Ottimizzazione Non Lineare

G. Liuzzi[1]

Mercoledì 14 Novembre 2018

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

Pseudo-code del metodo "Fermi-Metropolis"

INPUT: x_0, Δ_0, Δ_{min}, $\texttt{maxit}$

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

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

  $k \leftarrow k+1$, $\tilde{x} \leftarrow x$

  for $i = 1, 2, \ldots, n$

    if $f(\tilde{x} + \Delta e_i) < f(\tilde{x})$ **then**

      while $f(\tilde{x} + \Delta e_i) < f(\tilde{x})$ **do** $\tilde{x} \leftarrow \tilde{x} + \Delta e_i$ **end while**

    else if $f(\tilde{x} - \Delta e_i) < f(\tilde{x})$ **then**

      while $f(\tilde{x} - \Delta e_i) < f(\tilde{x})$ **do** $\tilde{x} \leftarrow \tilde{x} - \Delta e_i$ **end while**

    end if

  end for

  if $f(\tilde{x}) = f(x)$ **then** $\Delta \leftarrow \Delta/2$

  else $x \leftarrow \tilde{x}$

  end if

end while

RETURN: x (miglior punto determinato)

Compass Search "debole"

INPUT: x_0, Δ_0, Δ_{min}, `maxit`

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

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

  $k \leftarrow k + 1$, $\tilde{x} \leftarrow x$

  for $i = 1, 2, \ldots, n$

    if $f(\tilde{x} + \Delta e_i) < f(\tilde{x})$ **then**

      $\tilde{x} \leftarrow \tilde{x} + \Delta e_i$, **break**

    else if $f(\tilde{x} - \Delta e_i) < f(\tilde{x})$ **then**

      $\tilde{x} \leftarrow \tilde{x} - \Delta e_i$, **break**

    end if

  end for

  if $f(\tilde{x}) = f(x)$ **then** $\Delta \leftarrow \Delta/2$

  else $x \leftarrow \tilde{x}$

  end if

end while

RETURN: x (miglior punto determinato)

Compass Search "debole"

INPUT: x_0, Δ_0, Δ_{min}, $\mathtt{maxit}$, $D = \{\pm e_1, \ldots, \pm e_n\}$

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

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

$k \leftarrow k + 1$

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

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

else

$\Delta \leftarrow \Delta/2$, $\tilde{x} \leftarrow x$

endif

$x \leftarrow \tilde{x}$

end while

RETURN: x (miglior punto determinato)

Compass Search "debole"

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**

$\quad k \leftarrow k + 1$

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

$\quad\quad \tilde{x}_k \leftarrow x_k + \Delta_k \bar{d}$, $\Delta_{k+1} \leftarrow \Delta_k$

$\quad$ **else**

$\quad\quad \Delta_{k+1} \leftarrow \Delta_k/2$, $\tilde{x}_k \leftarrow x_k$

$\quad$ **endif**

$\quad x_{k+1} \leftarrow \tilde{x}_k$, $k \leftarrow k + 1$

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. 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 $\{x_k\}_{k \geq \bar{k}}$ (e $\{x_k\}$) è composta 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.

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

e quest contraddice il fatto che $f(x_{\tilde{k}+1}) < f(x_{\tilde{k}})$ e

allora, $\bar{\Delta} = 0$.

Convergenza a zero del passo (segue)

N.B. la convergenza a zero del passo Δ_k segue da:

- compattezza di $L(x_0)$
- struttura dei punti generati ovvero che per Δ_k fisso i punti appartengono ad una ben precisa griglia

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*

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

Convergenza a punti stazionari

Sotto l'ipotesi (A1), faremo vedere che

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

Seguiremo due strade.

1. Assumendo che ∇f sia continuo;
2. Assumendo che ∇f sia Lipschitz continuo con costante L.

Convergenza a punti stazionari

Teorema

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

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

(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). Percui, esiste $K_2 \subseteq K_1$ s.t.

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

Convergenza a punti stazionari (segue)

Per ogni $k \in K_2$ e $i = 1, \dots, n$, 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, $\nabla f(\bar{x}) = 0$.

Lipschitz Continuità di ∇f

Definizione

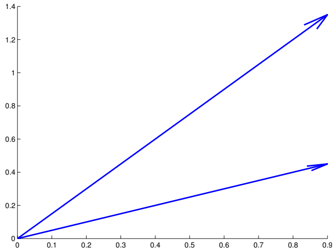
∇f è Lipschitz continuo (con costante L) su un insieme $A \subseteq \mathbb{R}^n$ quando, comunque presi due punti $x, y \in A$, risulta

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|.$$

Angolo tra due vettori

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

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

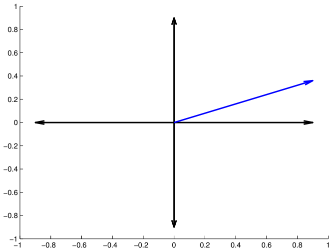


Angolo vettore – insieme di vettori

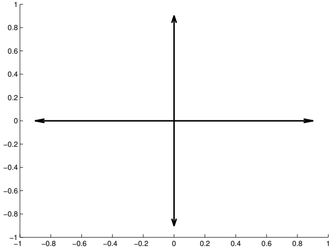
Dati:

- 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)

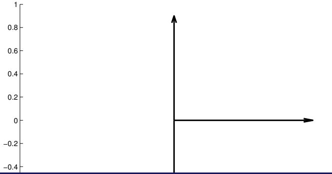


$$\phi(v) = \max_{d \in D} \frac{v^\top d}{\|v\|\|d\|} = \max_{d \in D} \cos\theta(d, v)$$

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

Dato un insieme finito di vettori (direzioni) D

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



Una proprietà importante

Limitiamoci a considerare $D = \{\pm e_1, \ldots, \pm e_n\}$. Allora

$$\begin{aligned}\kappa(D) &= \min_{v\in\mathbb{R}^n}\max_{d\in D}\frac{v^\top d}{\|v\|\|d\|} = \min_{v\in\mathbb{R}^n}\max_{i=1,\ldots,n}\frac{|v^\top e_i|}{\|v\|} \\ &= \min_{v\in\mathbb{R}^n,\|v\|=1}\max_{i=1,\ldots,n}|v^\top e_i|\end{aligned}$$

Lemma

$\kappa(D) = 1/\sqrt{n}$

dim. Per ogni $v \in \mathbb{R}^n$, $|v^\top e_i| = |v_i|$. Se $\|v\|^2 = \sum_{i=1}^n |v_i|^2 = 1$, allora $|v^\top e_j| \geq 1/\sqrt{n}$ per qualche j. Quindi, $\kappa(D) \geq 1/\sqrt{n}$.
Se consideriamo $v = \alpha_1 e_1 + \cdots + \alpha_n e_n$ con $\alpha_i = \pm 1/\sqrt{n}$, $\kappa(D) = 1/\sqrt{n}$. $\square$

Convergenza a punti stazionari

Teorema

Se vale (A1) e se $\nabla f(x)$ è Lipschitz continuo modulo L, 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). Percui, esiste $K_2 \subseteq K_1$ s.t.

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

Convergenza a punti stazionari (segue)

Dal fatto che $\kappa(D) = 1/\sqrt{n}$, abbiamo che, per ogni $k \in K_2$, esiste almeno una direzione $d \in D$ tale che

$$\frac{1}{\sqrt{n}}\|\nabla f(x_k)\| \leq -\nabla f(x_k)^\top d \tag{1}$$

inoltre, siccome l'iterazione $k \in K_2$ è di fallimento:

$$0 \leq f(x_k + \Delta_k d) - f(x_k) = \nabla f(x_k + \beta_k \Delta_k d)^\top \Delta_k d$$

con $\beta_k \in (0, 1)$. Cioè, sommando $-\Delta_k \nabla f(x_k)^\top d$ ad ambo i membri,

$$-\Delta_k \nabla f(x_k)^\top d \leq \Delta_k \left(\nabla f(x_k + \beta_k \Delta_k d) - \nabla f(x_k)\right)^\top d$$

ovvero, per la (1),

$$\frac{1}{\sqrt{n}}\|\nabla f(x_k)\| \leq \left(\nabla f(x_k + \beta_k \Delta_k d) - \nabla f(x_k)\right)^\top d.$$

Convergenza a punti stazionari (segue)

Ora, usando la Lipschitz continuità di ∇f

$$\frac{1}{\sqrt{n}}\|\nabla f(x_k)\| \leq L\|\beta_k \Delta_k d\|\|d\| \leq L\Delta_k,$$

quindi

$$\|\nabla f(x_k)\| \leq \sqrt{n} L\Delta_k.$$

Il risultato segue ricordando che $\lim_{k\to\infty, k\in K_2} \Delta_k = 0$ $\square$

Variante di Compass Search "debole"

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

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

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

$\quad k \leftarrow k + 1$

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

$\quad\quad \tilde{x}_k \leftarrow x_k + \Delta_k \bar{d}$, $\Delta_{k+1} \leftarrow \Delta_k$

$\quad$ **else**

$\quad\quad \Delta_{k+1} \leftarrow \Delta_k / 2$, $\tilde{x}_k \leftarrow x_k$

$\quad$ **endif**

$\quad x_{k+1} \leftarrow \tilde{x}_k$ Find x_{k+1} s.t. $f(x_{k+1}) \leq f(\tilde{x}_k)$, $k \leftarrow k + 1$

end while

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

ATTENZIONE, quando Δ_k costante, i punti non appartengono più ad una griglia !!

Variante di Compass Search "debole"

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

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

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

  $k \leftarrow k + 1$

  if $\exists \bar{d} \in D$ s.t. $f(x_k + \Delta_k \bar{d}) < f(x_k) - \gamma \Delta_k^2$ **then**

    $\tilde{x}_k \leftarrow x_k + \Delta_k \bar{d}$, $\Delta_{k+1} \leftarrow \Delta_k$

  else

    $\Delta_{k+1} \leftarrow \Delta_k / 2$, $\tilde{x}_k \leftarrow x_k$

  endif

  Find x_{k+1} s.t. $f(x_{k+1}) \leq f(\tilde{x}_k)$, $k \leftarrow k + 1$

end while

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