

# Ottimizzazione dei Sistemi Complessi

G. Liuzzi<sup>1</sup>

Venerdì 16 Novembre 2018

---

<sup>1</sup>Istituto di Analisi dei Sistemi ed Informatica IASI - CNR

# Vi ricordo che ...

... abbiamo considerato il problema:

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

- problema senza vincoli
- $f(x)$  continuamente differenziabile ma
- $\nabla f(x)$  non disponibile (o troppo costoso)
- ricerca di un minimo locale o punto stazionario

# Perché assumiamo che $f(x)$ sia continua?

Perché se così non fosse, **non avremmo garanzia di convergenza** a punti stazionari

Esempio:

$$f(x) = \begin{cases} (x - 1)^2 & \text{se } x \neq 0 \\ 5 & \text{altrimenti} \end{cases}$$

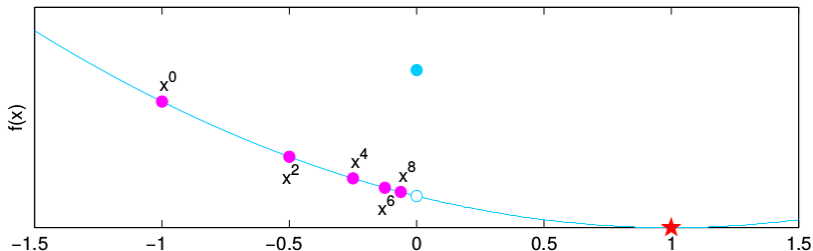
Supponiamo che un metodo “compass search” parta dal punto  $x_0 = -1$  con passo  $\Delta_0 = 1$  e sia  $D = \{+1, -1\}$

# Perché assumiamo che $f(x)$ sia continua?

$k$	$x_k$	$f(x_k)$	$\Delta_k$
0	-1	4	1
1	-1	4	0.5
2	-0.5	2.25	0.5
3	-0.5	2.25	0.25
4	-0.25	1.5625	0.25
⋮	⋮	⋮	⋮
$2h$	$-2^{-h}$	$(2^{-h} + 1)^2$	$2^{-h}$
$2h + 1$	$-2^{-h}$	$(2^{-h} + 1)^2$	$2^{-h-1}$
⋮	⋮	⋮	⋮

Come si vede, quando  $k \rightarrow \infty$ ,  $x_k \rightarrow 0$  che è il punto di discontinuità ma **non stazionario**

# Perché assumiamo che $f(x)$ sia continua?



# Perché assumiamo che $\nabla f(x)$ sia continuo?

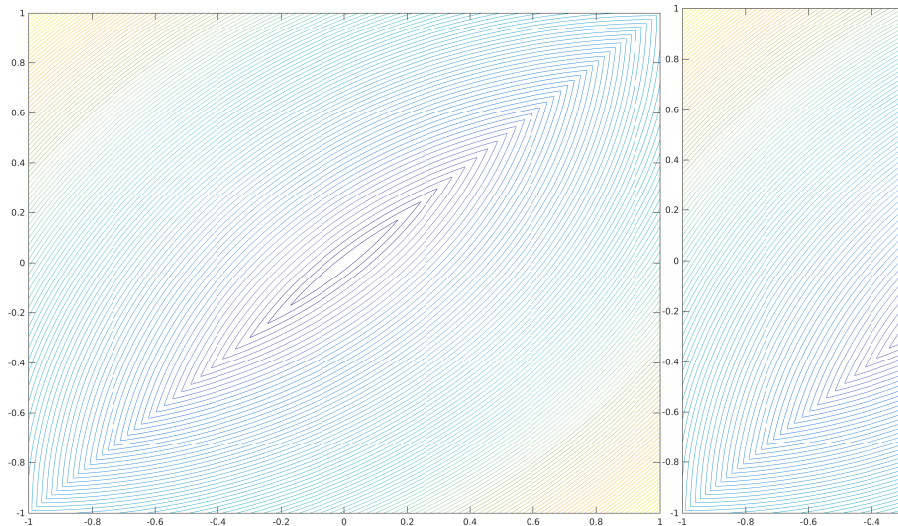
Perché se così non fosse, **non avremmo garanzia di convergenza** a punti stazionari

Esempio:

$$f(x) = \max \{ \|x - c_1\|^2, \|x - c_2\|^2 \}$$

con  $c_1 = (1, -1)^\top$  e  $c_2 = -c_1$

# Perché assumiamo che $\nabla f(x)$ sia continuo?



# Situazioni più complesse

- presenza di vincoli
  - “facili” (limitazioni sulle var.)
  - “difficili” ( $g(x) \leq 0$ , nonlineari generali)
- $f(x)$  e/o  $g(x)$  continue ma non differenziabili
- presenza di più funzioni obiettivo
- presenza di variabili “discrete”
- ricerca di un minimo globale

DFL (Derivative-Free Library), download gratuito di metodi che non usano derivate prime

<http://www.iasi.cnr.it/~liuzzi/DFL>



# Approfondimenti

- **Introduction to Derivative-Free Optimization**

A.R.Conn, K. Scheinberg, L.N. Vicente, MPS-SIAM series on Optimization (2009)

- **Derivative-Free and Blackbox Optimization**

C. Audet, W. Hare, Springer Series in Operations Research and Financial Engineering (2017)

# Definizione del problema

Consideriamo il problema

$$\min_{x \in \mathcal{D}} f(x)$$

con  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $\mathcal{D} \subseteq \mathbb{R}^n$ .

## Definizione (minimo globale)

*Un punto  $x^* \in \mathcal{D}$  è un minimo globale di  $f$  su  $\mathcal{D}$  quando*

$$f(x^*) \leq f(x), \quad \text{per ogni } x \in \mathcal{D}.$$

## Definizione (minimo globale stretto)

*Un punto  $x^* \in \mathcal{D}$  è un minimo globale stretto di  $f$  su  $\mathcal{D}$  quando*

$$f(x^*) < f(x), \quad \text{per ogni } x \in \mathcal{D}, x \neq x^*.$$

# Esistenza della soluzione

Attenzione, per essere certi che il problema ammette soluzione globale, assumiamo:

- $f$  almeno continua su  $\mathcal{D}$
- $\mathcal{D}$  insieme non vuoto e compatto

# Esistenza della soluzione

## Proposizione (Teorema di Weierstrass)

*Sia  $\mathcal{D} \neq \emptyset$  e compatto. Se  $f$  è continua su  $\mathcal{D}$ , allora esiste un punto di minimo globale di  $f$  su  $\mathcal{D}$ .*

# Una proprietà interessante

Alcuni metodi di ottimizzazione locale godono della seguente

## Proprietà (Attrazione dei minimi globali)

*Sia  $\{x_k\}$  la succ. di punti prodotti da un metodo di ottimizzazione locale e  $x^*$  un minimo globale di  $f$  su  $\mathbb{R}^n$ . Allora, esiste un  $\epsilon > 0$  tale che, se per un indice  $\bar{k}$  si ha  $x_{\bar{k}} \in B(x^*; \epsilon)$  allora:*

- $x_k \in B(x^*; \epsilon)$  per ogni  $k \geq \bar{k}$ ;
- $\lim_{k \rightarrow \infty} x_k = x^*$ .

Se  $f \in C^2$  e nel minimo globale  $x^*$  risulta  $\nabla^2 f(x^*)$  definita positiva, allora

- la maggior parte dei metodi senza derivate;
- la maggior parte delle modifiche glob. convergenti del metodo di Newton

godono della Prop. 1

# Classificazione

La maggior parte dei metodi di ottimizzazione globale considera il problema

$$\min_{x \in \mathcal{D}} f(x)$$

con  $\mathcal{D} \subset \mathbb{R}^n$  sottoinsieme compatto e non vuoto di  $\mathbb{R}^n$ .

La totalità dei metodi di ottimizzazione globale rientra in una delle seguenti due categorie:

- metodi probabilistici;
- metodi deterministici.

# Algoritmo di campionamento uniforme

Se  $\mathcal{A}$  è un intorno di un minimo globale  $x^*$  di  $f$  su  $\mathcal{D}$ , e si generano molti punti a caso su  $\mathcal{D}$ , allora è sempre più probabile che uno di questi punti cada in  $\mathcal{A}$ .

Questa considerazione è alla base del primo metodo di ottimizzazione globale che vediamo.

INPUT:  $\text{maxit} > 0$

Genera  $x_0 \in \mathcal{U}(\mathcal{D})$  e poni  $f_{best} \leftarrow f(x_0)$  e  $x_0^* \leftarrow x_0$

**for**  $k = 1, \dots, \text{maxit}$

    Genera  $x_k \in \mathcal{U}(\mathcal{D})$

**if**  $f(x_k) < f_{best}$  **then**  $f_{best} \leftarrow f(x_k)$ ,  $x_k^* \leftarrow x_k$

**else** poni  $x_k^* \leftarrow x_{k-1}^*$

**end for**

OUTPUT:  $\{x_k^*\}$

# Generazione di punti a caso su $\mathcal{D}$

Il risultato fondamentale sul quale si basano molti metodi probabilistici è il seguente.

## Proposizione

*Sia  $\{x_k\}$  una succ. di punti aleatori scelti a caso su  $\mathcal{D}$ . Allora, per ogni sottoinsieme  $\mathcal{A} \subset \mathcal{D}$  tale che*

$$\text{meas}(\mathcal{D}) > \text{meas}(\mathcal{A}) > 0$$

*si ha*

$$\lim_{k \rightarrow \infty} \text{Prob} \{X_k \cap \mathcal{A} \neq \emptyset\} = 1 \quad \text{con } X_k = \{x_0, x_1, \dots, x_k\}.$$

**dim.** Sia

$$p = \frac{\text{meas}(\mathcal{A})}{\text{meas}(\mathcal{D})}$$

per cui risulta  $0 < p < 1$ . È facile riconoscere in  $p$  la probabilità che un punto scelto a caso in  $\mathcal{D}$  appartenga ad  $\mathcal{A}$ .



Generazione di punti a caso su  $\mathcal{D}$ 

Ora, la probabilità che, generati  $k$  vettori, almeno uno appartenga ad  $\mathcal{A}$  è

$$\text{Prob}\{X_k \cap \mathcal{A} \neq \emptyset\} = 1 - \text{Prob}\{X_k \cap \mathcal{A} = \emptyset\} = 1 - (1 - p)^k$$

da questo segue

$$\lim_{k \rightarrow \infty} \text{Prob}\{X_k \cap \mathcal{A} \neq \emptyset\} = 1.$$



# Convergenza dell' Algoritmo di campionamento uniforme

## Proposizione

Sia  $\max_{\mathcal{D}} f = +\infty$  e  $\{x_k^*\}$  la successione (infinita) di punti prodotti dall'algoritmo. Allora, se  $f \in C$ , per ogni  $\epsilon > 0$  si ha:

$$\lim_{k \rightarrow \infty} \text{Prob} \{x_k^* \in \{x \in \mathcal{D} : f(x) < f^* + \epsilon\}\} = 1,$$

con  $f^*$  valore ottimo globale di  $f$  su  $\mathcal{D}$ .

**dim.** Definiamo

$$\mathcal{A} = \{x \in \mathcal{D} : f(x) < f^* + \epsilon\}.$$

La continuità di  $f$  implica che  $\text{meas}(\mathcal{A}) > 0$  e si nota facilmente che

$$\text{Prob} \{x_k^* \in \mathcal{A}\} = \text{Prob} \{X_k \cap \mathcal{A} \neq \emptyset\},$$

dove  $X_k = \{x_0, x_1, \dots, x_k\}$ .

□

# Nota Bene

Quante iterazioni  $k$  occorre eseguire affinché  $Prob\{x_k^* \in \mathcal{A}\} > \alpha \in (0, 1)$ ?

$$Prob\{x_k^* \in \mathcal{A}\} = 1 - \left(1 - \frac{meas(\mathcal{A})}{meas(\mathcal{D})}\right)^k > \alpha$$

$$1 - \alpha > \left(1 - \frac{meas(\mathcal{A})}{meas(\mathcal{D})}\right)^k$$

$$\log(1 - \alpha) > k \log\left(1 - \frac{meas(\mathcal{A})}{meas(\mathcal{D})}\right)$$

$$\frac{\log(1 - \alpha)}{\log\left(1 - \frac{meas(\mathcal{A})}{meas(\mathcal{D})}\right)} < k$$

# Nota Bene

Quante iterazioni  $k$  occorre eseguire affinché  $Prob\{x_k^* \in \mathcal{A}\} > \alpha \in (0, 1)$ ?

$$k > \frac{\log(1 - \alpha)}{\log\left(1 - \frac{meas(\mathcal{A})}{meas(\mathcal{D})}\right)}$$

Se ipotizziamo  $\mathcal{D}$  t.c.  $meas(\mathcal{D}) = 1$  e  $\mathcal{A}$  un ipercubo di lato  $\ell$  quindi  $meas(\mathcal{A}) = \ell^n$ ,

$$k > \frac{\log(1 - \alpha)}{\log(1 - \ell^n)} = \mathcal{O}\left(\frac{1}{\ell^n}\right)!!!$$