Ottimizzazione dei Sistemi Complessi

G. Liuzzi[1]

Giovedì 3 Maggio 2018

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

Un bel passo indietro ...

Metodo del gradiente (rivisto)

Consideriamo il problema *NON vincolato*

$$\begin{array}{l} \min\ f(x) \\ s.t.\ x \in \mathbb{R}^n \end{array}$$

Disponiamo di una stima x_k della soluzione, quindi scriviamo l'app. di Taylor (troncata al I ordine) di $f(x)$ in x_k

$$\bar{f}(x; x_k) = f(x_k) + \nabla f(x_k)^\top (x - x_k) = \nabla f(x_k)^\top d + f(x_k)$$

Vogliamo determinare una nuova stima x_{k+1} minimizzando $\bar{f}(x; x_k)$

Metodo del Gradiente

Purtroppo, il problema

$$\begin{aligned} &\min \;\; \bar{f}(x; x_k) \\ &s.t. \; x \in \mathbb{R}^n \end{aligned}$$

se $\nabla f(x_k) \neq 0$, non ammette soluzione!!

Metodo del Gradiente

Sia $q_k(d) = f(x_k) + \nabla f(x_k)^\top d + \|d\|^2/2$ e consideriamo il problema

$$\begin{array}{l} \min \;\; q_k(d) \\ s.t. \; x \in \mathbb{R}^n \end{array}$$

La direzione $d^* = -\nabla f(x_k)$ è tale che

- $\nabla q_k(d^*) = 0$
- $\nabla^2 q_k(d^*) = \mathbb{I} \succ 0$

Metodo del Gradiente

Quindi, si definisce l'iterazione

$$x_{k+1} = x_k + d^* = x_k - \alpha_k \nabla f(x_k)$$

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 (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 (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

$$F(x,\mu) = \begin{bmatrix} \nabla f(x) + \nabla h(x)\mu \\ h(x) \end{bmatrix} = 0$$

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

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

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

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^*)

Proprietà di convergenza di SQP

Proposizione

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

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

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

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

Vincoli di disuguaglianza

Come si gestiscono in questo contesto vincoli di disuguaglianza $g(x) \leq 0$?

1) Aggiunta di variabili slack: $g_i(x) + s_i = 0$, $s_i \geq 0$ e gestione "esplicita" dei vincoli di bound sulla variabili (`SNOPT`, `KNITRO`)
2) Approccio S**i**QP - SQP with inequalities
3) Approccio S**e**QP - SQP with equalities (`SNOPT`)

Vincoli di disuguaglianza

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

Allora, data una stima corrente (x_k, μ_k, λ_k) della soluzione, definisci il problema

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

N.B. ogni sottoproblema ha esattamente lo stesso numero di vincoli del problema originario

Vincoli di disuguaglianza

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

Immaginiamo di poter conoscere quali vincoli di disuguaglianza sono attivi in una soluzione del problema x^*.
Indichiamo I^* l'insieme degli indici dei vincoli attivi in x^*

$$I^* = \{i : g_i(x^*) = 0\}$$

Quindi, $g_i(x^*) < 0$, per ogni $i \notin I^*$

Vincoli di disuguaglianza

Sotto questa ipotesi, anche il problema

$$
\begin{array}{ll}
\min_x & f(x) \\
c.v & h(x) = 0 \\
 & g_i(x) = 0, \quad i \in I^*
\end{array}
$$

ammette x^* come soluzione locale.
N.B. questo problema ha (nuovamente) tutti e soli vincoli di uguaglianza. Pertanto, posso usare il metodo di soluzione che abbiamo visto per problemi con vincoli di uguaglianza

Vincoli di disuguaglianza

Cosa manca? Normalmente non sappiamo come è fatto I^*. Quindi:

- si dà una stima I_k di I^*
- si risolve il sotto problema
- si aggiorna la stima di I^* definendo I_{k+1}

Per questo motivo questa classe di metodi è anche nota come "active-set" SQP

Problema approssimante

Data una stima dei vincoli attivi I_k ed una stima della soluzione (x_k, μ_k, λ_k), il problema "approssimante" è

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

Soluzione del problema approssimante

In base a ragionamenti del tutto analoghi a quelli appena visti, bisogna risolvere il sistema lineare:

$$\begin{bmatrix} \nabla^2_{xx} L_k & \nabla h(x_k) & \nabla g_{I_k}(x_k) \\ \nabla h(x_k)^\top & 0 & 0 \\ \nabla g_{I_k}(x_k)^\top & 0 & 0 \end{bmatrix} \begin{bmatrix} \bar{d} \\ \bar{\mu} \\ \bar{\lambda} \end{bmatrix} = - \begin{bmatrix} \nabla f(x_k) \\ h(x_k) \\ g_{I_k}(x_k) \end{bmatrix}$$

Metodo di soluzione (vincoli di uguaglianza)

Algoritmo `SQP`

Dati: (x_0, μ_0, λ_0), `maxit`, $\gamma > 0$

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

$I_k = \{i : g_i(x_k) \geq -\gamma(\lambda_k)_i\}$

Calcola $\bar{d}_k$, $\bar{\mu}_k$, $\bar{\lambda}_k$

Poni $x_{k+1} = x_k + \bar{d}_k$, $\mu_{k+1} = \bar{\mu}_k$,
$(\lambda_{k+1})_{I_k} = \bar{\lambda}_k$, $(\lambda_{k+1})i = 0$, $i \notin I_k$

if $(x_{k+1}, \mu_{k+1}, \lambda_{k+1})$ è KKT **then**

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

endif

endfor

Return: miglior punto trovato (x^*, μ^*, λ^*)