Ottimizzazione dei Sistemi Complessi

G. Liuzzi[1]

Giovedì 15 Marzo 2018

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

pseudo-codice "implicit filtering"

INPUT: x_0, Δ_0, Δ_{min}, $\mathtt{maxit}$, $D = \{\pm e_i, i = 1, \dots, n\}$

$k \leftarrow 0$

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

  if $\exists\ \bar{d} \in D$ s.t. $f(x_k + \Delta_k \bar{d}) < f(x_k) - \gamma\Delta_k$ **then**

    $x_{k+1} \leftarrow x_k + \Delta_k \bar{d}$, $\Delta_{k+1} \leftarrow \Delta_k$

  else

    implicit filtering step

  endif

  $k \leftarrow k + 1$

end while

RETURN: $\{x_k\}, \{\Delta_k\}$ successioni di punti e passi

pseudo-codice "implicit filtering"

Implicit filtering step:

$d_k = -\nabla_{\Delta_k} f(x_k)$

if $\|d_k\| > \tau\Delta_k$ **and** $f(x_k + \Delta_k d_k) \le f(x_k) - \gamma\Delta_k\|d_k\|^2$

compute $\alpha_k = 2^\beta \Delta_k$ s.t.

$$\begin{aligned} f(x_k + \alpha_k d_k) &\le f(x_k) - \gamma\alpha_k\|d_k\|^2 \\ f(x_k + 2\alpha_k d_k) &> f(x_k) - \gamma 2\alpha_k\|d_k\|^2 \end{aligned}$$

if $\alpha_k\|d_k\|^2 > \tau\Delta_k$ **then**

$x_{k+1} \leftarrow x_k + \alpha_k d_k$, $\Delta_{k+1} \leftarrow \Delta_k$

else

$\Delta_{k+1} \leftarrow \Delta_k/2$, $x_{k+1} \leftarrow x_k$

endif

else

$\Delta_{k+1} \leftarrow \Delta_k/2$, $x_{k+1} \leftarrow x_k$

endif

Analisi di Convergenza

Quale proprietà si dimostra per prima ?

Assunzione (A1)

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

Lemma

Se (A1) è soddisfatta, $\Delta_{min} \leq 0$ *e* $\texttt{maxit} = +\infty$*, allora* $\{\Delta_k\}$ *è una succ. infinita e*

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

Convergenza a zero del passo (segue)

Dim.: Se $\Delta_{min} \leq 0$ e $\mathtt{maxit} = +\infty$, la condizione del **while** più esterno è sempre vera, quindi l'algoritmo non termina mai, producendo una succ. $\{\Delta_k\}$ infinita.

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 $\bar{\Delta} > 0$.
Esiste un intero $\bar{k}$ tale che, per $k \geq \bar{k}$ il passo non viene più ridotto. Quindi, per ogni $k \geq \bar{k}$ risulta $\Delta_{k+1} = \Delta_k = \Delta_{\bar{k}} = \bar{\Delta}$.
Per $k \geq \bar{k}$, o

- $f(x_{k+1}) \leq f(x_k) - \gamma\bar{\Delta} \leq f(x_k) - \gamma\tau\bar{\Delta}$ oppure
- $f(x_{k+1}) \leq f(x_k) - \gamma\alpha_k\|d_k\|^2 \leq f(x_k) - \gamma\tau\bar{\Delta}$

Convergenza a zero del passo (segue)

Quindi, per ogni $k \geq \bar{k}$, possiamo scrivere

$$f(x_{k+1}) \leq f(x_k) - \gamma\tau\bar{\Delta} \leq f(x_{k-1}) - 2\gamma\tau\bar{\Delta} \leq \cdots \leq f(x_{\bar{k}}) - (k - \bar{k} + 1)\gamma\tau\bar{\Delta}$$

Quando $k \to \infty$, la relazione di sopra comporta che $\lim_{k\to\infty} f(x_k) = -\infty$ che è in contraddizione con la continuità di f e compattezza di $L(x_0)$. $\square$

Convergenza a punti stazionari

Teorema

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

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

Dim.

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

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

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

Convergenza a punti stazionari (segue)

Possiamo partizionare l'insieme K in due sottoinsiemi, K_1 e $K_2 = K \setminus K_1$ con

$$K_1 = \{k \in K : \|d_k\| \leq \tau \Delta_k\}$$

(N.B. K è infinito quindi almeno uno dei due insiemi deve essere infinito)

Supponiamo che K_1 sia infinito. Allora possiamo prendere il limite per $k \to \infty$, $k \in K_1$ nella relazione $\|d_k\| \leq \tau \Delta_k$

$$0 = \lim_{k \to \infty, k \in K_1} \tau \Delta_k \geq \lim_{k \to \infty, k \in K_1} \|d_k\| = \|\nabla f(\bar{x})\| \geq 0$$

che prova che $\bar{x}$ è stazionario per f.

Convergenza a punti stazionari (segue)

Supponiamo ora che K_1 sia finito. Allora K_2 è infinito. In questo caso, per $k \in K_2$ può succedere:

(1) $f(x_k + \Delta_k d_k) > f(x_k) - \gamma \Delta_k \|d_k\|^2$ oppure

(2) $f(x_k + \Delta_k d_k) \leq f(x_k) - \gamma \Delta_k \|d_k\|^2$ cioè il metodo calcola $\alpha_k = 2^\beta \Delta_k$ tale che

$$\begin{aligned} f(x_k + \alpha_k d_k) &\leq f(x_k) - \gamma \alpha_k \|d_k\|^2 \\ f(x_k + 2\alpha_k d_k) &> f(x_k) - \gamma 2\alpha_k \|d_k\|^2 \end{aligned}$$

e, siccome $k \in K$,

$$\alpha_k \|d_k\|^2 \leq \tau \Delta_k$$

Convergenza a punti stazionari (segue)

Procediamo per contraddizione assumendo che $\|\nabla f(\bar{x})\| > 0$.

Supponiamo che infinite volte capiti (1). Applichiamo il Teorema della media e otteniamo

$$f(x_k + \Delta_k d_k) - f(x_k) = \Delta_k \nabla f(\xi_k)^\top d_k > -\gamma \Delta_k \|d_k\|^2$$

con $\xi_k = x_k + t_k \Delta_k d_k$ e $t_k \in (0,1)$.
Semplificando e prendendo il limite per $k \to \infty$ otteniamo

$$-\|\nabla f(\bar{x})\|^2 \geq -\gamma \|\nabla f(\bar{x})\|^2$$

ovvero $1 - \gamma \leq 0$ che contraddice l'ipotesi $\gamma < 1$.

Convergenza a punti stazionari (segue)

Supponiamo che infinite volte capiti (2). Questa volta

$$\alpha_k \|d_k\|^2 \leq \tau \Delta_k$$

Prendendo il limite per $k \to \infty$ e ricordando che $\Delta_k \to 0$ e $\|\nabla f(\bar{x})\| > 0$, otteniamo

$$\lim_{k \to \infty, (2)} \alpha_k = 0.$$

Il punto $x_k + \alpha_k d_k$ non è accettato dal metodo questa volta ma vale ugualmente

$$f(x_k + 2\alpha_k d_k) > f(x_k) - \gamma 2\alpha_k \|d_k\|^2.$$

per cui si può ragionare come nel punto precedente e concludere la dimostrazione. $\square$

Testo d'esame

Si consideri il problema non vincolato seguente:

$$\min_{x,y} \; f(x,y),$$

con $f(x,y) = \max\{x^2 + y^2, (x-1)^2 + y^2\}$.

Siano $x_0 = (0,0)^\top$ e $\Delta_0 = 1$, il punto ed il passo iniziali del metodo **Compass Search** (forte).

- Scrivere i punti di tentativo (e relativi valori di funzione) del metodo nella sua prima iterazione.
- Scrivere il punto x_1, determinato dal metodo alla fine della prima iterazione.

Siano ora $x_0 = (0,0)^\top$ e $\Delta_0 = 0.5$.

- Scrivere i punti di tentativo (e relativi valori di funzione) del metodo nella sua prima iterazione.
- Scrivere il punto x_1, determinato dal metodo alla fine della prima iterazione.

Il metodo di Nelder-Mead

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

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

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

$$f_1 \leq f_2 \leq \cdots \leq f_{n+1}.$$

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

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

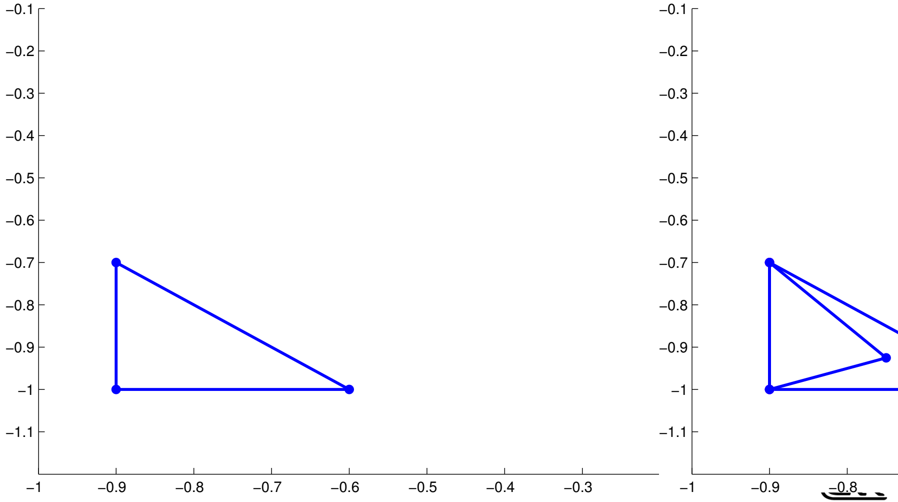
Il metodo usa: $-1 < \mu_{ic} < 0 < \mu_{oc} < \mu_r < \mu_e$ e

(inner contraction)	x^{ic}	$=$	$x(\mu_{ic})$,	f^{ic}	$=$	$f(x^{ic})$
(outer contraction)	x^{oc}	$=$	$x(\mu_{oc})$,	f^{oc}	$=$	$f(x^{oc})$
(reflect)	x^{r}	$=$	$x(\mu_{r})$,	f^{r}	$=$	$f(x^{r})$
(expand)	x^{e}	$=$	$x(\mu_{e})$,	f^{e}	$=$	$f(x^{e})$

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



Esercizio

Sia

$$X = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$$

$$F = \{ \quad 3, \quad 7, \quad 1, \quad \}$$

$$X = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

$$F = \{ \quad 1, \quad 3, \quad 7, \quad \}$$

Il centroide $\bar{x}$ di x_1 e x_2 (dopo il riordino) è

$$\bar{x} = \begin{pmatrix} 1 \\ 1/2 \end{pmatrix}$$

Il metodo considera i punti

$$x(\mu) = \bar{x} + \mu d = \bar{x} + \mu(\bar{x} - x_3)$$

Esercizio

- si calcola $d = \bar{x} - x_3 = (1,\ -1/2)^\top$
- si calcola $x_r = x(1) = \bar{x} + d = (2,\ 0)^\top$
- si calcola $x_e = x(2) = \bar{x} + 2d = (3,\ -1/2)^\top$
- si calcola $x_{oc} = x(1/2) = \bar{x} + 1/2d = (3/2,\ 1/4)^\top$
- si calcola $x_{ic} = x(-1/2) = \bar{x} - 1/2d = (1/2,\ 3/4)^\top$

Nelder&Mead

In $\mathbb{R}^2$ siano dati i seguenti punti:

$$x_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, x_2 = \begin{pmatrix} 3 \\ 2 \end{pmatrix}, x_3 = \begin{pmatrix} 1 \\ 3 \end{pmatrix},$$

a cui corrispondono: $f(x_1) = 1$, $f(x_2) = 3$, $f(x_3) = 5$.

Determinare i punti x_r, x_e, x_{oc} e x_{ic} utilizzati nel metodo di Nelder&Mead

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

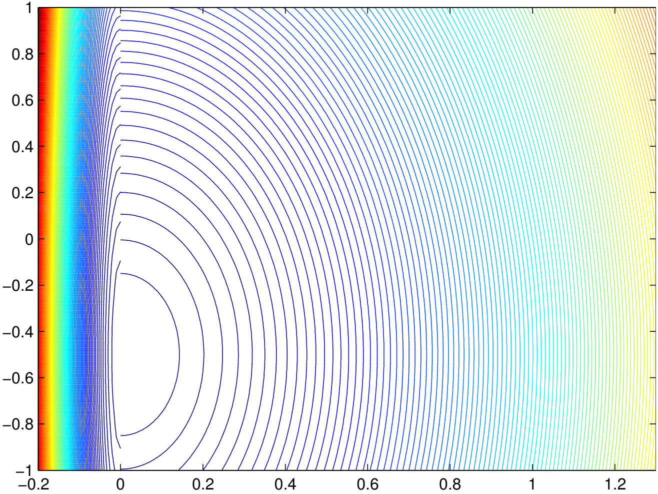
La funzione di McKinnon

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

- 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

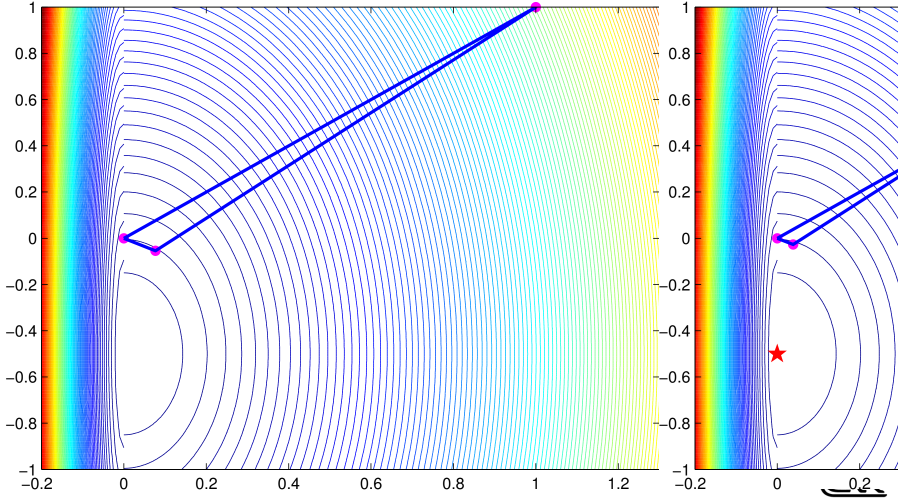


La funzione di McKinnon

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

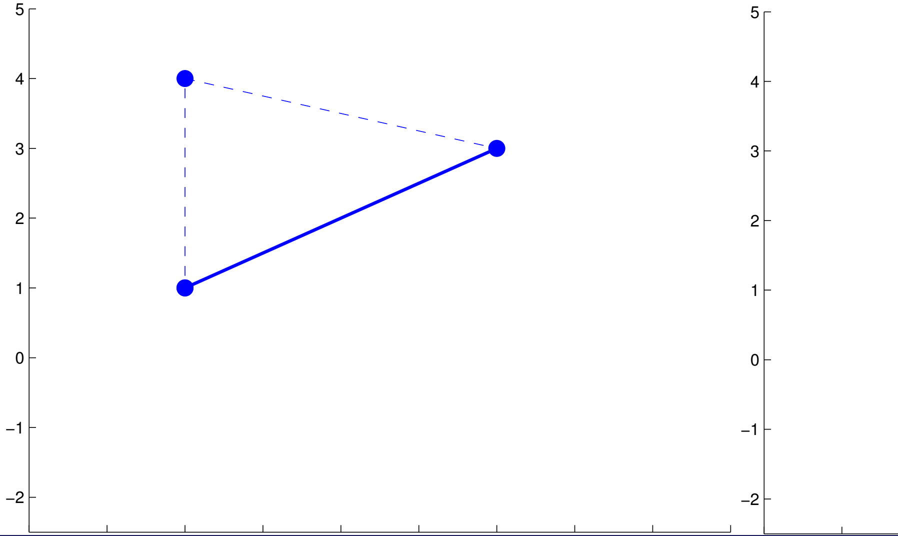


Perchè...

- Compass Search "funziona" ?
- Hooke & Jeeves "funziona" ?
- Nelder & Mead "**non** funziona" (teoricamente) ?

Cosa "davvero" distingue i primi due dal terzo ?

Perchè...



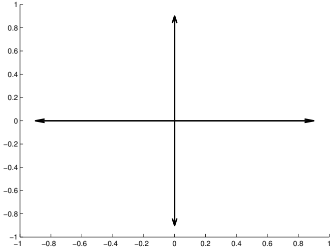
... e quindi ...

Quando $D = \{\pm e_1, \pm e_2, \ldots, \pm e_n\}$

$$\kappa(D) = \frac{1}{\sqrt{n}} > 0$$

Comunque scelto un $v \in \mathbb{R}^n$, esiste $d \in D$ tale che

$$\frac{v^\top d}{\|v\|\|d\|} = \cos\theta(d,v) \geq \kappa(D) > 0$$



Quando $v = -\nabla f(x) \neq 0$, esiste sempre una direzione $d \in D$ **di discesa**

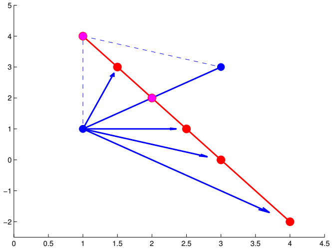
... e quindi ...

Quando $D = \{\text{vedi figura}\}$

$$\kappa(D) \leq 0$$

esistono $v \in \mathbb{R}^n$ tali che, per ogni $d \in D$, si ha

$$\frac{v^\top d}{\|v\|\|d\|} = \cos\theta(d, v) \leq 0$$



Potrebbe capitare che $\nabla f(x) \neq 0$, e ogni $d \in D$ è tale che

$$-\nabla f(x)^\top d \leq 0$$

cioè **nessuna** dir. in D è **di discesa**

... e quindi ...

per il metodo **compass search** che usa

- $D = \{\pm e_1, \ldots, \pm e_n\}$,
- $\kappa(D) = 1/\sqrt(n) > 0$

quando il punto corrente **non** è stazionario ($\nabla f(x) \neq 0$)

- esiste sempre (almeno) una direzione $d \in D$ che è **di discesa**

$$\nabla f(x)^\top d < 0$$

- per passi Δ sufficientemente piccoli **deve accadere**

$$f(x + \Delta d) < f(x)$$

cioè il punto x corrente (se non-stazionario) prima o dopo viene aggiornato

... e quindi ...

per il metodo **Nelder&Mead** che usa

- D t.c. $\kappa(D) \leq 0$

anche se il punto corrente **non** è stazionario ($\nabla f(x) \neq 0$)

- le direzioni $d \in D$ potrebbero tendere a direzioni $\bar{d}$ che **non** sono di discesa
$$\nabla f(x)^\top \bar{d} \geq 0$$
- anche per spostamenti piccolissimi (i.e. punti in X vicinissimi) nessuno dei punti x_r, x_e, x_{oc}, x_{ic} migliora il valore $f(x)$

cioè il punto x corrente non-stazionario non viene **mai** aggiornato!!