Ottimizzazione dei Sistemi Complessi

G. Liuzzi[1]

Lunedì 29 Febbraio 2016

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

Ricapitolando

Fin qui abbiamo visto:

- Compass search (Fermi-Metropolis 1952, forte)
- Compass search rivisto (o debole)
- Nelder-Mead (1965, controesempio)

Ora vediamo una "variante" del compass search rivisto, introdotta da **R. Hooke** e **T. Jeeves** nel 1961

Compass search "rivisto" o debole

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

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

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

  $k \leftarrow k + 1$

  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 nuovo metodo ?

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

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

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

  $k \leftarrow k + 1$, $y \leftarrow x$
  for $i = 1, \dots, r$
    if $f(y + \Delta d_i) < f(y)$ **then** $y \leftarrow y + \Delta d_i$
  end for
  if $f(y) < f(x)$ **then**
    $x \leftarrow y$
  else $\Delta \leftarrow \Delta/2$

end while

RETURN: x (miglior punto determinato)

Domande? Lo riconoscete?

Il metodo di Hooke&Jeeves

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**

$k \leftarrow k + 1$, $y \leftarrow x$
for $i = 1, \dots, r$ (*exploratory moves from x*)
  if $f(y + \Delta d_i) < f(y)$ **then** $y \leftarrow y + \Delta d_i$
end for
if $f(y) < f(x)$ **then** (*pattern move along* $y - x$)
  $z \leftarrow y + (y - x)$
  for $i = 1, \dots, r$ (*exploratory moves from z*)
    if $f(z + \Delta d_i) < f(z)$ **then** $z \leftarrow z + \Delta d_i$
  end for
  if $f(z) < f(y)$ **then** $x \leftarrow z$ **else** $x \leftarrow y$
else $\Delta \leftarrow \Delta/2$

end while

RETURN: x (miglior punto determinato)

Convergenza di Hooke&Jeeves

Il metodo di H&J altro non è se non una *evoluzione* del compass search debole (o rivisto)

Non stupisce che H&J abbia le medesime proprietà teoriche e cioè

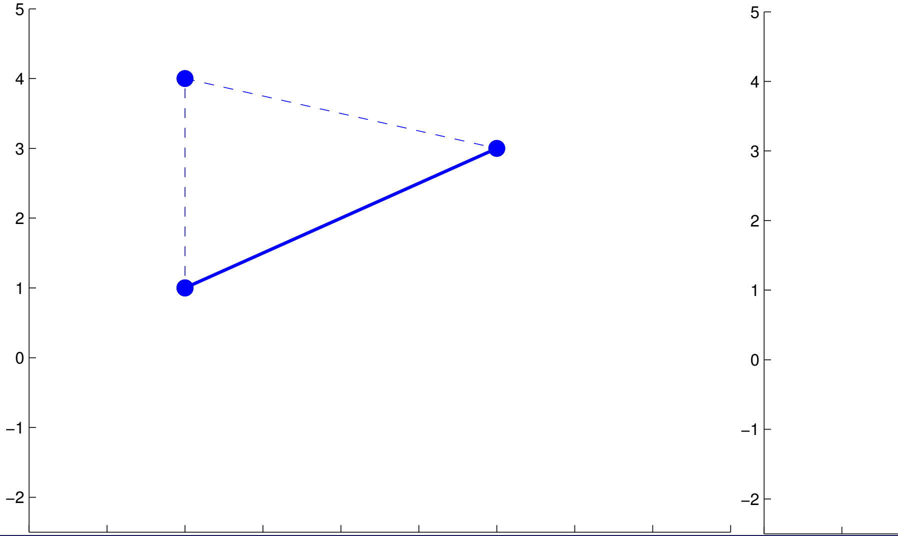
- $\lim_{k\to\infty} \Delta_k = 0$
- almeno un punto limite è stazionario
  - $\nabla f(\bar{x}) = 0$
  - $\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0$

Perchè...

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

Cosa “davvero” distingue i primi due dal terzo ?

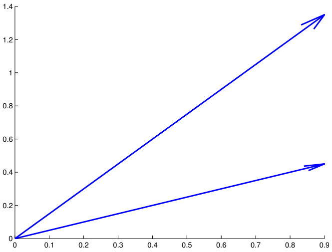
Perchè...



Angolo tra due vettori

Coseno dell'angolo tra due vettori in $\mathbb{R}^n$:

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

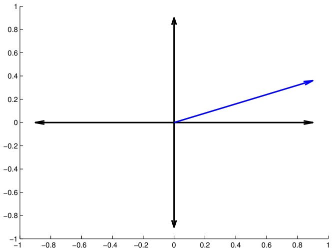


Angolo vettore – insieme di vettori

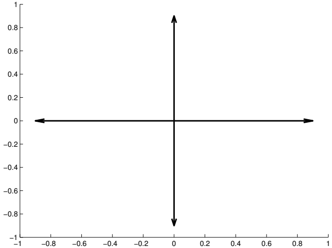
Dati:

- un insieme finito di vettori (direzioni) D
- un vettore $v \in \mathbb{R}^n$

$$\phi(v) = \max_{d \in D} \frac{v^\top d}{\|v\|\|d\|} = \max_{d \in D} \cos\theta(d, v)$$



Misura del coseno (di un ins. di vettori)

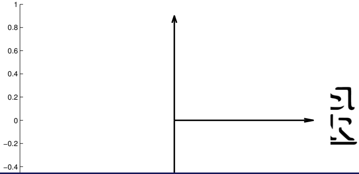


$$\phi(v) = \max_{d \in D} \frac{v^\top d}{\|v\| \|d\|} = \max_{d \in D} \cos \theta(d, v)$$

$$\kappa(D) = \frac{1}{\sqrt{2}}$$

Dato un insieme finito di vettori (direzioni) D

$$\kappa(D) = \min_{v \in \mathbb{R}^n} \phi(v) = \min_{v \in \mathbb{R}^n} \max_{d \in D} \frac{v^\top d}{\|v\| \|d\|}$$



... e quindi ...

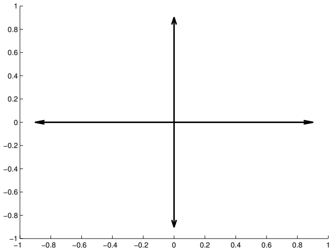
Quando $D = \{\pm e_1, \pm e_2, \ldots, \pm e_n\}$

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

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**



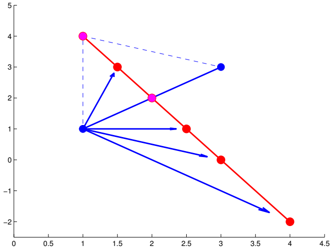
... e quindi ...

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!!

Un po' di definizioni ...

Sia $D = \{d_1, \dots, d_r\} \subset \mathbb{R}^n$ un insieme finito di vettori

Definizione (Span positivo generato da D)

$$PSpan(D) = \left\{ x \in \mathbb{R}^n : \ x = \sum_{i=1}^{r} \lambda_i d_i, \qquad \lambda_i \geq 0 \right\}$$

Definizione (Generatore positivo di $\mathbb{R}^n$)

Se $PSpan(D) = \mathbb{R}^n$, allora D è un generatore positivo di $\mathbb{R}^n$

Definizione (Base positiva di $\mathbb{R}^n$)

Una base positiva di $\mathbb{R}^n$ è un generatore positivo D minimale

... e qualche proprietà ...

$D = \{d_1, \ldots, d_r\} \subset \mathbb{R}^n$

- Se $PSpan(D) = \mathbb{R}^n$, allora D deve contenere $r-1$ generatori di $\mathbb{R}^n$
- Se D è una base positiva di $\mathbb{R}^n$, allora $n+1 \leq r \leq 2n$ in $\mathbb{R}^n$ sono basi positive, p.es.

$$\begin{aligned} D &= \{e_1, \ldots, e_n, \xi\},\ \xi = -\sum_{i=1}^{n} e_i \\ D &= \{\pm e_1, \ldots, \pm e_n\} \end{aligned}$$

- Se $PSpan(D) = \mathbb{R}^n$ e $v^\top d \geq 0$, per ogni $d \in D$, allora $v = 0$
- $PSpan(D) = \mathbb{R}^n$ **se e solo se**, per ogni $v \in \mathbb{R}^n$ $(v \neq 0)$, esiste $d \in D$: $v^\top d > 0$
- Se $PSpan(D) = \mathbb{R}^n$, allora $\kappa(D) > 0$