Ottimizzazione dei Sistemi Complessi

G. Liuzzi[1]

Martedì 23 Febbraio 2016

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

Pseudo-code del "compass search"

INPUT: x_0, Δ_0, Δ_{min}, $\mathtt{maxit}$, $D = \{d_i, i = 1, \dots, r\}$

$k \leftarrow 0$, $x \leftarrow x_0$, $\Delta \leftarrow \Delta_0$

while $k \leq \mathtt{maxit}$ **and** $\Delta \geq \Delta_{min}$ **do**

    $k \leftarrow k + 1$
    Let $\bar{d} \in D$ be s.t. $f(x + \Delta \bar{d}) = \min_{d_i \in D} f(x + \Delta d_i)$
    if $f(x + \Delta \bar{d}) < f(x)$ **then**
        $x \leftarrow x + \Delta \bar{d}$
    else
        $\Delta \leftarrow \Delta/2$
    endif

end while

RETURN: x (miglior punto determinato)

Pseudo-code di "compass search" rivisto

INPUT: x_0, Δ_0, Δ_{min}, $\mathtt{maxit}$, $D = \{d_i, i = 1, \dots, r\}$

$k \leftarrow 0$, $x \leftarrow x_0$, $\Delta \leftarrow \Delta_0$

while $k \leq \mathtt{maxit}$ **and** $\Delta \geq \Delta_{min}$ **do**

$k \leftarrow k + 1$

Let $\bar{d} \in D$ be s.t. $f(x + \Delta\bar{d}) = \min_{d_i \in D} f(x + \Delta d_i)$

if $f(x + \Delta\bar{d}) < f(x)$ **then**

if $\exists\ \bar{d} \in D$ s.t. $f(x + \Delta\bar{d}) < f(x)$ **then**

$x \leftarrow x + \Delta\bar{d}$

else

$\Delta \leftarrow \Delta/2$

endif

end while

RETURN: x (miglior punto determinato)

Un po' di analisi

Assunzione (A1)

L'insieme di livello $L(x_0) = \{x \in \mathbb{R}^n : \; f(x) \leq f(x_0)\}$ *è compatto*

N.B. nel metodo "compass search" il passo Δ, iterazione dopo iterazione,

- o diminuisce (di un fattore costante $\Delta \leftarrow \Delta/2$)
- o rimane costante

Lemma

Se (A1) è soddisfatta e $\Delta_{min} = 0$, *allora*

$$\lim_{k\to\infty} \Delta_k = 0.$$

Convergenza a zero del passo

Lemma

Se (A1) è soddisfatta e $\Delta_{min} = 0$*, allora*

$$\lim_{k\to\infty} \Delta_k = 0.$$

Dim.: Dalla def. dell'algoritmo segue che, per ogni k,

- $\Delta_k > 0$;
- $\Delta_{k+1} \leq \Delta_k$.

Quindi, $\lim_{k\to\infty} \Delta_k = \bar{\Delta} \geq 0$. Supponiamo che $\bar{\Delta} > 0$.
Per ogni k abbiamo che $\Delta_k = \Delta_0 / 2^{k_1}$ dove $k_1 \leq k + 1$ è il numero di volte che si riduce il passo nelle prime $k + 1$ iterazioni.
Esiste un intero $\bar{k}$ tale che, per $k \geq \bar{k}$ il passo non viene più ridotto. Quindi, per ogni $k \geq \bar{k}$

$$f(x_{k+1}) < f(x_k) \text{ e } \Delta_{k+1} = \Delta_k = \Delta_{\bar{k}} = \bar{\Delta}.$$

Convergenza a zero del passo (segue)

Quindi, per ogni $k \geq \bar{k}$, $\|x_{k+1} - x_k\| = \bar{\Delta}$.
Ora proviamo che la sottosucc. $\{x_k\}_{k \geq \bar{k}}$ è finita. Inidichiamo con

$$
\begin{aligned}
x_{\bar{k}+i} &= x_{\bar{k}} + \bar{\Delta} \sum_{r=0}^{i-1} d_{i_r}, \quad d_{i_r} \in D \\
x_{\bar{k}+j} &= x_{\bar{k}} + \bar{\Delta} \sum_{r=0}^{j-1} d_{j_r}, \quad d_{j_r} \in D
\end{aligned}
$$

$$
x_{\bar{k}+i} - x_{\bar{k}+j} = \bar{\Delta} \left(\sum_{r=0}^{i-1} d_{i_r} - \sum_{r=0}^{j-1} d_{j_r} \right)
$$

Convergenza a zero del passo (segue)

che implica (visto che $x_{\bar{k}+i}$ e $x_{\bar{k}+j}$ distinti e $\bar{\Delta} > 0$)

$$\left(\sum_{r=0}^{i-1} d_{i_r} - \sum_{r=0}^{j-1} d_{j_r}\right) \neq 0$$

ovvero

$$\left\|\sum_{r=0}^{i-1} d_{i_r} - \sum_{r=0}^{j-1} d_{j_r}\right\| \geq 1$$

e quindi

$$\|x_{\bar{k}+i} - x_{\bar{k}+j}\| \geq \bar{\Delta}.$$

Questo e (A1) implicano che $\{x_k\}_{k \geq \bar{k}}$ è finita. Quindi, devono esistere due interi k_1, k_2 tali che

$$x_{k_1} = x_{k_2}, \quad \bar{k} < k_1 < k_2$$

ma questo contraddice il fatto che $f(x_{k_2}) < f(x_{k_1})$.

Convergenza a punti stazionari

Per il metodo "compass search" rivisto vale il risultato

Teorema

Se vale (A1) e se $f(x)$ è continuamente differenziabile, allora

$$\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0.$$

(Ovvero almeno un punto limite di $\{x_k\}$ è stazionario)

Dim.

- $\{x_k\} \subset L(x_0)$;
- esiste $K_1 \subseteq \{0, 1, 2, \dots\}$ s.t. $\Delta_{k+1} = \Delta_k/2$.

Consideriamo la sottosuccessione $\{x_k\}_{k\in K_1}$ (contenuta in $L(x_0)$, ammette punti limite). $K_2 \subseteq K_1$ s.t.

- $\lim_{k\to\infty, k\in K_2} x_k = \bar{x}$;
- $\lim_{k\to\infty, k\in K_2} \Delta_k = 0$.

Convergenza a punti stazionari (segue)

Per ogni $k \in K_2$ e $i = 1, \ldots, n$, abbiamo che

$$\begin{aligned} f(x_k + \Delta_k e_i) &\geq f(x_k) \\ f(x_k - \Delta_k e_i) &\geq f(x_k) \end{aligned}$$

e, per il teorema della media,

$$\begin{aligned} f(x_k + \Delta_k e_i) &= f(x_k) + \Delta_k \nabla f(u_{k,i})^\top e_i \\ f(x_k - \Delta_k e_i) &= f(x_k) - \Delta_k \nabla f(v_{k,i})^\top e_i \end{aligned}$$

dove $u_{k,i} = x_k + \xi_{k,i} \Delta_k e_i$ e $v_{k,i} = x_k - \mu_{k,i} \Delta_k e_i$, con $\xi_{k,i}, \mu_{k,i} \in (0,1)$ e quindi $\lim_{k \to \infty, k \in K_2} u_{k,i} = \lim_{k \to \infty, k \in K_2} v_{k,i} = \bar{x}$.

Convergenza a punti stazionari (segue)

Segue anche che:

$$\begin{aligned} \Delta_k \nabla f(u_{k,i})^\top e_i &\geq 0 \\ -\Delta_k \nabla f(v_{k,i})^\top e_i &\geq 0. \end{aligned}$$

Dividendo per Δ_k e prendendo il limite per $k \to \infty, k \in K_2$ otteniamo

$$\begin{aligned} \lim_{k\to\infty, k\in K_2} \nabla f(u_{k,i})^\top e_i &= \nabla f(\bar{x})^\top e_i \geq 0 \\ \lim_{k\to\infty, k\in K_2} -\nabla f(v_{k,i})^\top e_i &= -\nabla f(\bar{x})^\top e_i \geq 0, \end{aligned}$$

e quindi, per indipendenza lineare di e_i, $i = 1, \ldots, n$, $\nabla f(\bar{x}) = 0$.

Convergenza del “compass search”

Il metodo “compass search” fa una esplorazione più approfondita dei punti vicini al punto corrente.

N.B. si sposta il punto corrente sostituendolo con quello, tra tutti, che grantisce il “miglior” decremento.

A questo corrisponde una proprietà di convergenza più forte. Infatti

Teorema

Se vale (A1) e se $f(x)$ è continuamente differenziabile, allora

$$\lim_{k\to\infty} \|\nabla f(x_k)\| = 0.$$

Ovvero **tutti** i punti limite di $\{x_k\}$ sono stazionari.

Il metodo di Nelder-Mead

Alla generica iterazione k lo stato dell'algoritmo è dato da:

$$X_k = \{x_1, x_2, \dots, x_{n+1}\}$$

indichiamo $f_i = f(x_i)$ e supponiamo X_k ordinato in modo che

$$f_1 \leq f_2 \leq \dots \leq fx_{n+1}.$$

Ricordiamo l'espressione del "centroide" dei primi n punti in X_k

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Idea: Sfruttare x_{n+1} e $\bar{x}$ per cercare un punto "migliore" di x_{n+1}

Il metodo di Nelder-Mead

Dato un parametro μ, definiamo

$$x(\mu) = \bar{x} + \mu(\bar{x} - x_{n+1})$$

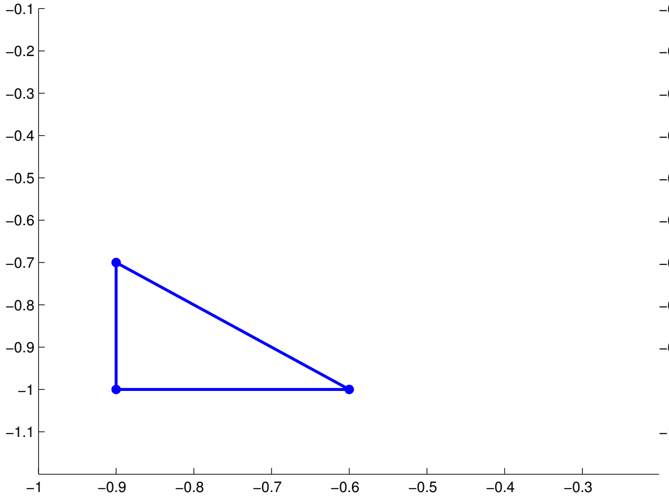
Il metodo usa: $-1 < \mu_{ic} < 0 < \mu_{oc} < \mu_r < \mu_e$ e

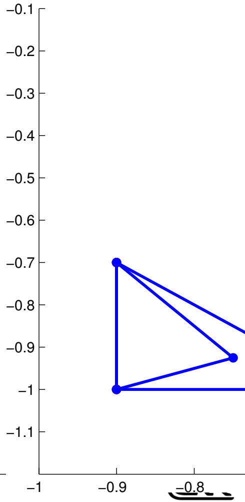
$$\begin{array}{llllll}
\text{(inner contraction)} & x^{ic} & = & x(\mu_{ic}), & f^{ic} & = & f(x^{ic}) \\
\text{(outer contraction)} & x^{oc} & = & x(\mu_{oc}), & f^{oc} & = & f(x^{oc}) \\
\text{(reflect)} & x^{r} & = & x(\mu_{r}), & f^{r} & = & f(x^{r}) \\
\text{(expand)} & x^{e} & = & x(\mu_{e}), & f^{e} & = & f(x^{e})
\end{array}$$

Iterazione k

- Se $\boxed{f_1 \le f^r < f_n}$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^r\}$ **FINE**
- Se $\boxed{f^r < f_1}$, allora
  - Se $f^e < f^r$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^e\}$ **FINE**
  - altrimenti $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^r\}$ **FINE**
- Se $\boxed{f_n \le f^r < f_{n+1}}$, allora
  - Se $f^{oc} \le f^r$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^{oc}\}$ **FINE**
  - altrimenti **shrink**
- Se $\boxed{f_{n+1} \le f^r}$, allora
  - Se $f^{ic} < f_{n+1}$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^{ic}\}$ **FINE**
  - altrimenti **shrink**
- **shrink**:
  - $X_{k+1} = \{x_1, \hat{x}_2, \dots, \hat{x}_{n+1}\}$ dove
  - $\hat{x}_i = x_1 + \gamma(x_i - x_1)$, $i = 2, \dots, n+1$, $\gamma \in (0, 1)$

Esempio su Funzione di Broyden





È un “buon” metodo?

La risposta è:

- generalmente **Si**
- però non ha proprietà di convergenza a punti stazionari
- esistono contro-esempi sui quali converge a punti **NON** stazionari
- p.es. la funzione di McKinnon

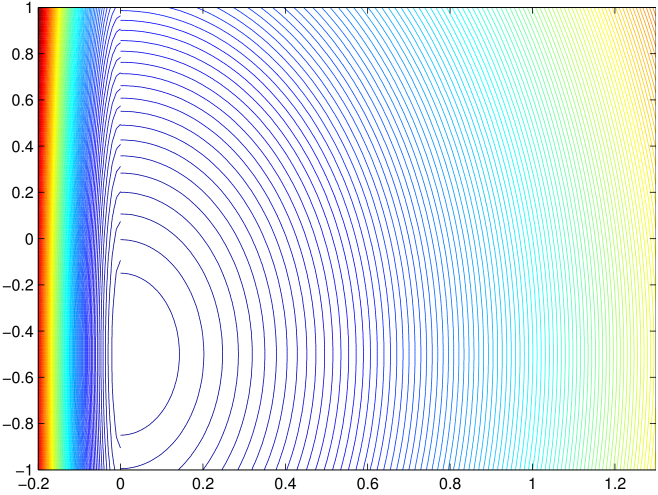
La funzione di McKinnon

$$f(x,y) = \begin{cases} \theta\phi|x|^\tau + y + y^2 & \text{se } x \le 0 \\ \theta x^\tau + y + y^2 & \text{se } x > 0 \end{cases}$$

- strettamente convessa per $\tau > 1$
- continuamente differenziabile per $\tau > 1$
- due volte cont. differenziabile per $\tau > 2$
- tre volte cont. differenziabile per $\tau > 3$

La funzione di McKinnon

Per $\tau = 2$, $\theta = 6$, $\phi = 60$, la funzione è



La funzione di McKinnon

Se inizializziamo il metodo di Nelder-Mead con il simplesso

$$X = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} \frac{1+\sqrt{33}}{88} \\ \frac{1-\sqrt{33}}{88} \end{pmatrix}, \right\}$$

La funzione di McKinnon