\top \nabla^2_{xx} L(x_k, \mu_k) d > 0$ *per ogni* $d \neq 0$ *e tale che* $\nabla h(x_k)^\top d = 0$

Metodo SQP

È possibile vedere il metodo di Newton-Lagrange in maniera alternativa. A tal fine, definiamo il problema

$$\begin{array}{ll} \min\limits_{d} & \frac{1}{2} d^\top \nabla^2_{xx} L(x_k, \mu_k) d + \nabla f(x_k)^\top d \\ c.v. & \nabla h(x_k)^\top d + h(x_k) = 0 \end{array}$$

Se vale l'assunzione di prima, questo problema ha una unica soluzione $(\bar{d}, \bar{\mu})$ che (ovviamente) soddisfa

$$\begin{array}{rcl} \nabla^2_{xx} L(x_k, \mu_k)\bar{d} + \nabla f(x_k) + \nabla h(x_k)\bar{\mu} & = & 0 \\ \nabla h(x_k)^\top \bar{d} + h(x_k) & = & 0 \end{array}$$

Metodo SQP

È possibile vedere il metodo di Newton-Lagrange in maniera alternativa. A tal fine, definiamo il problema

$$
\begin{array}{ll}
\min\limits_{d} & \frac{1}{2} d^\top \nabla^2_{xx} L(x_k, \mu_k) d + \nabla f(x_k)^\top d \\
c.v. & \nabla h(x_k)^\top d + h(x_k) = 0
\end{array}
$$

Se vale l'assunzione di prima, questo problema ha una unica soluzione $(\bar{d}, \bar{\mu})$ che (ovviamente) soddisfa

$$
\begin{aligned}
\nabla^2_{xx} L(x_k, \mu_k)\bar{d} + \nabla f(x_k) + \nabla h(x_k)\bar{\mu} &= 0 \\
\nabla h(x_k)^\top \bar{d} + h(x_k) &= 0
\end{aligned}
$$

Metodo SQP

È possibile vedere il metodo di Newton-Lagrange in maniera alternativa. A tal fine, definiamo il problema

$$
\begin{array}{ll}
\min_{d} & \frac{1}{2} d^\top \nabla^2_{xx} L(x_k, \mu_k) d + \nabla f(x_k)^\top d \\
c.v. & \nabla h(x_k)^\top d + h(x_k) = 0
\end{array}
$$

Se vale l'assunzione di prima, questo problema ha una unica soluzione $(\bar{d}, \bar{\mu})$ che (ovviamente) soddisfa

$$
\begin{array}{rcl}
\nabla^2_{xx} L(x_k, \mu_k)\bar{d} + \nabla f(x_k) + \nabla h(x_k)\bar{\mu} & = & 0 \\
\nabla h(x_k)^\top \bar{d} + h(x_k) & = & 0
\end{array}
$$

Metodo SQP

$$
\begin{aligned}
\nabla^2_{xx} L(x_k, \mu_k)\bar{d} + \nabla f(x_k) + \nabla h(x_k)\bar{\mu} &= 0 \\
\nabla h(x_k)^\top \bar{d} + h(x_k) &= 0
\end{aligned}
$$

lo possiamo scrivere come

$$
\begin{pmatrix} \nabla^2_{xx} L(x_k, \mu_k) & \nabla h(x_k) \\ \nabla h(x_k)^\top & 0 \end{pmatrix} \begin{pmatrix} \bar{d} \\ \bar{\mu} \end{pmatrix} = - \begin{pmatrix} \nabla f(x_k) \\ h(x_k) \end{pmatrix}
$$

e, sottraendo $\nabla h(x_k)\mu_k$ dalla prima eq.

$$
\begin{pmatrix} \nabla^2_{xx} L(x_k, \mu_k) & \nabla h(x_k) \\ \nabla h(x_k)^\top & 0 \end{pmatrix} \begin{pmatrix} \bar{d} \\ \bar{\mu} - \mu_k \end{pmatrix} = - \begin{pmatrix} \nabla f(x_k) + \nabla h(x_k)\mu_k \\ h(x_k) \end{pmatrix}
$$

Metodo SQP

Ponendo $\bar{d}_\mu = \bar{\mu} - \mu_k$ otteniamo proprio l'iterazione del metodo di Newton-Lagrange

$$\begin{pmatrix} x_{k+1} \\ \mu_{k+1} \end{pmatrix} = \begin{pmatrix} \bar{d} \\ \bar{d}_\mu \end{pmatrix} + \begin{pmatrix} x_k \\ \mu_k \end{pmatrix}$$

dove risulta $\mu_{k+1} = \bar{\mu}$

Metodo di soluzione

Algoritmo `SQP`

Dati: (x_0, μ_0), `maxit`

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

Calcola $\bar{d}_k$ e $\bar{\mu}_k$

Poni $x_{k+1} = x_k + \bar{d}_k$ e $\mu_{k+1} = \bar{\mu}_k$

if (x_{k+1}, μ_{k+1}) è KKT **then**

$x^* \leftarrow x_{k+1}$, $\mu^* \leftarrow \mu_{k+1}$ e STOP

endif

endfor

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

Metodo di soluzione

Algoritmo `SQP`

Dati: (x_0, μ_0), `maxit`

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

Calcola $\bar{d}_k$ e $\bar{\mu}_k$

Poni $x_{k+1} = x_k + \bar{d}_k$ e $\mu_{k+1} = \bar{\mu}_k$

if (x_{k+1}, μ_{k+1}) è KKT **then**

$x^* \leftarrow x_{k+1}$, $\mu^* \leftarrow \mu_{k+1}$ e STOP

endif

endfor

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

Metodo di soluzione

Algoritmo `SQP`

Dati: (x_0, μ_0), `maxit`

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

Calcola $\bar{d}_k$ e $\bar{\mu}_k$

Poni $x_{k+1} = x_k + \bar{d}_k$ e $\mu_{k+1} = \bar{\mu}_k$

if (x_{k+1}, μ_{k+1}) è KKT **then**

$x^* \leftarrow x_{k+1}$, $\mu^* \leftarrow \mu_{k+1}$ e STOP

endif

endfor

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

Vincoli di disuguaglianza

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

Allora, definisci il problema

$$\begin{array}{ll} \min\limits_d & \frac{1}{2} d^\top \nabla^2_{xx} L(x_k, \mu_k) d + \nabla f(x_k)^\top d \\ c.v. & \nabla h(x_k)^\top d + h(x_k) = 0 \\ & \nabla g(x_k)^\top d + g(x_k) \leq 0 \end{array}$$

Vincoli di disuguaglianza

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

Allora, definisci il problema

$$
\begin{array}{ll}
\min\limits_{d} & \frac{1}{2} d^\top \nabla^2_{xx} L(x_k, \mu_k) d + \nabla f(x_k)^\top d \\
c.v. & \nabla h(x_k)^\top d + h(x_k) = 0 \\
 & \nabla g(x_k)^\top d + g(x_k) \leq 0
\end{array}
$$

Proprietà di convergenza di SQP

Proposizione

Sia (x^*, μ^*, λ^*) *soluzione del problema tale che siano soddisfatte*

- *LICQ*[a]
- *stretta complementarità*
- *SOSC*[b]

Allora, se (x_0, μ_0, λ_0) *è suff. vicino a* (x^*, μ^*, λ^*)*, l'algoritmo* SQP *è ben definito e converge alla soluzione del problema*

[a] Linear Independence Constraint Qualification
[b] Second Order Sufficient Condition