# Ottimizzazione dei Sistemi Complessi

G. Liuzzi[1]

Giovedì 22 Marzo 2018

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

# Un metodo generale

INPUT: $x_0$, $\Delta_0$, $\Delta_{min}$, $\texttt{maxit}$, $D = \{\pm e_i, i = 1, \dots, n\}$

$k \leftarrow 0$

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

**if** $\boxed{\exists\ y_k \text{ s.t. } f(y_k) \leq f(x_k) - \gamma\Delta_k}$ **then**

$\Delta_{k+1} \leftarrow \Delta_k$

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

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

**else**

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

**endif**

$\boxed{\text{Find } x_{k+1} \text{ s.t. } f(x_{k+1}) \leq f(y_k)}$

$k \leftarrow k + 1$

**end while**

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

# Metodi più avanzati

INPUT: $x_0$, $\tilde{\Delta}_0^i$, $\Delta_{min}$, $\texttt{maxit}$, $D = \{e_i, i = 1, \dots, n\}$

$k \leftarrow 0$

**while** $k \leq \texttt{maxit}$ **and** $\max\{\tilde{\Delta}_k^i\} \geq \Delta_{min}$ **do**

**if** $\exists\ y_k$ s.t. $f(y_k) \leq f(x_k) - \gamma \max\{\tilde{\Delta}_k^i\}$ **then**

$\tilde{\Delta}_{k+1}^i \leftarrow \tilde{\Delta}_k^i$, $\Delta_k^i \leftarrow \tilde{\Delta}_k^i$ (*passo "pattern"*)

**else** (*WARNING: bisogna forzare la convergenza*)

$\boxed{\text{usa le dir. in } D \text{ per ridurre } f}$

**endif**

Find $x_{k+1}$ s.t. $f(x_{k+1}) \leq f(y_k)$

$k \leftarrow k + 1$

**end while**

RETURN: $\{x_k\}, \{\tilde{\Delta}_k^i\}, \{\Delta_k^i\}$ successioni di punti e passi

# Metodi più avanzati

$\boxed{\text{usa le dir. in } D \text{ per ridurre } f}$

$y_k^1 \leftarrow x_k$

**for** $i = 1, \dots, n$

**if** $f(y_k^i + \tilde{\Delta}_k^i d_i) \leq f(y_k^i) - \gamma(\tilde{\Delta}_k^i)^2$ **then**

$p_k^i \leftarrow d_i$ e $\boxed{\text{calcola } \Delta_k^i \text{ e } \tilde{\Delta}_{k+1}^i \text{ mediante una espansione lungo } p_k^i}$

**elseif** $f(y_k^i - \tilde{\Delta}_k^i d_i) \leq f({y_k}^i) - \gamma(\tilde{\Delta}_k^i)^2$ **then**

$p_k^i \leftarrow -d_i$ e $\boxed{\text{calcola } \Delta_k^i \text{ e } \tilde{\Delta}_{k+1}^i \text{ mediante una espansione lungo } p_k^i}$

**else** $\Delta_k^i \leftarrow 0,\ p_k^i \leftarrow d_i,\ \tilde{\Delta}_{k+1}^i \leftarrow \tilde{\Delta}_k^i / 2$

$y_k^{i+1} \leftarrow y_k^i + \Delta_k^i p_k^i$

**end for**

$y_k \leftarrow y_k^{n+1}$

# Espansione lungo $d_k^i$ con suff. decremento

Determina il più piccolo intero $j \in \{0, 1, 2, \dots\}$ tale che

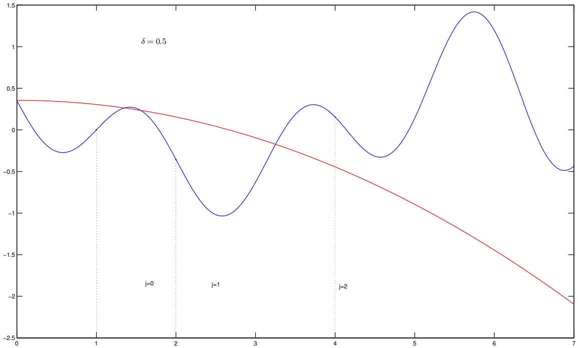
$$\begin{aligned} f(y_k^i + 2^j \tilde{\Delta}_k^i p_k^i) &\leq f(y_k^i) - \gamma(2^j \tilde{\Delta}_k^i)^2, \\ f(y_k^i + 2^{j+1} \tilde{\Delta}_k^i p_k^i) &> f(y_k^i) - \gamma(2^{j+1} \tilde{\Delta}_k^i)^2 \end{aligned}$$

Poni $\Delta_k^i \leftarrow 2^j \tilde{\Delta}_k^i$ e $\tilde{\Delta}_{k+1}^i \leftarrow \Delta_k^i$

# Ricerca di linea senza derivate

Siano

$$
\begin{aligned}
\psi(\beta) &= f(y_k^i + \beta p_k^i) \text{ e} \\
\Psi(\beta) &= f(y_k^i) - \gamma\beta^2
\end{aligned}
$$



# Proprietà teoriche

Se $L(x_0)$ è compatto, si può (facilmente) dimostrare che:

- La ricerca di linea restituisce un valore di $j$ finito
- $\lim_{k\to\infty} \max\{\tilde{\Delta}_k^i, \Delta_k^i\} = 0$

Se $L(x_0)$ è compatto e $f$ continuamente differenziabile, allora

- $\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0$ (almeno un p.to limite è stazionario)

# 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:

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

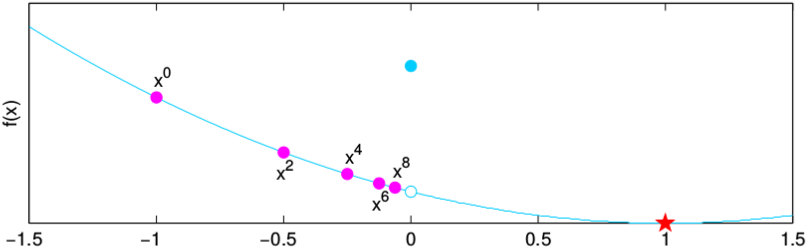
Supponiamo che un metodo "compass search" parta dal punto $x_0 = -1$ con passo $\Delta_0 = 1$ e sia $D = \{+1, -1\}$

# 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 |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $2h$ | $-2^{-h}$ | $(2^{-h}+1)^2$ | $2^{-h}$ |
| $2h+1$ | $-2^{-h}$ | $(2^{-h}+1)^2$ | $2^{-h-1}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |

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

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



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

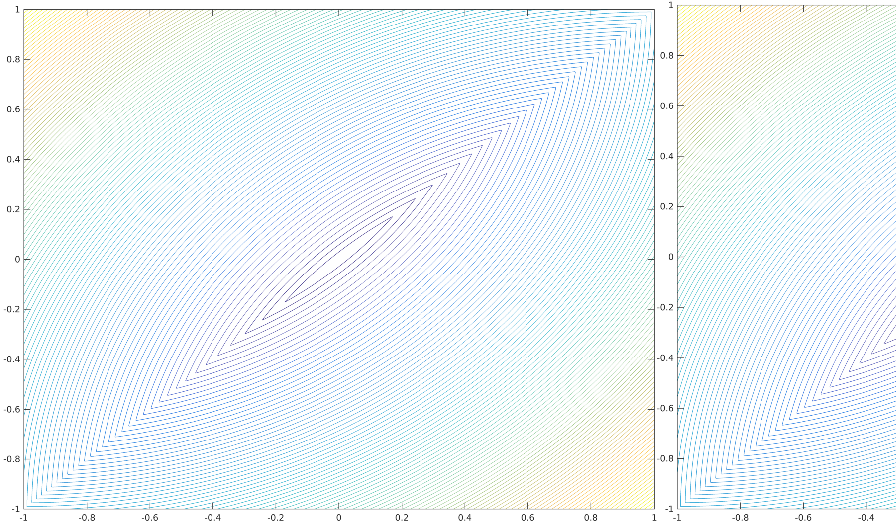
Perché se così non fosse, **non avremmo garanzia di convergenza** a punti stazionari

Esempio:

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

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`

`http://www.diag.uniroma1.it/~lucidi/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**)

# Problemi vincolati

Problema vincolato nonlineare generale

$$\begin{array}{llll} \min\limits_{x} & f(x) & f:\mathbb{R}^n \mapsto \mathbb{R} & \\ c.v & h(x)=0 & h:\mathbb{R}^n \mapsto \mathbb{R}^p & p \leq n \\ & g(x)\leq 0 & g:\mathbb{R}^n \mapsto \mathbb{R}^m & \end{array} \qquad (P_0)$$

Sia $\mathcal{F}$ l'insieme ammissibile di $(P_0)$
Problema con soli vincoli di uguaglianza o disuguaglianza

$$\begin{array}{ll} \min\limits_{x} & f(x) \\ c.v & h(x)=0 \end{array} \qquad\qquad \begin{array}{ll} \min\limits_{x} & f(x) \\ c.v & g(x)\leq 0 \end{array}$$

# Definizione di minimo

## Definizione (Minimo globale)

$x^* \in \mathcal{F}$ *è un punto di minimo globale* **stretto** *quando*

$$f(x^*) \leq < f(x), \qquad \forall x \in \mathcal{F},\ x \neq x^*$$

## Definizione (Minimo locale)

$x^* \in \mathcal{F}$ *è un punto di minimo locale* **stretto** *quando, esiste un* $\epsilon > 0$ *t.c.*

$$f(x^*) \leq < f(x), \qquad \forall x \in \mathcal{F} \cap \mathcal{B}(x^*, \epsilon),\ x \neq x^*$$

# Problemi vincolati

Sia $x \in \mathbb{R}^n$, definiamo

- $I_-(x) = \{i : g_i(x) < 0\}$
- $I_0(x) = \{i : g_i(x) = 0\}$
- $I_+(x) = \{i : g_i(x) > 0\}$

# Problemi vincolati

Trasformazioni:

$$\begin{array}{ll} \min\limits_{x} & f(x) \\ c.v & h(x) = 0 \end{array} \qquad \begin{array}{ll} \min\limits_{x} & f(x) \\ c.v & h(x) \leq 0 \\ & -h(x) \leq 0 \end{array}$$

$$\begin{array}{ll} \min\limits_{x} & f(x) \\ c.v & g(x) \leq 0 \end{array} \qquad \begin{array}{ll} \min\limits_{x,y} & f(x) \\ c.v & g_i(x) + y_i^2 = 0, \;\; i = 1, \dots, m \end{array}$$

$$\begin{array}{ll} \min\limits_{x} & f(x) \\ c.v & g(x) \leq 0 \end{array} \qquad \begin{array}{ll} \min\limits_{x,y} & f(x) \\ c.v & g_i(x) = y_i, \;\; y_i \leq 0, \;\; i = 1, \dots, m \end{array}$$

# Condizioni di ottimalità (richiami??)

Supponiamo:

- $f, g, h$ continuamente differenziabili
- di conoscere un punto di **minimo locale** $x^*$ del problema $(P_0)$

  Quindi, $x^*$ **ammissibile**, e.g. $g(x^*) \leq 0$, $h(x^*) = 0$

Allora

## Teorema (Fritz-John, 1948)

*Esiste un numero* $\lambda_0^* \geq 0$ *e dei moltiplicatori* $\lambda_1^*, \ldots, \lambda_m^* \geq 0$, $\mu_1^*, \ldots, \mu_p^*$, *non tutti nulli, tali che:*

$$\lambda_0^* \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla h_j(x^*) = 0,$$

$$\lambda_i^* g_i(x^*) = 0, \quad \textit{per ogni } i = 1, \ldots, m.$$

# Condizioni di ottimalità (richiami??)

Supponiamo:

- $f, g, h$ continuamente differenziabili
- di conoscere un punto di **minimo locale** $x^*$ del problema $(P_0)$
- che in $x^*$ i vincoli siano **regolari**

e.g. $\left\{\nabla h_j(x^*),\ j = 1, \dots, p,\ \nabla g_i(x^*),\ i \in I_0(x^*)\right\}$ lin. indip.

Allora

**Teorema (Kaursh, 1939, Kuhn-Tucker, 1951)**

*Esistono dei moltiplicatori* $\lambda_1^*, \dots, \lambda_m^* \geq 0$, $\mu_1^*, \dots, \mu_p^*$ *tali che:*

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla h_j(x^*) = 0,$$

$$\lambda_i^* g_i(x^*) = 0, \quad \textit{per ogni } i = 1, \dots, m.$$

# Esempio

$$\begin{array}{ll} \min\limits_{x} & x_1 + x_2 \\ c.v & x_1^2 + x_2^2 - 2 = 0 \end{array}$$

$$\begin{array}{ll} \min\limits_{x} & x_1 + x_2 \\ c.v & x_1^2 + x_2^2 - 2 \leq 0 \end{array}$$

$$\begin{array}{ll} \min\limits_{x} & x_1 + x_2 \\ c.v & x_1^2 + x_2^2 - 2 \leq 0, \;\; x_2 \geq 0 \end{array}$$

$$\begin{array}{ll} \min\limits_{x} & x_1 + x_2 \\ c.v & x_2 - x_1^3 \leq 0, \;\; x_2 \geq 0 \end{array}$$

# Condizioni di ottimalità (richiami??)

Supponiamo:

- $f, g$ differenziabili e **convesse**
- $h$ **affini**, e.g. $h_j(x) = a^\top x - b = 0$

**Teorema (Condizione Sufficiente)**

*Se* $x^* \in \mathcal{F}$ *e esistono dei moltiplicatori* $\lambda_1^*, \dots, \lambda_m^* \geq 0$, $\mu_1^*, \dots, \mu_p^*$ *tali che:*

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla h_j(x^*) = 0,$$

$$\lambda_i^* g_i(x^*) = 0, \quad \textit{per ogni } i = 1, \dots, m,$$

*allora* $x^*$ *è un minimo locale (globale) di* $(P_0)$

# Condizioni di ottimalità (richiami??)

Ci sono condizioni sufficienti di ottimo nel caso **non convesso**?
Supponiamo:

- $f, g, h$ **due volte** continuamente differenziabili
- $x^*, \lambda^*, \mu^*$ soddisfano KKT

**Teorema (Condizione Sufficiente)**

*Se* $d^\top \nabla^2_{xx} L(x^*, \lambda^*, \mu^*) d > 0$, *per ogni* $d \in Y^*$, $d \neq 0$

$$
\begin{aligned}
Y^* \;=\; \Big\{ d \in \mathbb{R}^n : \nabla h_j(x^*)^\top d = 0, j = 1, \dots, p \\
\nabla g_i(x^*)^\top d = 0,\; i \in I_0(x^*), \lambda_i^* > 0, \\
\nabla g_i(x^*)^\top d \geq 0,\; i \in I_0(x^*), \lambda_i^* = 0 \Big\}
\end{aligned}
$$

*allora* $x^*$ *è un minimo locale stretto di* $(P_0)$