

Ottimizzazione Non Lineare

G. Liuzzi¹

Lunedì 10 Dicembre 2018

¹Istituto di Analisi dei Sistemi ed Informatica IASI - CNR



Definizione del problema

$$\begin{aligned} \min f(x) \\ \text{s.t. } x \in \mathbb{R}^n S \end{aligned}$$

dove:

- $f : \mathbb{R}^n \rightarrow \mathbb{R}$;
- $S \subset \mathbb{R}^n$, convesso



Difinizioni

Definizione (Direzione di discesa)

Sia $x \in S$ e $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Si dice che $d \in \mathbb{R}^n$, $d \neq 0$, è una direzione di discesa per f in x se esiste un $\bar{t} > 0$ tale che

$$f(x + td) < f(x), \quad \text{per ogni } t \in (0, \bar{t}]$$

Definizione (Direzione ammissibile)

Sia $x \in S$. Si dice che $d \in \mathbb{R}^n$, $d \neq 0$, è una direzione ammissibile per S in x se esiste un $\bar{t} > 0$ tale che

$$x + td \in S, \quad \text{per ogni } t \in [0, \bar{t}]$$



C.N. di ottimo locale

Consideriamo un punto ammissibile $\bar{x} \in S$ e supponiamo che in \bar{x} esista una direzione $d \in \mathbb{R}^n$, $d \neq 0$ che sia contemporaneamente **ammissibile** e **di discesa**.

Allora, in ogni intorno di \bar{x} è possibile trovare, per $t > 0$ suff. piccolo, un punto $\bar{x} + td$ tale che

$$\bar{x} + td \in S$$

$$f(\bar{x} + td) < f(\bar{x})$$

Quindi \bar{x} **non può essere** minimo locale del problema



C.N. di ottimo locale (1)

Rimane quindi provato

Proposizione

Sia x^ un punto di minimo locale del problema, allora non può esistere in x^* una direzione ammissibile (per S) che sia anche di discesa (per f).*

Se f è differenziabile, la condizione

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

è sufficiente affinché d sia di discesa per f in x . Quindi

Proposizione

Sia x^ un punto di minimo locale del problema e f cont. differenziabile in un intorno di x^* .*

Allora non può esistere in x^ una direzione ammissibile tale che $\nabla f(x^*)^\top d < 0$, o equivalentemente,*

$$\nabla f(x^*)^\top d \geq 0, \quad \text{per ogni } d \text{ ammissibile in } x^*$$



Caratterizzazione delle dir. ammissibili

Supponiamo di conoscere un punto $\bar{x} \in S$ e che $S \neq \{\bar{x}\}$.

Comunque scelto un punto $x \in S$, $x \neq \bar{x}$, per la convessità di S abbiamo che

$$(1 - t)\bar{x} + tx \in S, \quad \text{per ogni } t \in [0, 1]$$

e quindi,

$$\bar{x} + t(x - \bar{x}) \in S, \quad \text{per ogni } t \in [0, 1],$$

ovvero $d = x - \bar{x} \neq 0$ è ammissibile per S in \bar{x} .

Inversamente, se $d \neq 0$ è ammissibile per S in \bar{x} , esiste un $t > 0$ tale che

$$\bar{x} + td = x \in S$$

ovvero $d = (x - \bar{x})/t = \lambda(x - \bar{x})$



Caratterizzazione delle dir. ammissibili

Proposizione

Sia $\bar{x} \in S$ e $S \neq \{\bar{x}\}$. Allora, per ogni $x \in S$, $x \neq \bar{x}$, la direzione $d = x - \bar{x}$ è ammissibile per S in \bar{x} , e viceversa



C.N. di ottimo locale (2)

Proposizione

Sia x^ un punto di minimo locale del problema e f cont. differenziabile in un intorno di x^* .*

Allora,

$$\nabla f(x^*)^\top (x - x^*) \geq 0, \quad \text{per ogni } x \in S$$

Dim. Se $S = \{x^*\}$, la condizione è banalmente soddisfatta.

Supponiamo quindi che $S \neq \{x^*\}$ cioè che esista $x \in S$, $x \neq x^*$.

Qualunque sia x , sappiamo che $d = x - x^*$ è ammissibile per S in x^* e che $\nabla f(x^*)^\top d \geq 0$. □



Ricerca lungo una direzione ammissibile

Supponiamo che:

- 1 $\bar{x} \in S$
- 2 $\bar{x} + \bar{d} \in S$, quindi \bar{d} è ammissibile per S in \bar{x}
- 3 $\nabla f(\bar{x})^\top \bar{d} < 0$, quindi \bar{d} è di discesa per f in \bar{x}

ricerca unidimensionale di Armijo con passo iniziale $\alpha = 1$

Dati: $\gamma \in (0, 1)$, $\delta \in (0, 1)$

Poni $\alpha \leftarrow 1$

while $f(\bar{x} + \alpha \bar{d}) > f(\bar{x}) + \gamma \alpha \nabla f(\bar{x})^\top \bar{d}$

 Poni $\alpha \leftarrow \delta \alpha$

end while

Poni $\bar{\alpha} \leftarrow \alpha$



La ricerca unidimensionale di Armijo è ben posta

Si nota facilmente che (sotto le condizioni poste) la ricerca di Armijo si arresta sempre in un numero finito di passi determinando un $\bar{\alpha} \in (0, 1]$ tale che

- $\bar{x} + \bar{\alpha}\bar{d} \in S$
- $f(\bar{x} + \bar{\alpha}\bar{d}) \leq f(\bar{x}) + \gamma\bar{\alpha}\nabla f(\bar{x})^\top \bar{d} < f(\bar{x})$



Il metodo di Armijo

Sia $\{d_k\}$ una successione infinita di direzioni tale che:

- 1 $x_k + d_k \in S$, quindi d_k è ammissibile per S in x_k
- 2 $\nabla f(x_k)^\top d_k < 0$, quindi d_k è di discesa per f in x_k

Dati: $x_0 \in S$, $\gamma \in (0, 1)$, $\delta \in (0, 1)$

for $k = 0, 1, \dots$

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

con α_k calcolato dalla ricerca unidimensionale di Armijo

end for



Convergenza del metodo di Armijo

Proposizione

Sia f continuamente differenziabile su un aperto contenente S . Supponiamo S compatto (oltre che convesso). Il metodo di Armijo produce una successione infinita di punti $\{x_k\}$. Se la successione $\{d_k\}$ è tale che, per ogni k ,

$$\text{ii) } x_k + d_k \in S$$

$$\text{iii) } \nabla f(x_k)^\top d_k < 0$$

allora,

$$\text{① } x_{k+1} \in S$$

$$\text{② } f(x_{k+1}) < f(x_k)$$

$$\text{③ } \lim_{k \rightarrow \infty} \nabla f(x_k)^\top d_k = 0$$



Convergenza del metodo di Armijo (segue)

Dim. La dimostrazione di (1) e (2) è banale. Quindi rimane da far vedere che vale la (3).

Osserviamo che, essendo $x_k \in S$ e $x_k + d_k \in S$, la compattezza di S implica che $\{\|d_k\|\}$ è limitata, ossia esiste $M > 0$ tale che $\|d_k\| \leq M$. Infatti, $\|d_k\| = \|d_k + x_k - x_k\| \leq \|x_k + d_k\| + \|x_k\| \leq M$
Dalle proprietà di α_k , possiamo scrivere

$$f(x_k) - f(x_{k+1}) \geq \gamma \alpha_k |\nabla f(x_k)^\top d_k|$$

e quindi

$$\lim_{k \rightarrow \infty} \alpha_k |\nabla f(x_k)^\top d_k| = 0.$$

Procediamo per assurdo e supponiamo che la (3) non sia vera. Allora deve esistere una sottosucc. e un numero $\eta > 0$ tale che

$$\lim_{k \rightarrow \infty} \nabla f(x_k)^\top d_k = -\eta < 0 \quad \text{e quindi} \quad \lim_{k \rightarrow \infty} \alpha_k = 0.$$



Convergenza del metodo di Armijo (segue)

Inoltre, devono esistere sottosuccessioni tali che

$$\lim_{k \rightarrow \infty} x_k = \hat{x} \in S \quad \text{e} \quad \lim_{k \rightarrow \infty} d_k = \hat{d}$$

da cui segue

$$\lim_{k \rightarrow \infty} \nabla f(x_k)^\top d_k = \nabla f(\hat{x})^\top \hat{d} = -\eta < 0$$

Siccome $\alpha_k \rightarrow 0$, per k suff. grande deve risultare $\alpha_k < 1$ e quindi

$$f\left(x_k + \frac{\alpha_k}{\delta} d_k\right) > f(x_k) + \gamma \frac{\alpha_k}{\delta} \nabla f(x_k)^\top d_k$$

per il Teorema della Media abbiamo

$$f\left(x_k + \frac{\alpha_k}{\delta} d_k\right) = f(x_k) + \frac{\alpha_k}{\delta} \nabla f(z_k)^\top d_k, \quad z_k = x_k + \theta_k \frac{\alpha_k}{\delta} d_k, \quad \theta_k \in (0, 1)$$

per cui otteniamo

$$\nabla f(z_k)^\top d_k > \gamma \nabla f(x_k)^\top d_k$$



Convergenza del metodo di Armijo (segue)

Prendendo il limite per $k \rightarrow \infty$, e ricordando che $\lim_{k \rightarrow \infty} \alpha_k \|d_k\| = 0$ per cui $\lim_{k \rightarrow \infty} z_k = \hat{x}$, otteniamo

$$\nabla f(\hat{x})^\top \hat{d} \geq \gamma \nabla f(\hat{x})^\top \hat{d}$$

ovvero $\eta \leq \gamma \eta$, da cui seguirebbe $\gamma \geq 1$. □

Per stabilire la convergenza dell'algoritmo a punti critici del problema (cioè punti che soddisfano C.N. di ottimo), è sufficiente mostrare che d_k sia scelta in modo tale che

- ① sia una direzione ammissibile
- ② sia una direzione di discesa
- ③ $\lim_{k \rightarrow \infty} \nabla f(x_k)^\top d_k = 0$ implichi la convergenza a punti critici.



Introduzione

Ricordiamo che S è un insieme convesso e compatto.

Dato $x_k \in S$, un metodo per cercare una direzione che sia contemporaneamente di discesa ed ammissibile è quello di risolvere il problema

$$\begin{aligned} \min \quad & \nabla f(x_k)^\top (x - x_k) \\ \text{s.t.} \quad & x \in S \end{aligned}$$

Il problema ammette sempre soluzione ottima $\hat{x}_k \in S$ e risulta $\nabla f(x_k)^\top (\hat{x}_k - x_k) \leq 0$.

Se $\nabla f(x_k)^\top (\hat{x}_k - x_k) = 0$ allora

$$0 = \nabla f(x_k)^\top (\hat{x}_k - x_k) \leq \nabla f(x_k)^\top (x - x_k), \quad \text{per ogni } x \in S$$

cioè x_k è un punto critico.

Se invece $\nabla f(x_k)^\top (\hat{x}_k - x_k) < 0$, la direzione $d_k = \hat{x}_k - x_k$ è una direzione ammissibile e di discesa in x_k e possiamo definire l'iterazione

$$x_{k+1} = x_k + \alpha_k d_k$$



Metodo di Frank-Wolfe

INPUT: $x_0 \in S$, $\gamma \in (0, 1)$, $\delta \in (0, 1)$

for $k = 0, 1, \dots$

Calcola una soluzione \hat{x}_k di $\min_{x \in S} \nabla f(x_k)^\top (x - x_k)$

if $\nabla f(x_k)^\top (\hat{x}_k - x_k) = 0$ **then** STOP

Poni $d_k = \hat{x}_k - x_k$ e calcola α_k con una ricerca di linea di Armijo

Poni $x_{k+1} = x_k + \alpha_k d_k$

end for



Convergenza

Proposizione

Sia f cont. differenziabile su un aperto contenente S e $\{x_k\}$ la succ. prodotta dal metodo di F-W. Allora,

- *o esiste una iterazione ν tale che x_ν è un punto critico;*
- *oppure $\{x_k\}$ è infinita ed ogni punto di accumulazione è critico.*

Dim. Supponiamo che l'algoritmo non si arresti mai. Allora, risultano verificate le ipotesi del metodo di Armijo, per cui

$$\lim_{k \rightarrow \infty} \nabla f(x_k)^\top d_k = 0.$$

Per la compattezza di S deve esistere un punto di accumulazione \bar{x} , inoltre d_k è limitata.

Possiamo quindi definire una sottosucc. K tale che

$$\lim_{k \rightarrow \infty, k \in K} x_k = \bar{x}, \quad \lim_{k \rightarrow \infty, k \in K} d_k = \bar{d}$$



Convergenza (segue)

da cui segue

$$\nabla f(\bar{x})^\top \bar{d} = 0.$$

In base alla definizione di d_k segue che

$$\nabla f(x_k)^\top d_k \leq \nabla f(x_k)^\top (x - x_k), \quad \text{per ogni } x \in S.$$

Quindi, prendendo il limite $k \rightarrow \infty, k \in K$

$$0 = \nabla f(\bar{x})^\top \bar{d} \leq \nabla f(\bar{x})^\top (x - \bar{x}), \quad \text{per ogni } x \in S$$

che prova che \bar{x} è un punto critico.

