Ottimizzazione dei Sistemi Complessi

G. Liuzzi[1]

Venerdì 9 Marzo 2018

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

Pseudo-codice del "compass search"

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**

  $k \leftarrow k + 1$

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

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

  else

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

  endif

  $k \leftarrow k + 1$

end while

RETURN: x (miglior punto determinato)

Pseudo-codice del "compass search"

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

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

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

  $k \leftarrow k + 1$

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

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

  else

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

  endif

  $k \leftarrow k + 1$

end while

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

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**

  $k \leftarrow k + 1$

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

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

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

  else

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

  endif

  $k \leftarrow k + 1$

end while

RETURN: x (miglior punto determinato)

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** $\Delta_k \geq \Delta_{min}$ **do**

    $k \leftarrow k + 1$

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

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

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

    else

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

    endif

    $k \leftarrow k + 1$

end while

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

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*

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

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*

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

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

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

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

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

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

Per i metodi visti 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). Percui, esiste $K_2 \subseteq K_1$ 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

Per i metodi visti 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). 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

Per i metodi visti 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). 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 $\displaystyle\lim_{k\to\infty, k\in K_2} u_{k,i} = \lim_{k\to\infty, k\in K_2} v_{k,i} = \bar{x}$.

Convergenza a punti stazionari (segue)

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)

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)

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

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

e, per il teorema della media,

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

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 $\lim\limits_{k\to\infty, k\in K_2} u_{k,i} = \lim\limits_{k\to\infty, k\in K_2} v_{k,i} = \bar{x}$.

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, $\nabla f(\bar{x}) = 0$.

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, $\nabla f(\bar{x}) = 0$.

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

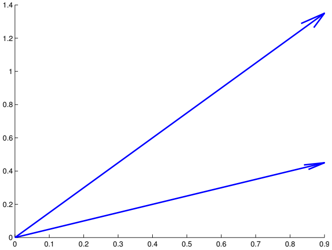
∇f è Lipschitz continuo (con costante L) su un insieme $A \subseteq \mathbb{R}^n$ quando, comunque presi due punti $x, y \in A$, risulta

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

Angolo tra due vettori

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

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

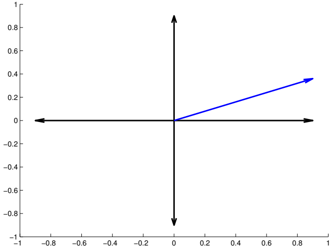


Angolo vettore – insieme di vettori

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)

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

Dato un insieme finito di vettori (direzioni) D

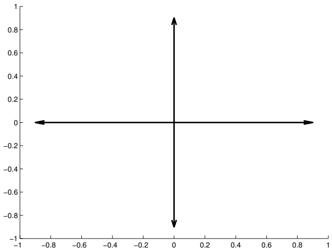
$$\kappa(D) = \min_{v \in \mathbb{R}^n} \phi(v) = \min_{v \in \mathbb{R}^n} \max_{d \in D} \frac{v^\top d}{\|v\|\|d\|}$$

Misura del coseno (di un ins. di vettori)

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

Dato un insieme finito di vettori (direzioni) D

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



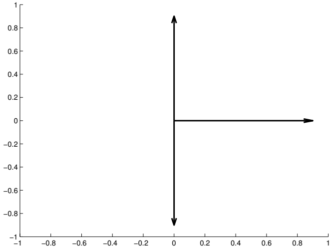
$$\kappa(D) = \frac{1}{\sqrt{2}}$$

Misura del coseno (di un ins. di vettori)

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

Dato un insieme finito di vettori (direzioni) D

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



$$\kappa(D) = 0$$

Una proprietà importante

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

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

$$= \min_{v\in\mathbb{R}^n, \|v\|=1} \max_{i=1,\ldots,n} |v^\top e_i|$$

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 $\alpha_i = \pm 1/\sqrt{n}$, $\kappa(D) = 1/\sqrt{n}$. $\square$

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 $\alpha_i = \pm 1/\sqrt{n}$, $\kappa(D) = 1/\sqrt{n}$. □

Una proprietà importante

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

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

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 $\alpha_i = \pm 1/\sqrt{n}$, $\kappa(D) = 1/\sqrt{n}$. $\square$

Una proprietà importante

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

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

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 $\alpha_i = \pm 1/\sqrt{n}$, $\kappa(D) = 1/\sqrt{n}$. $\square$

Una proprietà importante

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

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

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 $\alpha_i = \pm 1/\sqrt{n}$, $\kappa(D) = 1/\sqrt{n}$. $\square$

Una proprietà importante

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

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

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 $\alpha_i = \pm 1/\sqrt{n}$, $\kappa(D) = 1/\sqrt{n}$. □

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 $\alpha_i = \pm 1/\sqrt{n}$, $\kappa(D) = 1/\sqrt{n}$. $\square$

Una proprietà importante

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

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

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

Per i metodi visti vale il risultato

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.

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

Convergenza a punti stazionari

Per i metodi visti vale il risultato

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.

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

Convergenza a punti stazionari

Per i metodi visti vale il risultato

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.

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

Convergenza a punti stazionari (segue)

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)

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)

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

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

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)

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

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

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)

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

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

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 $\lim_{k\to\infty, k\in K_2} \Delta_k = 0$ $\square$

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 $\lim_{k\to\infty, k\in K_2} \Delta_k = 0$ $\square$

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 $\lim_{k\to\infty, k\in K_2} \Delta_k = 0$ $\square$

Introduzione

Definizione (Stencil)

Dato $x \in \mathbb{R}^n$ *ed un passo* $h > 0$*, uno* stencil *è l'insieme (finito) di punti*

$$S(x,h) = \{x + he_1, \ldots, x + he_n, x - he_1, \ldots, x - he_n\}$$

Definizione (Gradiente approssimato)

Dati $x \in \mathbb{R}^n$*,* $h > 0$ *e* $S(x,h)$*,*

$$\nabla_h f(x) = \left(\frac{\partial_h f(x)}{\partial x_1}, \ldots, \frac{\partial_h f(x)}{\partial x_n}\right)^\top$$

con

$$\frac{\partial_h f(x)}{\partial x_i} = \frac{f(x + he_i) - f(x - he_i)}{2h}$$

Introduzione

Definizione (Stencil)

Dato $x \in \mathbb{R}^n$ *ed un passo* $h > 0$, *uno* stencil *è l'insieme (finito) di punti*

$$S(x,h) = \{x + he_1, \ldots, x + he_n, x - he_1, \ldots, x - he_n\}$$

Definizione (Gradiente approssimato)

Dati $x \in \mathbb{R}^n$, $h > 0$ *e* $S(x,h)$,

$$\nabla_h f(x) = \left(\frac{\partial_h f(x)}{\partial x_1}, \ldots, \frac{\partial_h f(x)}{\partial x_n} \right)^\top$$

con

$$\frac{\partial_h f(x)}{\partial x_i} = \frac{f(x + he_i) - f(x - he_i)}{2h}$$

Proprietà di $\nabla_h f(x)$

Lemma

Sia f continuamente differenziabile e siano $\{z_k\} \subset \mathbb{R}^n$, $\{h_k\} \subset \mathbb{R}_+$ due succ. infinite tali che

$$\lim_{k\to\infty} z_k = x \qquad \lim_{k\to\infty} h_k = 0.$$

Allora

$$\lim_{k\to\infty} \nabla_{h_k} f(z_k) = \nabla f(x)$$

Proprietà di $\nabla_h f(x)$ (segue)

Dim. Per ogni $i \in \{1, \ldots, n\}$ risulta

$$\frac{\partial_h f(z_k)}{\partial x_i} = \frac{f(z_k + h_k e_i) - f(z_k - h_k e_i)}{2h_k}$$

e dal Teorema della media

$$\frac{\partial_h f(z_k)}{\partial x_i} = \frac{\partial f(\xi_k^i)}{\partial x_i}$$

dove $\xi_k^i = z_k + \theta_k^i h_k e_i$, con $\theta_k^i \in (-1, 1)$. Prendendo il limite per $k \to \infty$ e notando che $\lim_{k\to\infty} \xi_k^i = x$, otteniamo

$$\lim_{k\to\infty} \frac{\partial_h f(z_k)}{\partial x_i} = \frac{\partial f(x)}{\partial x_i}$$

che prova il risultato data la generalità di i. $\square$

Proprietà di $\nabla_h f(x)$ (segue)

Dim. Per ogni $i \in \{1, \dots, n\}$ risulta

$$\frac{\partial_h f(z_k)}{\partial x_i} = \frac{f(z_k + h_k e_i) - f(z_k - h_k e_i)}{2h_k}$$

e dal Teorema della media

$$\frac{\partial_h f(z_k)}{\partial x_i} = \frac{\partial f(\xi_k^i)}{\partial x_i}$$

dove $\xi_k^i = z_k + \theta_k^i h_k e_i$, con $\theta_k^i \in (-1, 1)$. Prendendo il limite per $k \to \infty$ e notando che $\lim_{k\to\infty} \xi_k^i = x$, otteniamo

$$\lim_{k\to\infty} \frac{\partial_h f(z_k)}{\partial x_i} = \frac{\partial f(x)}{\partial x_i}$$

che prova il risultato data la generalità di i. $\square$

Proprietà di $\nabla_h f(x)$ (segue)

Dim. Per ogni $i \in \{1, \dots, n\}$ risulta

$$\frac{\partial_h f(z_k)}{\partial x_i} = \frac{f(z_k + h_k e_i) - f(z_k - h_k e_i)}{2h_k}$$

e dal Teorema della media

$$\frac{\partial_h f(z_k)}{\partial x_i} = \frac{\partial f(\xi_k^i)}{\partial x_i}$$

dove $\xi_k^i = z_k + \theta_k^i h_k e_i$, con $\theta_k^i \in (-1, 1)$. Prendendo il limite per $k \to \infty$ e notando che $\lim_{k\to\infty} \xi_k^i = x$, otteniamo

$$\lim_{k\to\infty} \frac{\partial_h f(z_k)}{\partial x_i} = \frac{\partial f(x)}{\partial x_i}$$

che prova il risultato data la generalità di i. $\square$

Osservazione

Nelle due varianti di "compass search" che abbiamo visto, in quale situazione si aggiorna il passo?

Quando **non esiste** $y \in S(x_k, \Delta_k)$ t.c. $f(y) < f(x_k)$ (**stencil failure**)

In occasione di uno *stencil failure*, abbiamo tutto quello che serve per costruire $\nabla_h f(x_k)$ con $h = \Delta_k$

Il metodo **implicit filtering** tenta di usare $d_k = -\nabla_h f(x_k)$ per compiere uno spostamento prima di aggiornare il passo

Osservazione

Nelle due varianti di "compass search" che abbiamo visto, in quale situazione si aggiorna il passo?

Quando **non esiste** $y \in S(x_k, \Delta_k)$ t.c. $f(y) < f(x_k)$ (**stencil failure**)

In occasione di uno *stencil failure*, abbiamo tutto quello che serve per costruire $\nabla_h f(x_k)$ con $h = \Delta_k$

Il metodo **implicit filtering** tenta di usare $d_k = -\nabla_h f(x_k)$ per compiere uno spostamento prima di aggiornare il passo

Osservazione

Nelle due varianti di "compass search" che abbiamo visto, in quale situazione si aggiorna il passo?

Quando **non esiste** $y \in S(x_k, \Delta_k)$ t.c. $f(y) < f(x_k)$ (**stencil failure**)

In occasione di uno *stencil failure*, abbiamo tutto quello che serve per costruire $\nabla_h f(x_k)$ con $h = \Delta_k$

Il metodo **implicit filtering** tenta di usare $d_k = -\nabla_h f(x_k)$ per compiere uno spostamento prima di aggiornare il passo

Osservazione

Nelle due varianti di "compass search" che abbiamo visto, in quale situazione si aggiorna il passo?

Quando **non esiste** $y \in S(x_k, \Delta_k)$ t.c. $f(y) < f(x_k)$ (**stencil failure**)

In occasione di uno *stencil failure*, abbiamo tutto quello che serve per costruire $\nabla_h f(x_k)$ con $h = \Delta_k$

Il metodo **implicit filtering** tenta di usare $d_k = -\nabla_h f(x_k)$ per compiere uno spostamento prima di aggiornare il passo

pseudo-codice "compass search"

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

$k \leftarrow 0$

while $k \leq$ 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

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

  endif

  $k \leftarrow k + 1$

end while

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

pseudo-codice "compass search"

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

$k \leftarrow 0$

while $k \leq \texttt{maxit}$ **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

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

endif

$k \leftarrow k + 1$

end while

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

pseudo-codice "compass search"

$\vdots$

else

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

endif

$\vdots$

pseudo-codice "compass search"

$\vdots$

else

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

endif

$\vdots$

pseudo-codice "implicit filtering"

⋮

else

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

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

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

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

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

endif

⋮

pseudo-codice "implicit filtering"

$\vdots$

else

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

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

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

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

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

endif

$\vdots$

pseudo-codice "implicit filtering"

$\vdots$

else

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

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

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

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

    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

endif

$\vdots$

Analisi di Convergenza

Quale proprietà si dimostra per prima ?

Assunzione (A1)

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

Lemma

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

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

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 $\texttt{maxit} = +\infty$, 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 $\bar{\Delta} > 0$.
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)

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 $\bar{\Delta} > 0$.
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)

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.

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

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.

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

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.

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

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.

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

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)

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)

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)

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)

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. $\square$

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. $\square$