Ottimizzazione dei Sistemi Complessi

G. Liuzzi[1]

Giovedì 9 Marzo 2017

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

Il metodo di Nelder-Mead

Alla generica iterazione k lo stato dell'algoritmo è dato da:

$$X_k = \{x_1, x_2, \dots, x_{n+1}\}$$

indichiamo $f_i = f(x_i)$ e supponiamo X_k ordinato in modo che

$$f_1 \leq f_2 \leq \cdots \leq f_{n+1}.$$

Ricordiamo l'espressione del "centroide" dei primi n punti in X_k

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Idea: Sfruttare x_{n+1} e $\bar{x}$ per cercare un punto "migliore" di x_{n+1}

Il metodo di Nelder-Mead

Alla generica iterazione k lo stato dell'algoritmo è dato da:

$$X_k = \{x_1, x_2, \dots, x_{n+1}\}$$

indichiamo $f_i = f(x_i)$ e supponiamo X_k ordinato in modo che

$$f_1 \leq f_2 \leq \cdots \leq f_{n+1}.$$

Ricordiamo l'espressione del "centroide" dei primi n punti in X_k

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Idea: Sfruttare x_{n+1} e $\bar{x}$ per cercare un punto "migliore" di x_{n+1}

Il metodo di Nelder-Mead

Alla generica iterazione k lo stato dell'algoritmo è dato da:

$$X_k = \{x_1, x_2, \dots, x_{n+1}\}$$

indichiamo $f_i = f(x_i)$ e supponiamo X_k ordinato in modo che

$$f_1 \leq f_2 \leq \cdots \leq f_{n+1}.$$

Ricordiamo l'espressione del "centroide" dei primi n punti in X_k

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Idea: Sfruttare x_{n+1} e $\bar{x}$ per cercare un punto "migliore" di x_{n+1}

Il metodo di Nelder-Mead

Dato un parametro μ, definiamo

$$x(\mu) = \bar{x} + \mu(\bar{x} - x_{n+1})$$

Il metodo usa: $-1 < \mu_{ic} < 0 < \mu_{oc} < \mu_r < \mu_e$ e

$$\begin{array}{llll} \text{(inner contraction)} & x^{ic} = x(\mu_{ic}), & f^{ic} = f(x^{ic}) \\ \text{(outer contraction)} & x^{oc} = x(\mu_{oc}), & f^{oc} = f(x^{oc}) \\ \text{(reflect)} & x^{r} = x(\mu_{r}), & f^{r} = f(x^{r}) \\ \text{(expand)} & x^{e} = x(\mu_{e}), & f^{e} = f(x^{e}) \end{array}$$

Il metodo di Nelder-Mead

Dato un parametro μ, definiamo

$$x(\mu) = \bar{x} + \mu(\bar{x} - x_{n+1})$$

Il metodo usa: $-1 < \mu_{ic} < 0 < \mu_{oc} < \mu_r < \mu_e$ e

$$\begin{array}{lrclrcl}
\text{(inner contraction)} & x^{ic} & = & x(\mu_{ic}), & f^{ic} & = & f(x^{ic}) \\
\text{(outer contraction)} & x^{oc} & = & x(\mu_{oc}), & f^{oc} & = & f(x^{oc}) \\
\text{(reflect)} & x^{r} & = & x(\mu_{r}), & f^{r} & = & f(x^{r}) \\
\text{(expand)} & x^{e} & = & x(\mu_{e}), & f^{e} & = & f(x^{e})
\end{array}$$

Iterazione k

- Se $\boxed{f_1 \leq f^r < f_n}$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^r\}$ **FINE**
- Se $\boxed{f^r < f_1}$, allora
  - Se $f^e < f^r$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^e\}$ **FINE**
  - altrimenti $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^r\}$ **FINE**
- Se $\boxed{f_n \leq f^r < f_{n+1}}$, allora
  - Se $f^{oc} \leq f^r$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^{oc}\}$ **FINE**
  - altrimenti **shrink**
- Se $\boxed{f_{n+1} \leq f^r}$, allora
  - Se $f^{ic} < f_{n+1}$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^{ic}\}$ **FINE**
  - altrimenti **shrink**
- **shrink**:
  - $X_{k+1} = \{x_1, \hat{x}_2, \dots, \hat{x}_{n+1}\}$ dove
  - $\hat{x}_i = x_1 + \gamma(x_i - x_1)$, $i = 2, \dots, n+1$, $\gamma \in (0,1)$

Iterazione k

- Se $\boxed{f_1 \leq f^r < f_n}$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^r\}$ **FINE**
- Se $\boxed{f^r < f_1}$, allora
  - Se $f^e < f^r$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^e\}$ **FINE**
  - altrimenti $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^r\}$ **FINE**
- Se $\boxed{f_n \leq f^r < f_{n+1}}$, allora
  - Se $f^{oc} \leq f^r$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^{oc}\}$ **FINE**
  - altrimenti **shrink**
- Se $\boxed{f_{n+1} \leq f^r}$, allora
  - Se $f^{ic} < f_{n+1}$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^{ic}\}$ **FINE**
  - altrimenti **shrink**
- **shrink**:
  - $X_{k+1} = \{x_1, \hat{x}_2, \ldots, \hat{x}_{n+1}\}$ dove
  - $\hat{x}_i = x_1 + \gamma(x_i - x_1)$, $i = 2, \ldots, n+1$, $\gamma \in (0,1)$

Iterazione k

- Se $\boxed{f_1 \le f^r < f_n}$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^r\}$ **FINE**
- Se $\boxed{f^r < f_1}$, allora
  - Se $f^e < f^r$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^e\}$ **FINE**
  - altrimenti $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^r\}$ **FINE**
- Se $\boxed{f_n \le f^r < f_{n+1}}$, allora
  - Se $f^{oc} \le f^r$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^{oc}\}$ **FINE**
  - altrimenti **shrink**
- Se $\boxed{f_{n+1} \le f^r}$, allora
  - Se $f^{ic} < f_{n+1}$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^{ic}\}$ **FINE**
  - altrimenti **shrink**
- **shrink**:
  - $X_{k+1} = \{x_1, \hat{x}_2, \dots, \hat{x}_{n+1}\}$ dove
  - $\hat{x}_i = x_1 + \gamma(x_i - x_1)$, $i = 2, \dots, n+1$, $\gamma \in (0,1)$

Iterazione k

- Se $\boxed{f_1 \leq f^r < f_n}$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^r\}$ **FINE**
- Se $\boxed{f^r < f_1}$, allora
  - Se $f^e < f^r$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^e\}$ **FINE**
  - altrimenti $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^r\}$ **FINE**
- Se $\boxed{f_n \leq f^r < f_{n+1}}$, allora
  - Se $f^{oc} \leq f^r$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^{oc}\}$ **FINE**
  - altrimenti **shrink**
- Se $\boxed{f_{n+1} \leq f^r}$, allora
  - Se $f^{ic} < f_{n+1}$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^{ic}\}$ **FINE**
  - altrimenti **shrink**
- **shrink**:
  - $X_{k+1} = \{x_1, \hat{x}_2, \ldots, \hat{x}_{n+1}\}$ dove
  - $\hat{x}_i = x_1 + \gamma(x_i - x_1)$, $i = 2, \ldots, n+1$, $\gamma \in (0,1)$

Iterazione k

- Se $\boxed{f_1 \leq f^r < f_n}$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^r\}$ **FINE**
- Se $\boxed{f^r < f_1}$, allora
  - Se $f^e < f^r$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^e\}$ **FINE**
  - altrimenti $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^r\}$ **FINE**
- Se $\boxed{f_n \leq f^r < f_{n+1}}$, allora
  - Se $f^{oc} \leq f^r$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^{oc}\}$ **FINE**
  - altrimenti **shrink**
- Se $\boxed{f_{n+1} \leq f^r}$, allora
  - Se $f^{ic} < f_{n+1}$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^{ic}\}$ **FINE**
  - altrimenti **shrink**
- **shrink**:
  - $X_{k+1} = \{x_1, \hat{x}_2, \dots, \hat{x}_{n+1}\}$ dove
  - $\hat{x}_i = x_1 + \gamma(x_i - x_1)$, $i = 2, \dots, n+1$, $\gamma \in (0,1)$

Iterazione k

- Se $\boxed{f_1 \leq f^r < f_n}$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^r\}$ **FINE**
- Se $\boxed{f^r < f_1}$, allora
  - Se $f^e < f^r$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^e\}$ **FINE**
  - altrimenti $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^r\}$ **FINE**
- Se $\boxed{f_n \leq f^r < f_{n+1}}$, allora
  - Se $f^{oc} \leq f^r$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^{oc}\}$ **FINE**
  - altrimenti **shrink**
- Se $\boxed{f_{n+1} \leq f^r}$, allora
  - Se $f^{ic} < f_{n+1}$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^{ic}\}$ **FINE**
  - altrimenti **shrink**
- **shrink**:
  - $X_{k+1} = \{x_1, \hat{x}_2, \ldots, \hat{x}_{n+1}\}$ dove
  - $\hat{x}_i = x_1 + \gamma(x_i - x_1)$, $i = 2, \ldots, n+1$, $\gamma \in (0,1)$

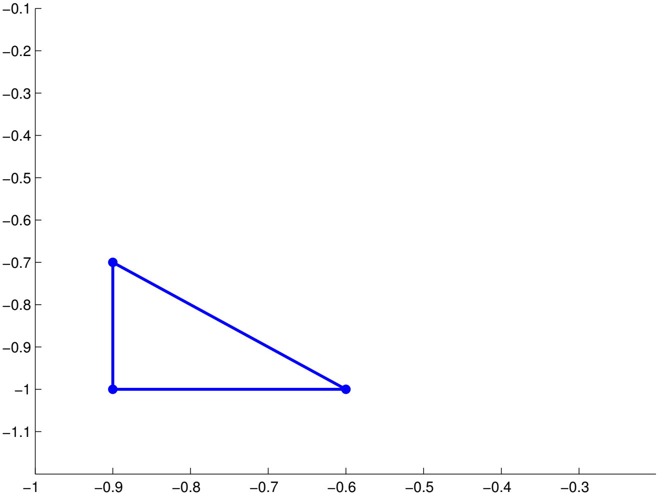
Iterazione k

- Se $\boxed{f_1 \le f^r < f_n}$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^r\}$ **FINE**
- Se $\boxed{f^r < f_1}$, allora
  - Se $f^e < f^r$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^e\}$ **FINE**
  - altrimenti $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^r\}$ **FINE**
- Se $\boxed{f_n \le f^r < f_{n+1}}$, allora
  - Se $f^{oc} \le f^r$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^{oc}\}$ **FINE**
  - altrimenti **shrink**
- Se $\boxed{f_{n+1} \le f^r}$, allora
  - Se $f^{ic} < f_{n+1}$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^{ic}\}$ **FINE**
  - altrimenti **shrink**
- **shrink**:
  - $X_{k+1} = \{x_1, \hat{x}_2, \dots, \hat{x}_{n+1}\}$ dove
  - $\hat{x}_i = x_1 + \gamma(x_i - x_1)$, $i = 2, \dots, n+1$, $\gamma \in (0,1)$

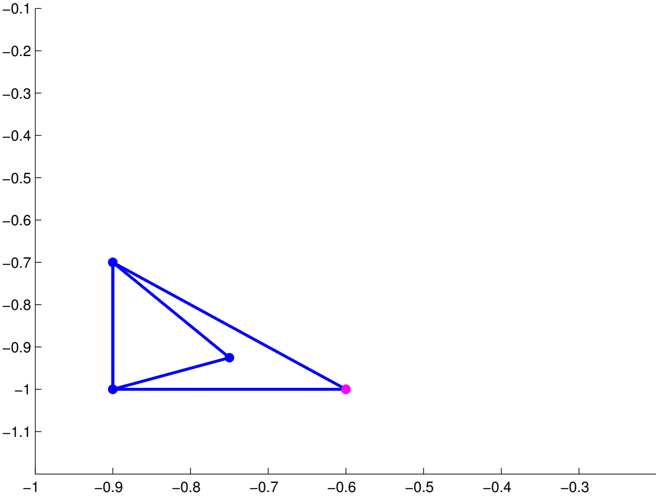
Iterazione k

- Se $\boxed{f_1 \leq f^r < f_n}$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^r\}$ **FINE**
- Se $\boxed{f^r < f_1}$, allora
  - Se $f^e < f^r$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^e\}$ **FINE**
  - altrimenti $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^r\}$ **FINE**
- Se $\boxed{f_n \leq f^r < f_{n+1}}$, allora
  - Se $f^{oc} \leq f^r$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^{oc}\}$ **FINE**
  - altrimenti **shrink**
- Se $\boxed{f_{n+1} \leq f^r}$, allora
  - Se $f^{ic} < f_{n+1}$, allora $X_{k+1} = X_k \setminus \{x_{n+1}\} \cup \{x^{ic}\}$ **FINE**
  - altrimenti **shrink**
- **shrink**:
  - $X_{k+1} = \{x_1, \hat{x}_2, \dots, \hat{x}_{n+1}\}$ dove
  - $\hat{x}_i = x_1 + \gamma(x_i - x_1)$, $i = 2, \dots, n+1$, $\gamma \in (0,1)$

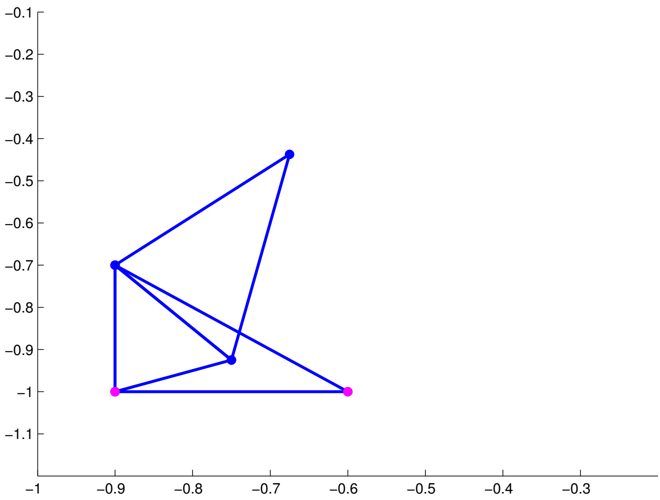
Esempio su Funzione di Broyden



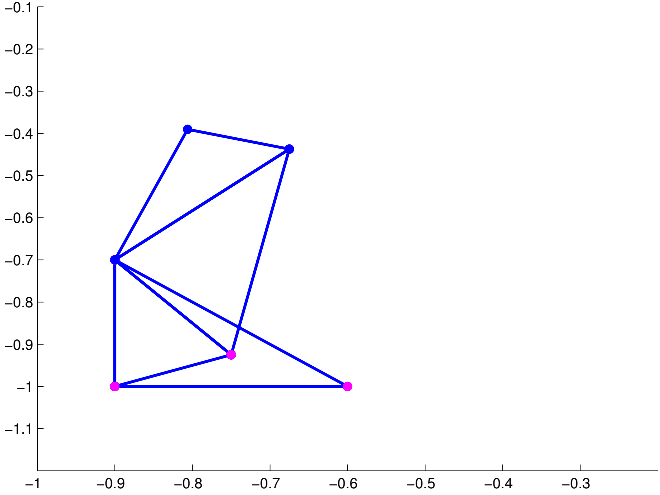
Esempio su Funzione di Broyden



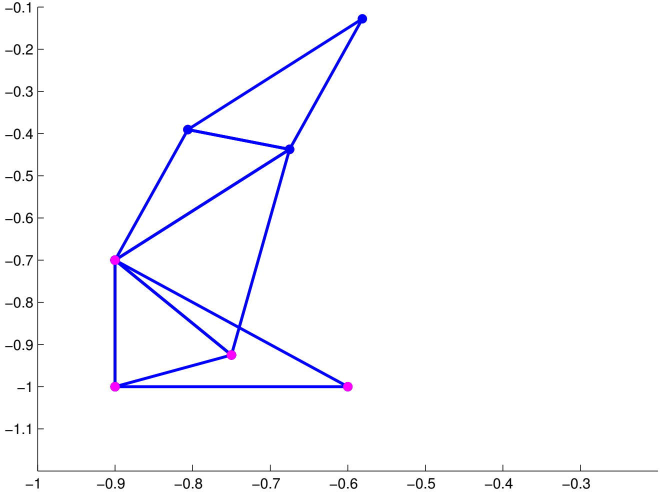
Esempio su Funzione di Broyden



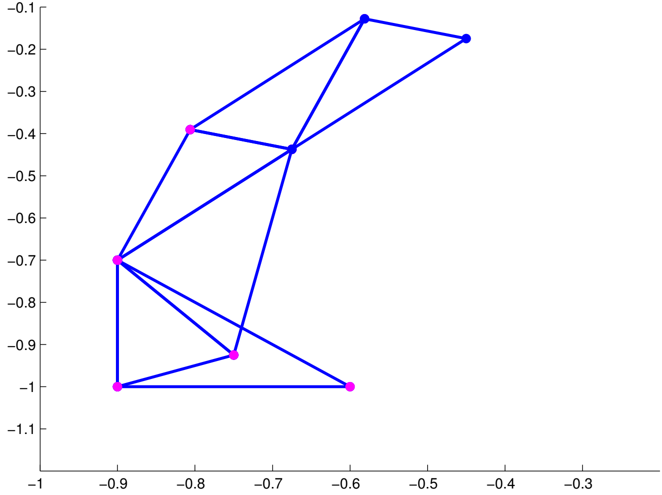
Esempio su Funzione di Broyden



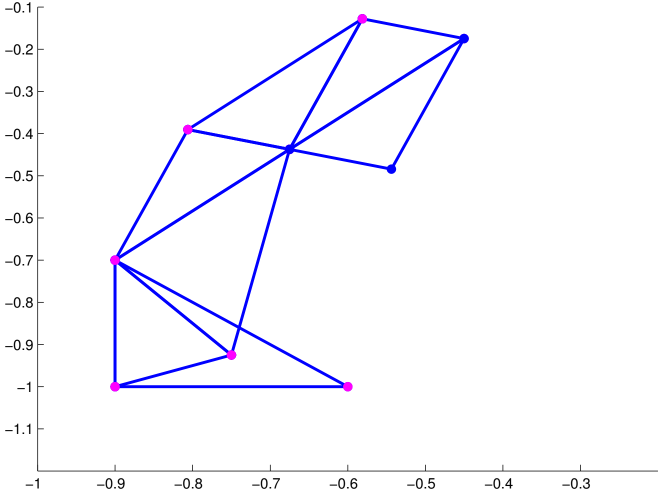
Esempio su Funzione di Broyden



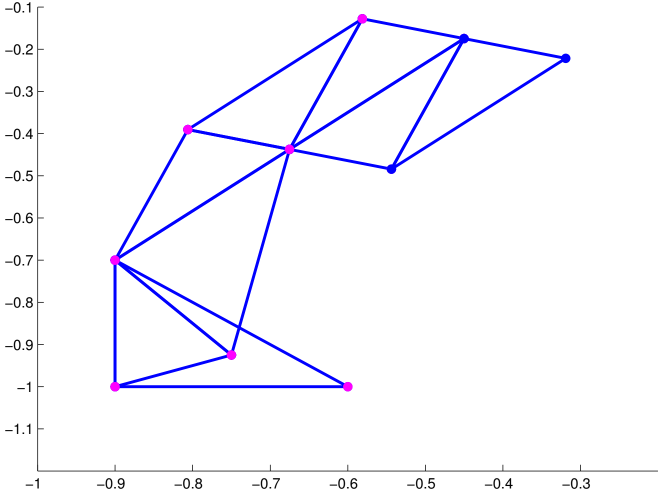
Esempio su Funzione di Broyden



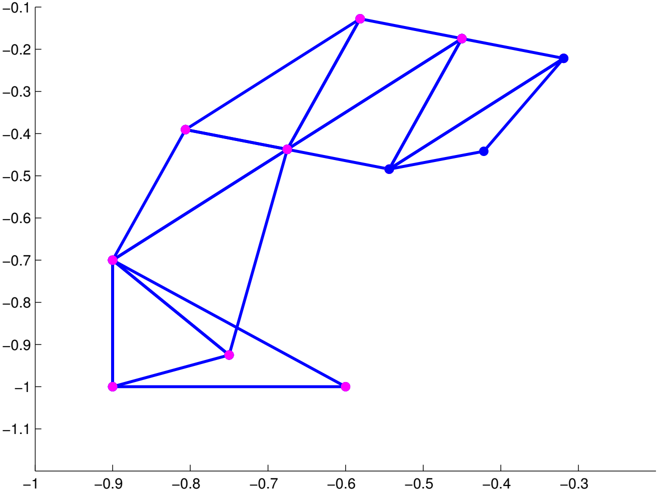
Esempio su Funzione di Broyden



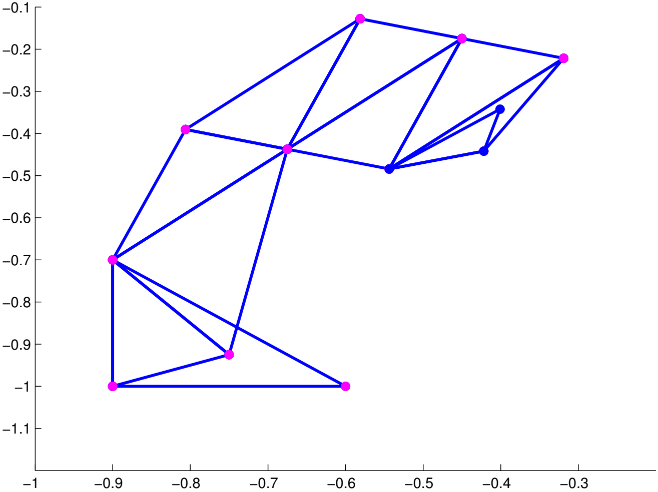
Esempio su Funzione di Broyden



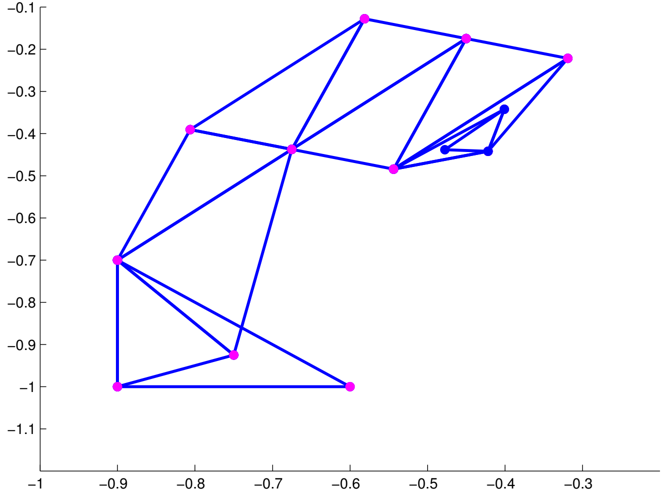
Esempio su Funzione di Broyden



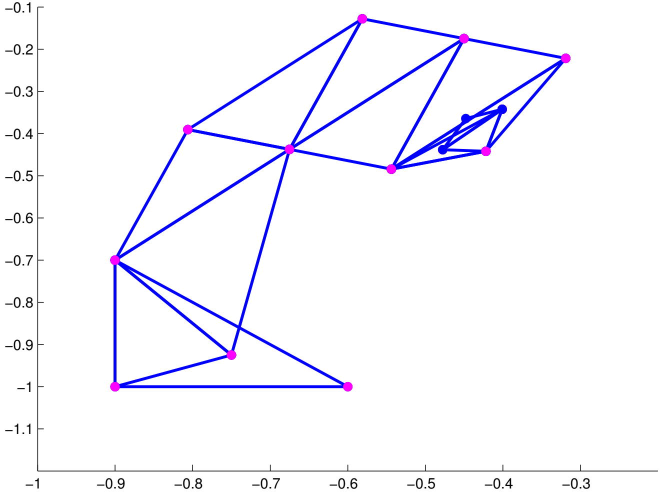
Esempio su Funzione di Broyden



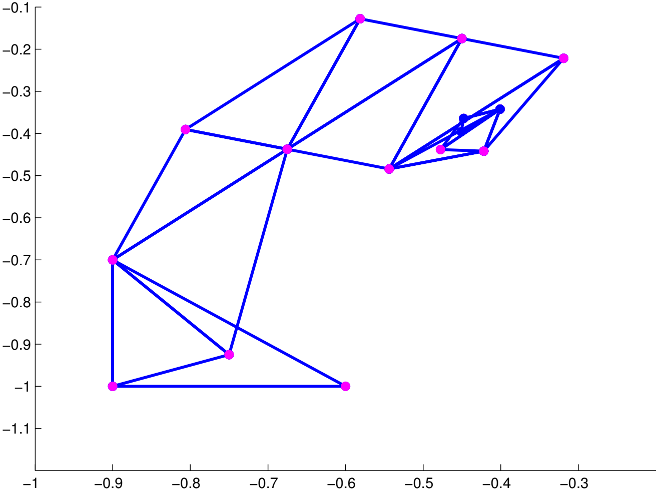
Esempio su Funzione di Broyden



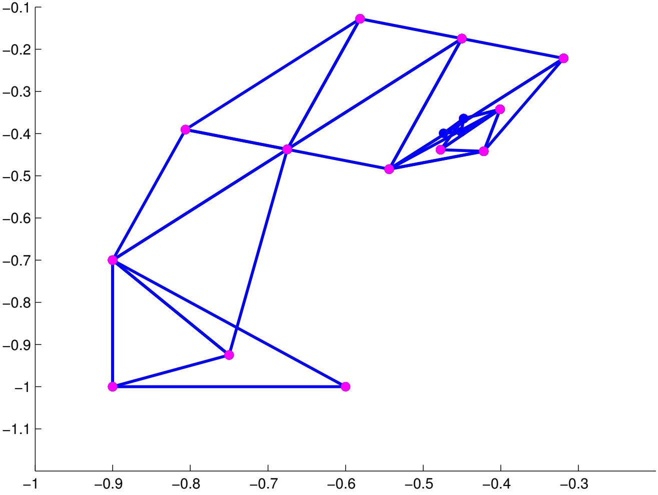
Esempio su Funzione di Broyden



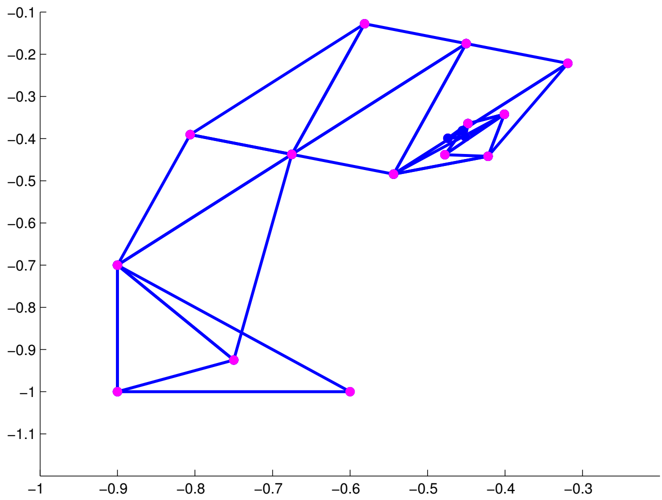
Esempio su Funzione di Broyden



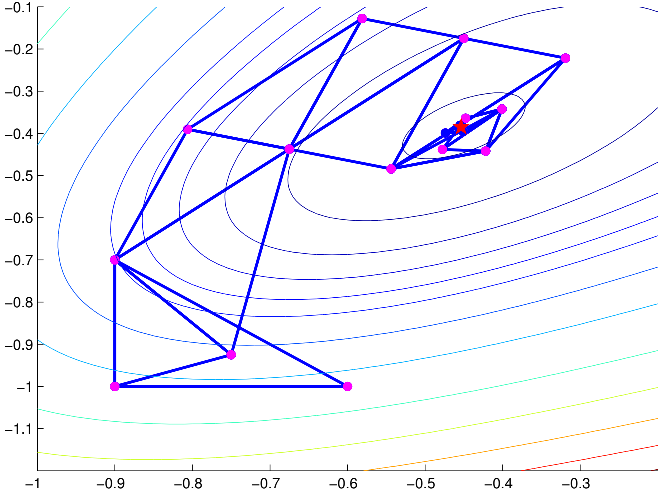
Esempio su Funzione di Broyden



Esempio su Funzione di Broyden



Esempio su Funzione di Broyden



Esercizio

Sia

$$
\begin{aligned}
X &= \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\} \\
F &= \{ \quad 3, \qquad 7, \qquad 1, \quad \}
\end{aligned}
$$

Il centroide $\bar{x}$ di x_1 e x_2 (dopo il riordino) è

$$\bar{x} = \begin{pmatrix} 1 \\ 1/2 \end{pmatrix}$$

Il metodo considera i punti

$$x(\mu) = \bar{x} + \mu d = \bar{x} + \mu(\bar{x} - x_3)$$

per valori $\mu = 1,\ 2,\ 1/2,\ -1/2$

Esercizio

Sia

$$
\begin{aligned}
X &= \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \\
F &= \{ \quad 1, \qquad 3, \qquad 7, \quad \}
\end{aligned}
$$

Il centroide $\bar{x}$ di x_1 e x_2 (dopo il riordino) è

$$\bar{x} = \begin{pmatrix} 1 \\ 1/2 \end{pmatrix}$$

Il metodo considera i punti

$$x(\mu) = \bar{x} + \mu d = \bar{x} + \mu(\bar{x} - x_3)$$

per valori $\mu = 1,\ 2,\ 1/2,\ -1/2$

Esercizio

Sia

$$
\begin{aligned}
X &= \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \\
F &= \{ \quad 1, \qquad 3, \qquad 7, \quad \}
\end{aligned}
$$

Il centroide $\bar{x}$ di x_1 e x_2 (dopo il riordino) è

$$\bar{x} = \begin{pmatrix} 1 \\ 1/2 \end{pmatrix}$$

Il metodo considera i punti

$$x(\mu) = \bar{x} + \mu d = \bar{x} + \mu(\bar{x} - x_3)$$

per valori $\mu = 1,\ 2,\ 1/2,\ -1/2$

Esercizio

Sia

$$
\begin{aligned}
X &= \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \\
F &= \{ \quad 1, \qquad 3, \qquad 7, \quad \}
\end{aligned}
$$

Il centroide $\bar{x}$ di x_1 e x_2 (dopo il riordino) è

$$
\bar{x} = \begin{pmatrix} 1 \\ 1/2 \end{pmatrix}
$$

Il metodo considera i punti

$$
x(\mu) = \bar{x} + \mu d = \bar{x} + \mu(\bar{x} - x_3)
$$

per valori $\mu = 1,\ 2,\ 1/2,\ -1/2$

Esercizio

- si calcola $d = \bar{x} - x_3 = (1,\ -1/2)^\top$
- si calcola $x_r = x(1) = \bar{x} + d = (2,\ 0)^\top$
- si calcola $x_e = x(2) = \bar{x} + 2d = (3,\ -1/2)^\top$
- si calcola $x_{oc} = x(1/2) = \bar{x} + 1/2d = (3/2,\ 1/4)^\top$
- si calcola $x_{ic} = x(-1/2) = \bar{x} - 1/2d = (1/2,\ 3/4)^\top$

Esercizio

- si calcola $d = \bar{x} - x_3 = (1,\ -1/2)^\top$
- si calcola $x_r = x(1) = \bar{x} + d = (2,\ 0)^\top$
- si calcola $x_e = x(2) = \bar{x} + 2d = (3,\ -1/2)^\top$
- si calcola $x_{oc} = x(1/2) = \bar{x} + 1/2d = (3/2,\ 1/4)^\top$
- si calcola $x_{ic} = x(-1/2) = \bar{x} - 1/2d = (1/2,\ 3/4)^\top$

Esercizio

- si calcola $d = \bar{x} - x_3 = (1,\ -1/2)^\top$
- si calcola $x_r = x(1) = \bar{x} + d = (2,\ 0)^\top$
- si calcola $x_e = x(2) = \bar{x} + 2d = (3,\ -1/2)^\top$
- si calcola $x_{oc} = x(1/2) = \bar{x} + 1/2d = (3/2,\ 1/4)^\top$
- si calcola $x_{ic} = x(-1/2) = \bar{x} - 1/2d = (1/2,\ 3/4)^\top$

Esercizio

- si calcola $d = \bar{x} - x_3 = (1,\ -1/2)^\top$
- si calcola $x_r = x(1) = \bar{x} + d = (2,\ 0)^\top$
- si calcola $x_e = x(2) = \bar{x} + 2d = (3,\ -1/2)^\top$
- si calcola $x_{oc} = x(1/2) = \bar{x} + 1/2d = (3/2,\ 1/4)^\top$
- si calcola $x_{ic} = x(-1/2) = \bar{x} - 1/2d = (1/2,\ 3/4)^\top$

Esercizio

- si calcola $d = \bar{x} - x_3 = (1,\ -1/2)^\top$
- si calcola $x_r = x(1) = \bar{x} + d = (2,\ 0)^\top$
- si calcola $x_e = x(2) = \bar{x} + 2d = (3,\ -1/2)^\top$
- si calcola $x_{oc} = x(1/2) = \bar{x} + 1/2d = (3/2,\ 1/4)^\top$
- si calcola $x_{ic} = x(-1/2) = \bar{x} - 1/2d = (1/2,\ 3/4)^\top$

Nelder&Mead

In $\mathbb{R}^2$ siano dati i seguenti punti:

$$x_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, x_2 = \begin{pmatrix} 3 \\ 2 \end{pmatrix}, x_3 = \begin{pmatrix} 1 \\ 3 \end{pmatrix},$$

a cui corrispondono: $f(x_1) = 1$, $f(x_2) = 3$, $f(x_3) = 5$.

Determinare i punti x_r, x_e, x_{oc} e x_{ic} utilizzati nel metodo di Nelder&Mead

Nelder&Mead

In $\mathbb{R}^2$ siano dati i seguenti punti:

$$x_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, x_2 = \begin{pmatrix} 3 \\ 2 \end{pmatrix}, x_3 = \begin{pmatrix} 1 \\ 3 \end{pmatrix},$$

a cui corrispondono: $f(x_1) = 1$, $f(x_2) = 3$, $f(x_3) = 5$.

Determinare i punti x_r, x_e, x_{oc} e x_{ic} utilizzati nel metodo di Nelder&Mead

È un "buon" metodo?

La risposta è:

- generalmente **Si**
- però non ha proprietà di convergenza a punti stazionari
- esistono contro-esempi sui quali converge a punti **NON** stazionari
- p.es. la funzione di McKinnon

È un "buon" metodo?

La risposta è:

- generalmente **Si**
- però non ha proprietà di convergenza a punti stazionari
- esistono contro-esempi sui quali converge a punti **NON** stazionari
- p.es. la funzione di McKinnon

È un “buon” metodo?

La risposta è:

- generalmente **Si**
- però non ha proprietà di convergenza a punti stazionari
- esistono contro-esempi sui quali converge a punti **NON** stazionari
- p.es. la funzione di McKinnon

La funzione di McKinnon

$$f(x,y) = \begin{cases} \theta\phi|x|^\tau + y + y^2 & \text{se } x \leq 0 \\ \theta x^\tau + y + y^2 & \text{se } x > 0 \end{cases}$$

- strettamente convessa per $\tau > 1$
- continuamente differenziabile per $\tau > 1$
- due volte cont. differenziabile per $\tau > 2$
- tre volte cont. differenziabile per $\tau > 3$

La funzione di McKinnon

$$f(x,y) = \begin{cases} \theta\phi|x|^\tau + y + y^2 & \text{se } x \leq 0 \\ \theta x^\tau + y + y^2 & \text{se } x > 0 \end{cases}$$

- strettamente convessa per $\tau > 1$
- continuamente differenziabile per $\tau > 1$
- due volte cont. differenziabile per $\tau > 2$
- tre volte cont. differenziabile per $\tau > 3$

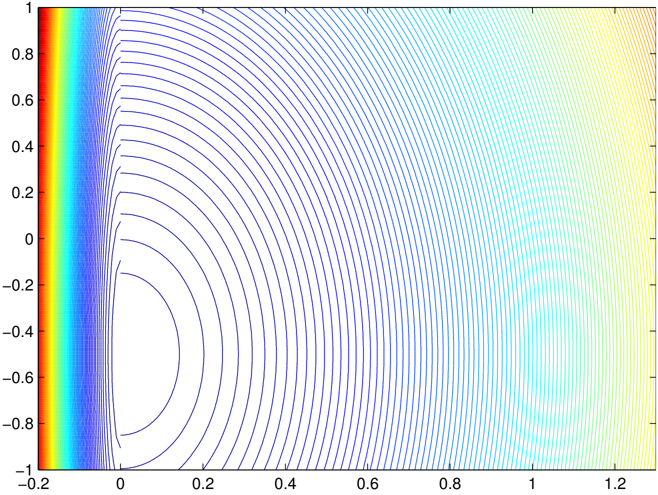
La funzione di McKinnon

$$f(x,y) = \begin{cases} \theta\phi|x|^\tau + y + y^2 & \text{se } x \leq 0 \\ \theta x^\tau + y + y^2 & \text{se } x > 0 \end{cases}$$

- strettamente convessa per $\tau > 1$
- continuamente differenziabile per $\tau > 1$
- due volte cont. differenziabile per $\tau > 2$
- tre volte cont. differenziabile per $\tau > 3$

La funzione di McKinnon

Per $\tau = 2$, $\theta = 6$, $\phi = 60$, la funzione è

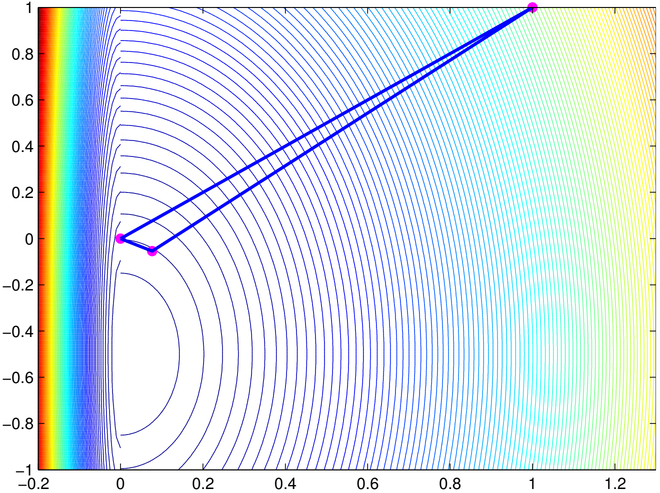


La funzione di McKinnon

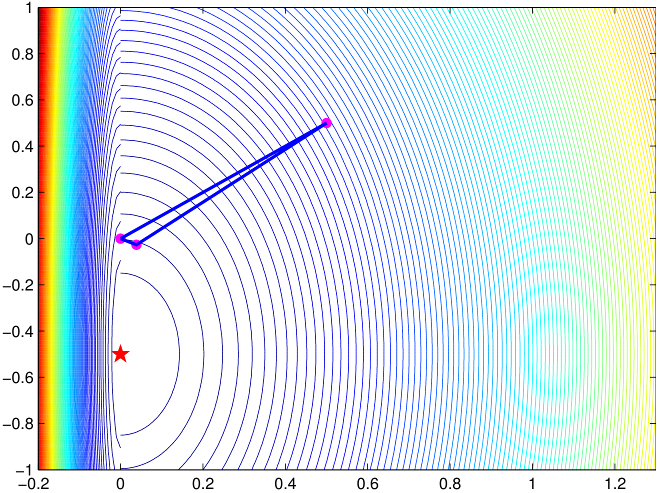
Se inizializziamo il metodo di Nelder-Mead con il simplesso

$$X = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} \frac{1+\sqrt{33}}{88} \\ \frac{1-\sqrt{33}}{88} \end{pmatrix}, \right\}$$

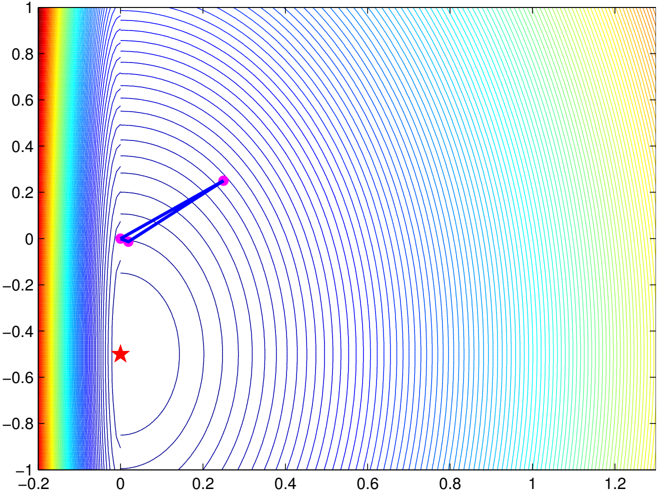
La funzione di McKinnon



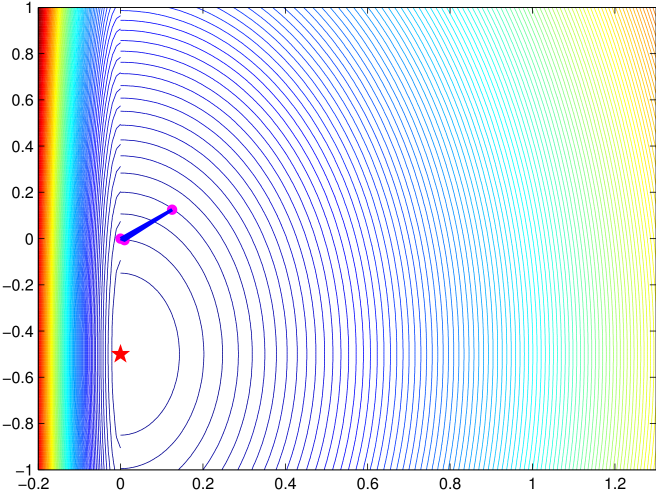
La funzione di McKinnon



La funzione di McKinnon



La funzione di McKinnon



Perchè...

- Compass Search "funziona" ?
- Hooke & Jeeves "funziona" ?
- Nelder & Mead "**non** funziona" (teoricamente) ?

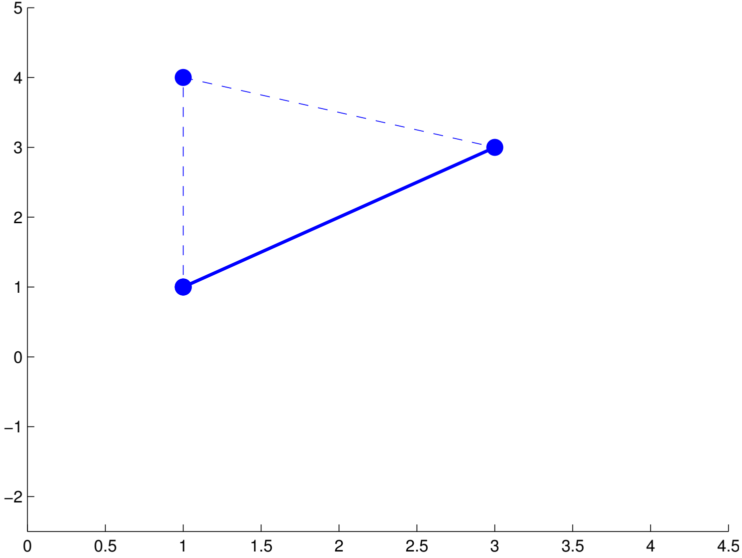
Cosa "davvero" distingue i primi due dal terzo ?

Perchè...

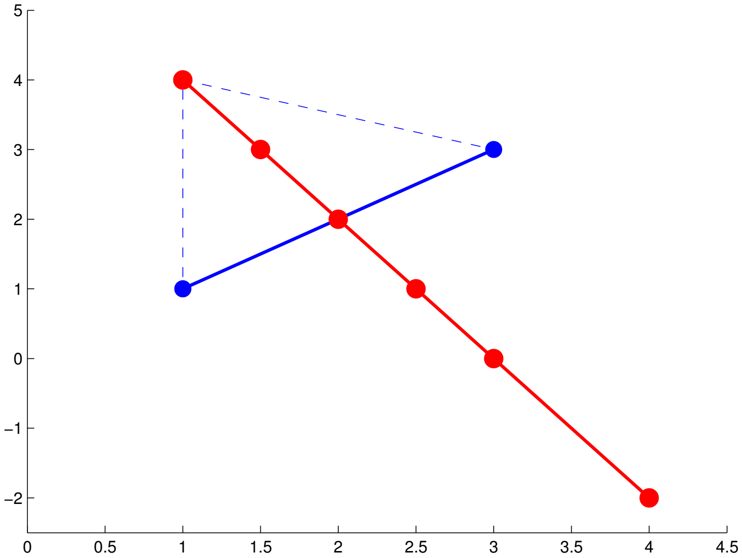
- Compass Search "funziona" ?
- Hooke & Jeeves "funziona" ?
- Nelder & Mead "**non** funziona" (teoricamente) ?

Cosa "davvero" distingue i primi due dal terzo ?

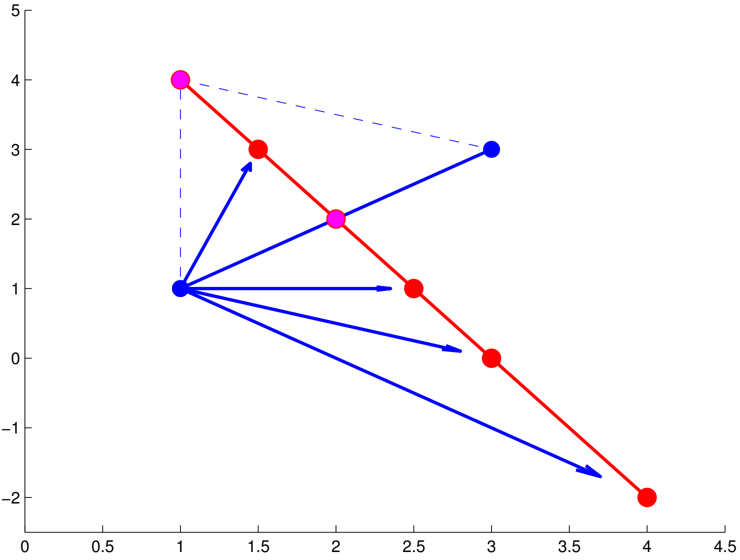
Perchè...



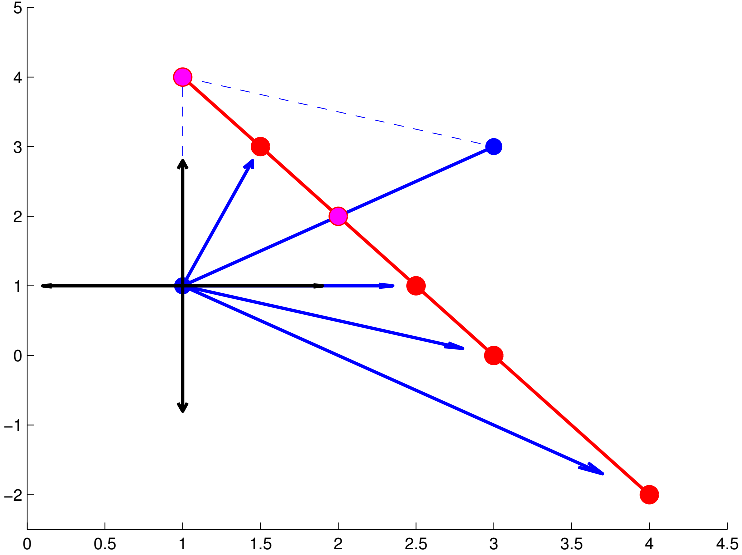
Perchè...



Perchè...



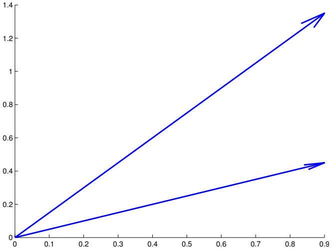
Perchè...



Angolo tra due vettori

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

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

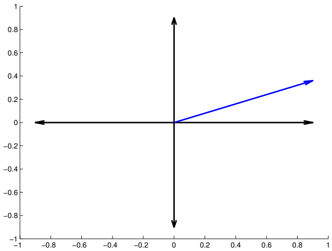


Angolo vettore – insieme di vettori

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

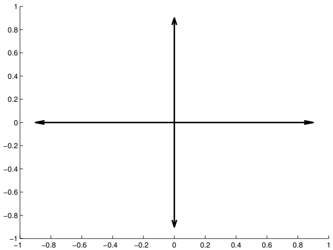
$$\kappa(D) = \min_{v \in \mathbb{R}^n} \phi(v) = \min_{v \in \mathbb{R}^n} \max_{d \in D} \frac{v^\top d}{\|v\|\|d\|}$$

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

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



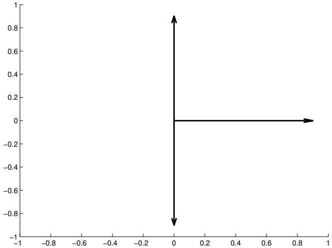
$$\kappa(D) = \frac{1}{\sqrt{2}}$$

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

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



$$\kappa(D) = 0$$

... e quindi ...

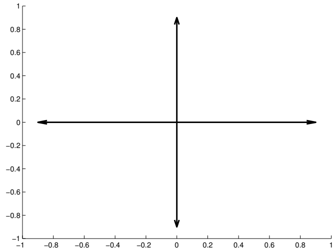
Quando $D = \{\pm e_1, \pm e_2, \dots, \pm e_n\}$

$$\kappa(D) = \frac{1}{\sqrt{n}} > 0$$

Comunque scelto un $v \in \mathbb{R}^n$, esiste $d \in D$ tale che

$$\frac{v^\top d}{\|v\|\|d\|} = \cos\theta(d, v) \geq \kappa(D) > 0$$

Quando $v = -\nabla f(x) \neq 0$, esiste sempre una direzione $d \in D$ **di discesa**



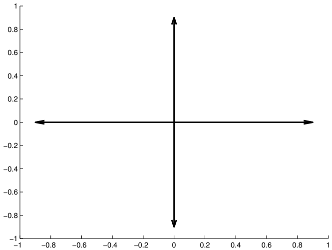
... e quindi ...

Quando $D = \{\pm e_1, \pm e_2, \ldots, \pm e_n\}$

$$\kappa(D) = \frac{1}{\sqrt{n}} > 0$$

Comunque scelto un $v \in \mathbb{R}^n$, esiste $d \in D$ tale che

$$\frac{v^\top d}{\|v\|\|d\|} = \cos\theta(d, v) \geq \kappa(D) > 0$$



Quando $v = -\nabla f(x) \neq 0$, esiste sempre una direzione $d \in D$ **di discesa**

... e quindi ...

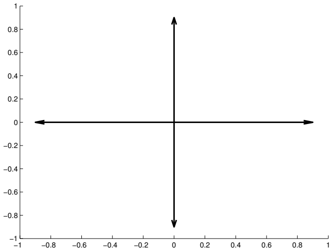
Quando $D = \{\pm e_1, \pm e_2, \ldots, \pm e_n\}$

$$\kappa(D) = \frac{1}{\sqrt{n}} > 0$$

Comunque scelto un $v \in \mathbb{R}^n$, esiste $d \in D$ tale che

$$\frac{v^\top d}{\|v\|\|d\|} = \cos\theta(d, v) \geq \kappa(D) > 0$$

Quando $v = -\nabla f(x) \neq 0$, esiste sempre una direzione $d \in D$ **di discesa**



... e quindi ...

Quando $D = \{\text{vedi figura}\}$

$$\kappa(D) \leq 0$$

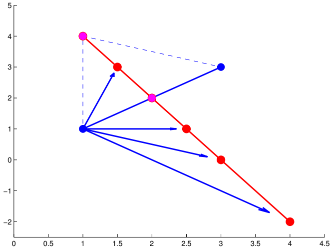
esistono $v \in \mathbb{R}^n$ tali che, per ogni $d \in D$, si ha

$$\frac{v^\top d}{\|v\|\|d\|} = \cos\theta(d, v) \leq 0$$

Potrebbe capitare che $\nabla f(x) \neq 0$, e ogni $d \in D$ è tale che

$$-\nabla f(x)^\top d \leq 0$$

cioè **nessuna** dir. in D è **di discesa**



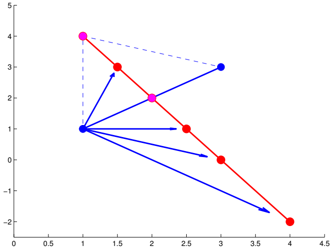
... e quindi ...

Quando $D = \{\text{vedi figura}\}$

$$\kappa(D) \leq 0$$

esistono $v \in \mathbb{R}^n$ tali che, per ogni $d \in D$, si ha

$$\frac{v^\top d}{\|v\|\|d\|} = \cos\theta(d, v) \leq 0$$



Potrebbe capitare che $\nabla f(x) \neq 0$, e ogni $d \in D$ è tale che

$$-\nabla f(x)^\top d \leq 0$$

cioè **nessuna** dir. in D è **di discesa**

... e quindi ...

Quando $D = \{\text{vedi figura}\}$

$$\kappa(D) \leq 0$$

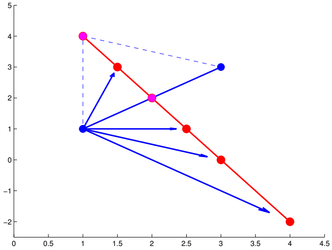
esistono $v \in \mathbb{R}^n$ tali che, per ogni $d \in D$, si ha

$$\frac{v^\top d}{\|v\|\|d\|} = \cos\theta(d, v) \leq 0$$

Potrebbe capitare che $\nabla f(x) \neq 0$, e ogni $d \in D$ è tale che

$$-\nabla f(x)^\top d \leq 0$$

cioè **nessuna** dir. in D è **di discesa**



... e quindi ...

per il metodo **compass search** che usa

- $D = \{\pm e_1, \dots, \pm e_n\}$,
- $\kappa(D) = 1/\sqrt{(n)} > 0$

quando il punto corrente **non** è stazionario $(\nabla f(x) \neq 0)$

- esiste sempre (almeno) una direzione $d \in D$ che è **di discesa**

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

- per passi Δ sufficientemente piccoli **deve accadere**

$$f(x + \Delta d) < f(x)$$

cioè il punto x corrente (se non-stazionario) prima o dopo viene aggiornato

... e quindi ...

per il metodo **compass search** che usa

- $D = \{\pm e_1, \dots, \pm e_n\}$,
- $\kappa(D) = 1/\sqrt(n) > 0$

quando il punto corrente **non** è stazionario ($\nabla f(x) \neq 0$)

- esiste sempre (almeno) una direzione $d \in D$ che è **di discesa**

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

- per passi Δ sufficientemente piccoli **deve accadere**

$$f(x + \Delta d) < f(x)$$

cioè il punto x corrente (se non-stazionario) prima o dopo viene aggiornato

... e quindi ...

per il metodo **compass search** che usa

- $D = \{\pm e_1, \ldots, \pm e_n\}$,
- $\kappa(D) = 1/\sqrt(n) > 0$

quando il punto corrente **non** è stazionario ($\nabla f(x) \neq 0$)

- esiste sempre (almeno) una direzione $d \in D$ che è **di discesa**

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

- per passi Δ sufficientemente piccoli **deve accadere**

$$f(x + \Delta d) < f(x)$$

cioè il punto x corrente (se non-stazionario) prima o dopo viene aggiornato

... e quindi ...

per il metodo **compass search** che usa

- $D = \{\pm e_1, \ldots, \pm e_n\}$,
- $\kappa(D) = 1/\sqrt(n) > 0$

quando il punto corrente **non** è stazionario ($\nabla f(x) \neq 0$)

- esiste sempre (almeno) una direzione $d \in D$ che è **di discesa**

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

- per passi Δ sufficientemente piccoli **deve accadere**

$$f(x + \Delta d) < f(x)$$

cioè il punto x corrente (se non-stazionario) prima o dopo viene aggiornato

... e quindi ...

per il metodo **Nelder&Mead** che usa

- D t.c. $\kappa(D) \leq 0$

anche se il punto corrente **non** è stazionario $(\nabla f(x) \neq 0)$

- le direzioni $d \in D$ potrebbero tendere a direzioni $\bar{d}$ che **non** sono di discesa

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

- anche per spostamenti piccolissimi (i.e. punti in X vicinissimi) nessuno dei punti x_r, x_e, x_{oc}, x_{ic} migliora il valore $f(x)$

cioè il punto x corrente non-stazionario non viene **mai** aggiornato!!

... e quindi ...

per il metodo **Nelder&Mead** che usa

- D t.c. $\kappa(D) \leq 0$

anche se il punto corrente **non** è stazionario $(\nabla f(x) \neq 0)$

- le direzioni $d \in D$ potrebbero tendere a direzioni $\bar{d}$ che **non** sono di discesa
$$\nabla f(x)^\top \bar{d} \geq 0$$

- anche per spostamenti piccolissimi (i.e. punti in X vicinissimi) nessuno dei punti x_r, x_e, x_{oc}, x_{ic} migliora il valore $f(x)$

cioè il punto x corrente non-stazionario non viene **mai** aggiornato!!

... e quindi ...

per il metodo **Nelder&Mead** che usa

- D t.c. $\kappa(D) \leq 0$

anche se il punto corrente **non** è stazionario ($\nabla f(x) \neq 0$)

- le direzioni $d \in D$ potrebbero tendere a direzioni $\bar{d}$ che **non** sono di discesa

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

- anche per spostamenti piccolissimi (i.e. punti in X vicinissimi) nessuno dei punti x_r, x_e, x_{oc}, x_{ic} migliora il valore $f(x)$

cioè il punto x corrente non-stazionario non viene **mai** aggiornato!!

... e quindi ...

per il metodo **Nelder&Mead** che usa

- D t.c. $\kappa(D) \leq 0$

anche se il punto corrente **non** è stazionario ($\nabla f(x) \neq 0$)

- le direzioni $d \in D$ potrebbero tendere a direzioni $\bar{d}$ che **non** sono di discesa
$$\nabla f(x)^\top \bar{d} \geq 0$$

- anche per spostamenti piccolissimi (i.e. punti in X vicinissimi) nessuno dei punti x_r, x_e, x_{oc}, x_{ic} migliora il valore $f(x)$

cioè il punto x corrente non-stazionario non viene **mai** aggiornato!!