Ottimizzazione dei Sistemi Complessi

G. Liuzzi[1]

Giovedì 10 Marzo 2016

[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: $\nabla L(x,\mu) = 0$

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

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

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

Proposizione

Per ogni valore $\epsilon > 0$, ogni punto $(\bar{x}, \bar{\mu})$ di minimo locale di $L_a(x, \mu; \epsilon)$ è un punto di minimo locale del problema originario

È dunque possibile minimizzare L_a per determinare minimi locali del problema originario

Cosa manca?

Chi ci grarantisce che, se il problema originario "ammette soluzione", allora L_a ammette punti di minimo?

Terza proprietà

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

Sotto opportune ipotesi sulla coppia di KKT $(\bar{x}, \bar{\mu})$, è possibile dimostrare che

Proposizione

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

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

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

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

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

endif

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

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

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 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 – 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 $\mu_{k+1} = \mu_k$

Il solutore MINOS (5.x) – 1

Algoritmi implementati:

- Simplesso (Dantzig, 1947, 1963)
- quasi-Newton (Davidon, 1959, ...)
- Gradiente ridotto (Wolfe, 1962)
- **Lagrangiano aumentato** (proiettato) (Robinson, 1972)

Problemi che risolve:

- problemi lineari
- problemi nonlineari non vincolati
- problemi nonlineari con vincoli lineari
- **problemi nonlineari vincolati**

Il solutore MINOS (5.x) – 2

Considera problemi nella forma

$$\begin{array}{lllll} \min\limits_{x,y} & & F(x) + c^\top x + d^\top y & & \\ s.t. & b_1 & \leq f(x) + A_1 y \leq & b_2 & \text{(vincoli nonlin. in } x\text{)} \\ & b_3 & \leq A_2 x + A_3 y \leq & b_4 & \text{(vincoli lineari)} \\ & \ell & \leq (x, y) \leq & u & \text{(vincoli di bound)} \end{array}$$

Pertanto,

- x sono le variabilil **non lineari**
- y sono le variabili **lineari**

Il solutore MINOS (5.x) – 3

Sia (x_k, y_k) il punto all'inizio della k-esima iterazione.

$$\begin{aligned}\bar{f}(x, x_k) &= f(x_k) + \nabla f(x_k)^\top (x - x_k) = f_k + J_k(x - x_k) = \bar{f}_k(x) \\ f_d(x, x_k) &= f(x) - \bar{f}_k(x)\end{aligned}$$

N.B. $\bar{f}_k(x)$ è **lineare**

Considera il problema "quasi-linearizzato"

$$\begin{array}{lrcl}\min\limits_{x,y} & & F(x) + c^\top x + d^\top y & \\ s.t. & b_1 \leq & \bar{f}_k + A_1 y & \leq b_2 \\ & b_3 \leq & A_2 x + A_3 y & \leq b_4 \\ & \ell \leq & (x, y) & \leq u \\ & & f_d(x, x_k) & = 0\end{array}$$

Il solutore MINOS (5.x) – 4

A questo punto, interviene la funzione **Lagrangiana aumentata** ...

... e si considera il sottoproblema

$$
\begin{array}{lrcl}
\min\limits_{x,y} & F(x) + c^\top x + d^\top y & \boxed{+\mu_k^\top f_d + \frac{1}{2}\rho_k \|f_d\|^2} & \\
s.t. & b_1 \leq & \bar{f} + A_1 y & \leq b_2 \\
 & b_3 \leq & A_2 x + A_3 y & \leq b_4 \\
 & \ell \leq & (x, y) & \leq u
\end{array}
$$

N.B. per valori fissati di μ_k e ρ_k, è un "normale" problema con f.ob. nonlineare e vincoli lineari (metodo del Gradiente ridotto)

Il solutore MINOS (5.x) – 5

Output a video su un problema esempio (britgas.mod[2]):

```
Itn 0 -- linear constraints satisfied.
funcon sets 1466 out of 1466 constraint gradients.
funobj sets 407 out of 407 objective gradients.

Major minor step objective Feasible Optimal nsb ncon penalty BSwap
    1 0T 0.0E+00 0.00000E+00 1.0E+00 0.0E+00 0 3 1.0E+00 0
    2 260T 5.9E-01 2.30898E-01 9.5E-02 4.2E-01 4 59 1.0E+00 0
    3 63T 1.0E+00 6.54950E-03 4.6E-02 6.7E-01 2 124 1.0E+00 4
    4 44T 1.0E+00 0.00000E+00 5.1E-02 1.3E+00 6 205 1.0E+00 2
    5 43T 1.0E+00 0.00000E+00 5.4E-02 3.7E+00 7 294 1.0E+00 6
    6 45T 1.0E+00 0.00000E+00 5.1E-02 2.2E+00 10 373 1.0E+00 6
    7 46T 1.0E+00 0.00000E+00 1.2E-02 5.5E-01 10 452 1.0E+00 0
    8 42T 1.0E+00 0.00000E+00 2.9E-03 1.3E-01 13 536 1.0E+00 0
    9 43T 1.0E+00 0.00000E+00 7.1E-04 2.9E-02 15 624 1.0E+00 0
    10 41T 1.0E+00 0.00000E+00 9.5E-05 3.6E-03 17 718 1.0E+00 10

Major minor step objective Feasible Optimal nsb ncon penalty BSwap
    11 13 1.0E+00 0.00000E+00 1.8E-05 7.4E-05 16 749 1.0E+00 0
Completion Full now requested
    12 1 1.0E+00 0.00000E+00 3.2E-06 5.0E-06 16 752 1.0E-01 0
    13 1 1.0E+00 0.00000E+00 2.3E-08 1.2E-06 16 755 1.0E-02 0
    14 0 1.0E+00 0.00000E+00 4.4E-09 2.9E-07 16 756 1.0E-03 0
    15 0 1.0E+00 0.00000E+00 8.9E-10 7.0E-08 16 757 1.0E-04 0

    EXIT -- optimal solution found
```

[2] http://orfe.princeton.edu/~rvdb/ampl/nlmodels/cute