Ottimizzazione dei Sistemi Complessi

G. Liuzzi[1]

Giovedì 8 Marzo 2018

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

Il problema

Obiettivo: minimizzare una funzione $f(x)$ di n variabili reali per la quale non è data una procedura che restituisce le derivate di f in un punto x

- Metodi che usano il gradiente (naturalmente **approssimato**)
- Metodi che **NON** usano approssimazioni del gradiente

Formule di approssimazione delle derivate

- differenze finite in avanti (FFD):

$$\frac{\partial f}{\partial x_i} \simeq \frac{f(x+\epsilon e_i) - f(x)}{\epsilon} \quad \text{con } e_i \text{ } i\text{-esima colonna di } I$$

- differenze finite all'indietro (BFD):

$$\frac{\partial f}{\partial x_i} \simeq \frac{f(x) - f(x-\epsilon e_i)}{\epsilon}$$

- differenze finite centrali (CFD):

$$\frac{\partial f}{\partial x_i} \simeq \frac{f(x+\epsilon e_i) - f(x-\epsilon e_i)}{2\epsilon}$$

Metodo che usa il gradiente approssimato

- Calcolo $d = -\nabla_\epsilon f(x)$
- Determino un passo α t.c.

$$\begin{aligned} f(x+\alpha d) &\le f(x) + \gamma\alpha\nabla_\epsilon f(x)^\top d \quad \text{(suff. riduzione)} \\ f(x+\delta\alpha d) &> f(x) + \gamma\delta\alpha\nabla_\epsilon f(x)^\top d \quad \text{(suff. spostamento)} \end{aligned}$$

  con $\gamma \in (0, 1/2)$ e $\delta > 1$
- Pongo $x \leftarrow x + \alpha d$

Pseudo-codice

INPUT: x_0, Δ_0, Δ_{min}, $\gamma \in (0, 1/2)$, $\beta \in (0, 1)$, maxit

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

while $k \leq$ maxit **and** $\|\nabla_\epsilon f(x)\| \geq \Delta_{min}$ **do**

  $k \leftarrow k + 1$, $d = -\nabla_\epsilon f(x)$

  Let $\alpha \leftarrow \Delta$

  if $f(x + \alpha d) > f(x) + \gamma \alpha \nabla_\epsilon f(x)^\top d$ **then**

    while $f(x + \alpha d) > f(x) + \gamma \alpha \nabla_\epsilon f(x)^\top d$

      $\alpha \leftarrow \beta \alpha$

  else

    $\tilde{\alpha} \leftarrow \alpha / \beta$

    while $f(x + \tilde{\alpha} d) \leq f(x) + \gamma \tilde{\alpha} \nabla_\epsilon f(x)^\top d$

      $\alpha \leftarrow \tilde{\alpha}$, $\tilde{\alpha} \leftarrow \alpha / \beta$

  endif

  $x \leftarrow x + \alpha d$, $\Delta \leftarrow \alpha$

end while

RETURN: x (miglior punto determinato)

Pseudo-codice del "compass search"

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

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

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

  $k \leftarrow k + 1$

  If $\exists\ \bar{d} \in D$ be s.t. $f(x + \Delta\bar{d}) < f(x)$

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

  else

    $\Delta \leftarrow \Delta/2$

  endif

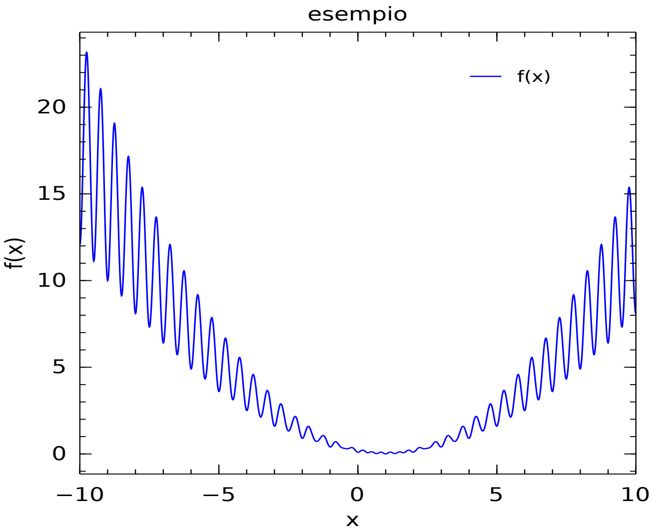
end while

RETURN: x (miglior punto determinato)

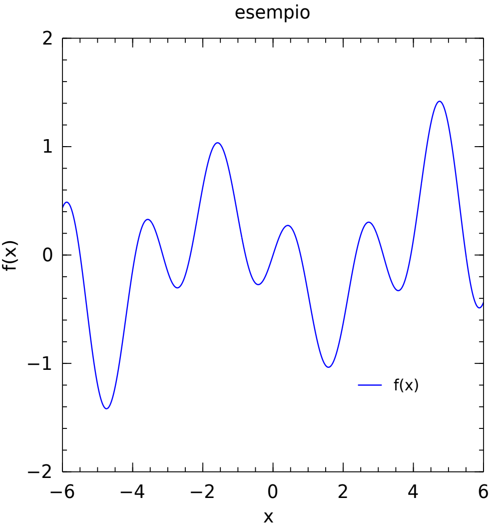
Esempio

Confronto su un problema in Julia

Esempio, molti minimi locali



Esempio



Perché?

Supponiamo $f(x)$ sia affetta da rumore additivo quindi

$$\tilde{f}(x) = f(x) + \epsilon$$

Differenze finite su $\tilde{f}$:

$$\begin{aligned} \frac{\partial \tilde{f}}{\partial x_i}(x) &\cong \frac{\tilde{f}(x+\Delta e_i) - \tilde{f}(x)}{\Delta} = \frac{f(x+\Delta e_i) + \epsilon_1 - f(x) - \epsilon_2}{\Delta} \\ &\cong \frac{\partial f}{\partial x_i}(x) + \frac{\epsilon_1 - \epsilon_2}{\Delta} \end{aligned}$$

Problema: per avere una buona approssimazione devo usare $\Delta \ll 1$ **ma** in questo caso

$$\frac{\epsilon_1 - \epsilon_2}{\Delta} \gg 1!!!$$

Compass Search

Consideriamo il problema:

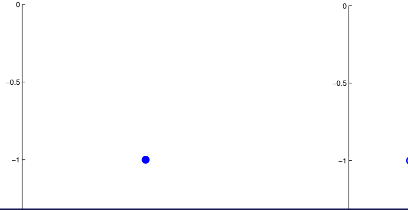
$$\min_{x \in \mathbb{R}^2} f(x)$$

Punto iniziale: $x_0 = (-0.9; -1.0)^\top$
$f(x)$ iniziale: $f(x_0) = 11.3524$
Passo iniziale: $\Delta = 0.3$

Punto corrente: $x_k = (-0.9; -0.7)^\top$
$f(x)$ corrente: $f(x_k) = 5.0788$
Passo corrente: $\Delta = 0.3$

Punto corrente: $x_k = (-0.6; -0.7)^\top$
$f(x)$ corrente: $f(x_k) = 2.2048$
Passo corrente: $\Delta = 0.3$

Punto corrente: $x_k = (-0.6; -0.4)^\top$
$f(x)$ corrente: $f(x_k) = 0.5248$



Ricapitolando

$f(x_k)$	Δ_k
11.352400	0.300000
5.078800	0.300000
2.204800	0.300000
0.524800	0.300000
0.524800	0.150000
0.006925	0.150000
0.006925	0.075000
0.006925	0.037500
0.006925	0.018750
0.000298	0.018750
0.000298	0.009375
0.000298	0.004687
0.000298	0.002344
0.000173	0.002344
0.000054	0.002344
0.000043	0.002344
0.000033	0.002344

Pseudo-codice del “compass search” forte

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

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

while $k \leq$ 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-codice del "compass search" forte

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

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

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

$\quad k \leftarrow k + 1$

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

$\quad$ **if** $f(x_k + \Delta_k \bar{d}) < f(x_k)$ **then**

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

$\quad$ **else**

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

$\quad$ **endif**

$\quad k \leftarrow k + 1$

end while

RETURN: x (miglior punto determinato)$\{x_k\}, \{\Delta_k\}$ successioni di punti e passi

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. nei (due) metodi visti, il passo Δ_k, iterazione dopo iterazione,

- o diminuisce (di un fattore costante $\Delta_{k+1} \leftarrow \Delta_k/2$)
- o rimane costante ($\Delta_{k+1} \leftarrow \Delta_k$)

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

Lemma

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

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

Dim.: Se $\Delta_{min} \leq 0$ e $\texttt{maxit} = +\infty$, 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 $\bar{\Delta} > 0$.

Convergenza a zero del passo (segue)

Esiste un intero $\bar{k}$ tale che, per $k \geq \bar{k}$ il passo non viene più ridotto. Quindi, per ogni $k \geq \bar{k}$

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

Quindi, per ogni $k \geq \bar{k}$, i punti x_k appartengono ad una griglia, ovvero, per $i > 0$

$$x_{\bar{k}+i} = x_{\bar{k}} + \bar{\Delta} \sum_{r=0}^{i-1} d_{i_r} \in L(x_0) \quad d_{i_r} \in D$$

Questo e (A1) implicano che la sottosuccessione $\{x_k\}_{k \geq \bar{k}}$ (e quindi $\{x_k\}$) è costituita da un numero **finito** di punti distinti.

Convergenza a zero del passo (segue)

Ma allora, deve necessariamente esistere un intero $\tilde{k} > \bar{k}$ t.c.

$$x_{\tilde{k}+1} = x_{\tilde{k}},$$

ma questo contraddice il fatto che $f(x_{\tilde{k}+1}) < f(x_{\tilde{k}})$ e allora, $\bar{\Delta} = 0$.

Un po' di analisi (matematica)

Teorema (della media o di Lagrange in $\mathbb{R}$)

$f(x) : \mathbb{R} \to \mathbb{R}$ continua e derivabile in $[a, b]$. Esiste sempre un $\xi \in (a, b)$ tale che

$$f(b) - f(a) = f'(\xi)(b - a)$$

Definizione (Segmento in $\mathbb{R}^n$)

Dati due punti $a, b \in \mathbb{R}^n$, il segmento di estremi a e b è

$$\begin{aligned} [a, b] &\triangleq \{x \in \mathbb{R}^n : x = ta + (1 - t)b,\ \ t \in [0, 1]\} \\ (a, b) &\triangleq \{x \in \mathbb{R}^n : x = ta + (1 - t)b,\ \ t \in (0, 1)\} \end{aligned}$$

Teorema (della media o di Lagrange in $\mathbb{R}^n$)

$f(x) : \mathbb{R}^n \to \mathbb{R}$ continua e derivabile in $[a, b]$. Esiste sempre un $\xi \in (a, b)$ tale che

$$f(b) - f(a) = \nabla f(\xi)^\top (b - a)$$