Ottimizzazione dei Sistemi Complessi

G. Liuzzi[1]

Lunedì 24 Febbraio 2020

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

Chi sono io ?

Giampaolo Liuzzi

- studio: IASI (CNR), Via dei Taurini 19 (00185, Roma), V piano, stanza 514 (poco utile)
- tel: 06 49937129 (poco utile)
- email: giampaolo.liuzzi@iasi.cnr.it (utile)
- didattica: (utilissimo!!)

http://www.iasi.cnr.it/~liuzzi/teachita.htm

https://groups.google.com/forum/#!forum/osc_sapienza_2018-2019

Chi sono io ?

Giampaolo Liuzzi

- studio: IASI (CNR), Via dei Taurini 19 (00185, Roma), V piano, stanza 514 (**poco utile**)
- tel: 06 49937129 (**poco utile**)
- email: `giampaolo.liuzzi@iasi.cnr.it` (**utile**)
- didattica: (**utilissimo!!**)

  `http://www.iasi.cnr.it/~liuzzi/teachita.htm`

  `https://groups.google.com/forum/#!forum/osc_sapienza_2018-2019`

Chi siete Voi ?

- Studenti del 1° anno della Laurea Magistrale in **Ingegneria Gestionale** (? Altri?)
- che hanno già sostenuto (almeno) l'esame di **Ricerca Operativa** (12CFU, Lucidi/De Santis) (?)
- che hanno già sostenuto l'esame di **Ottimizzazione** (9CFU, Lucidi) (?)

Notizie utili (che corso è questo?)

Ottimizzazione dei Sistemi Complessi

- per Ingegneria **Gestionale** (MGER)
- Le lezioni si svolgono in via Aristo 25 (DIAG) secondo il seguente orario:
  - **Lunedì** 9:00-12:00, aula A2
  - **Giovedì** 8:00-10:00, aula A4
- Ricevimento: **Giovedì** 10:00-11:00 presso **stanza Prof. Lucidi**
- Materiale didattico:
  - slides delle lezioni
  - dispense
  - `http://www.iasi.cnr.it/~liuzzi/teachita.htm`
  - `http://www.dis.uniroma1.it/~or/gestionale/osc/`

Notizie utili (che corso è questo?)

Ottimizzazione dei Sistemi Complessi

- Modalità d'esame:
  - prova scritta
  - negli appelli d'esame di Giugno, Luglio e Settembre è possibile sostituire un esercizio a scelta con lo svolgimento di una implementazione. Però:
    - lo devo sapere prima della fine del corso
    - l'implementazione va discussa prima di sostenere l'esame

Sistemi "complessi"?

$$\min \; f(x)$$
$$s.t. \; x \in \mathcal{F}$$

con, p.es., $\mathcal{F} = \{x \in \mathbb{R}^n : \; g(x) \leq 0\}$.

Cosa rende (o potrebbe rendere) complesso un problema di ottimo?

- complessità nelle funzioni $f(x)$, $g(x)$
  - derivate prime, seconde, ... non disponibili
  - non linearità nelle espressioni di $g(x)$
- complessità nelle variabili
  - variabili $x = x(t)$ sono "traiettorie" cioè funzioni del tempo
- complessità per la presenza di due o più funzioni obiettivo
  - $f(x) = (f_1(x), \ldots, f_q(x))^\top$

Sistemi "complessi"?

$$\begin{aligned} &\min \ f(x) \\ &s.t.\ x \in \mathcal{F} \end{aligned}$$

con, p.es., $\mathcal{F} = \{x \in \mathbb{R}^n : \ g(x) \leq 0\}$.

Cosa rende (o potrebbe rendere) complesso un problema di ottimo?

- complessità nelle funzioni $f(x)$, $g(x)$
  - derivate prime, seconde, ... non disponibili
  - non linearità nelle espressioni di $g(x)$
- complessità nelle variabili
  - variabili $x = x(t)$ sono "traiettorie" cioè funzioni del tempo
- complessità per la presenza di due o più funzioni obiettivo
  - $f(x) = (f_1(x), \dots, f_q(x))^\top$

Sistemi "complessi"?

$$\min \; f(x)$$
$$s.t. \; x \in \mathcal{F}$$

con, p.es., $\mathcal{F} = \{x \in \mathbb{R}^n : \; g(x) \leq 0\}$.

Cosa rende (o potrebbe rendere) complesso un problema di ottimo?

- complessità nelle funzioni $f(x)$, $g(x)$
  - derivate prime, seconde, ... non disponibili
  - non linearità nelle espressioni di $g(x)$
- complessità nelle variabili
  - variabili $x = x(t)$ sono "traiettorie" cioè funzioni del tempo
- complessità per la presenza di due o più funzioni obiettivo
  - $f(x) = (f_1(x), \ldots, f_q(x))^\top$

Sistemi "complessi"?

$$\min \ f(x)$$
$$s.t. \ x \in \mathcal{F}$$

con, p.es., $\mathcal{F} = \{x \in \mathbb{R}^n : \ g(x) \leq 0\}$.

Cosa rende (o potrebbe rendere) complesso un problema di ottimo?

- complessità nelle funzioni $f(x)$, $g(x)$
  - derivate prime, seconde, ... non disponibili
  - non linearità nelle espressioni di $g(x)$
- complessità nelle variabili
  - variabili $x = x(t)$ sono "traiettorie" cioè funzioni del tempo
- complessità per la presenza di due o più funzioni obiettivo
  - $f(x) = (f_1(x), \ldots, f_q(x))^\top$

Cosa faremo

- Ottimizzazione **senza** l'uso delle **derivate**
  - Esercitazioni in Julia
- Ottimizzazione per problemi con **vincoli non lineari**
  - Esercitazioni in Julia
- Ottimizzazione delle **traiettorie**
  - Esercitazioni in Julia ??
- Ottimizzazione per problemi con **più funzioni obiettivo**
  - Esercitazioni in Julia

Cosa faremo

- Ottimizzazione **senza** l'uso delle **derivate**
  - Esercitazioni in Julia
- Ottimizzazione per problemi con **vincoli non lineari**
  - Esercitazioni in Julia
- Ottimizzazione delle **traiettorie**
  - Esercitazioni in Julia ??
- Ottimizzazione per problemi con **più funzioni obiettivo**
  - Esercitazioni in Julia

Materiale didattico

- Ottimizzazione **senza** l'uso delle **derivate**
  - Dispense e trasparenze delle lezioni (`www.iasi.cnr.it/~liuzzi/teachita.htm`)
- Ottimizzazione per problemi con **vincoli non lineari**
  - `www.dis.uniroma1.it/~or/gestionale/osc/`
- Ottimizzazione delle **traiettorie**
  - `www.dis.uniroma1.it/~or/gestionale/osc/`
- Ottimizzazione per problemi con **più funzioni obiettivo**
  - Dispense e trasparenze delle lezioni (`www.iasi.cnr.it/~liuzzi/teachita.htm`)

Materiale didattico

- Ottimizzazione **senza** l'uso delle **derivate**
  - Dispense e trasparenze delle lezioni (`www.iasi.cnr.it/~liuzzi/teachita.htm`)
- Ottimizzazione per problemi con **vincoli non lineari**
  - `www.dis.uniroma1.it/~or/gestionale/osc/`
- Ottimizzazione delle **traiettorie**
  - `www.dis.uniroma1.it/~or/gestionale/osc/`
- Ottimizzazione per problemi con **più funzioni obiettivo**
  - Dispense e trasparenze delle lezioni (`www.iasi.cnr.it/~liuzzi/teachita.htm`)

Una “strana” funzione

Consideriamo il problema

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

ove però:

- non è nota l’espressione analitica di $f(x)$
- non sono disponibili $\nabla f(x)$, $\nabla^2 f(x)$, ...

Esempi

Visitando la URL

- `http://www.iasi.cnr.it/~liuzzi/OSC2018/testn.php?x=1,3,5,2`

  è possibile ottenere il valore di $f(x)$ quando $x = (1, 3, 5, 2)^\top$

- `http://www.iasi.cnr.it/~liuzzi/OSC2018/testn.php?x=1,1,-2,1,0.5`

  è possibile ottenere il valore di $f(x)$ quando $x = (1, 1, -2, 1, 0.5)^\top$

- ... e così via

Il problema

Stabilito un valore per il numero n di variabili, il problema è:

ridurre il più possibile $f(x)$ quando

$$-10 \leq x_i \leq 10, \quad \text{per ogni } i = 1, 2, \ldots, n$$

ovvero, risolvere

$$\begin{array}{l} \min \ f(x) \\ s.t. \ -10 \leq x_i \leq 10, \quad \text{per ogni } i = 1, 2, \ldots, n \end{array}$$

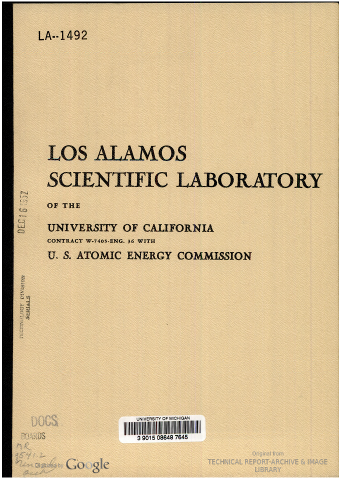
In particolare

Quando $n = 30$, partendo dal punto

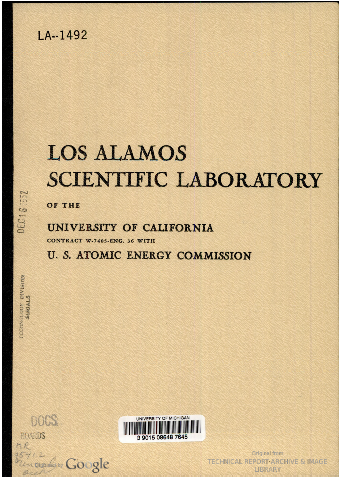
$$x_0 = \begin{pmatrix} 8.1 & 10.0 & -8.4 & -8.1 & -5.6 & -7.5 & \dots \\ -9.0 & -10.0 & -3.5 & 10.0 & 2.2 & 1.0 & \dots \\ -6.0 & 10.0 & -8.4 & 3.9 & 1.9 & 9.3 & \dots \\ 10.0 & 6.7 & -0.8 & -10.0 & -3.7 & 6.7 & \dots \\ -10.0 & -5.8 & -10.0 & -8.6 & 2.6 & -9.2 & \end{pmatrix}^\top$$

migliorare $f(x_0) = 1498.0922$ il più possibile.

Un problema di minimo (nella fisica delle particelle)

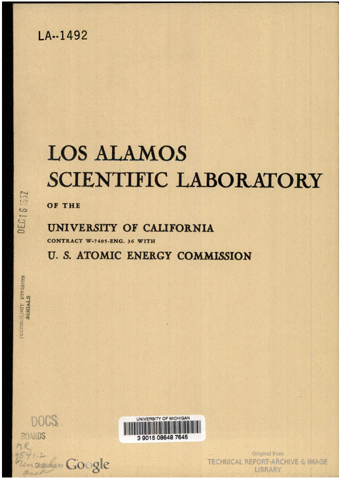


Un problema di minimo (nella fisica delle particelle)


LA--1492

LOS ALAMOS
SCIENTIFIC LABORATORY
OF THE
UNIVERSITY OF CALIFORNIA
CONTRACT W-7405-ENG. 36 WITH
U. S. ATOMIC ENERGY COMMISSION

Un problema di minimo (nella fisica delle particelle)


LA--1492

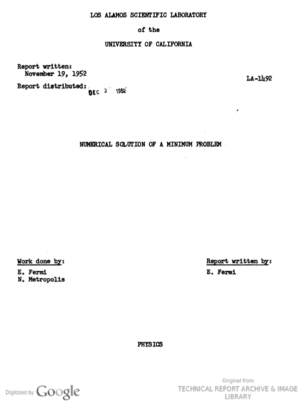
LOS ALAMOS SCIENTIFIC LABORATORY

OF THE

UNIVERSITY OF CALIFORNIA

CONTRACT W-7405-ENG. 36 WITH

U. S. ATOMIC ENERGY COMMISSION


LOS ALAMOS SCIENTIFIC LABORATORY

of the

UNIVERSITY OF CALIFORNIA

Report written:
November 19, 1952

LA-1492

Report distributed: DEC 3 1952

NUMERICAL SOLUTION OF A MINIMUM PROBLEM

Work done by:
E. Fermi
N. Metropolis

Report written by:
E. Fermi

PHYSICS

Un problema di minimo (nella fisica delle particelle)

Abstract

"A particular non-linear function of six independent variables is minimized, using the Los Alamos electronic computer. The values of the variables at the minimum correspond to the phase shift angles in the scattering of pions by hydrogen"

mesone = particella composta da quark + antiquark

pione = mesone π (il più leggero dei mesoni)

Ci sono tre tipi di pione distinti per la loro carica elettrica:

$$\pi^+, \quad \pi^0, \quad \pi^-$$

Dispersione (scattering) $\pi - H$

Due livelli di energia: 113 MeV e 135 MeV

Tre processi di dispersione (scattering) per collisione $\pi - H$

- $\pi^- + p \to \pi^- + p$
- $\pi^+ + p \to \pi^+ + p$
- $\pi^- + p \to \pi^0 + n \to \gamma\gamma + n$

I risultati degli esperimenti (cross sections) sono acquisiti tramite rivelatori posti ad angoli di: 45°, 90° e 135° attorno alla camera a idrogeno (dove avviene la dispersione).

- $\sigma_-^1, \sigma_-^2, \sigma_-^3$
- $\sigma_+^1, \sigma_+^2, \sigma_+^3$
- $\sigma_\gamma^1, \sigma_\gamma^2, \sigma_\gamma^3$

Stima dei parametri

La teoria della dispersione vuole che le **cross sections** $(\sigma_1, \ldots, \sigma_9)$ siano funzione di certi parametri teorici incogniti (**phase shifts** $\alpha_1, \ldots, \alpha_6$).

$$\begin{aligned} \sigma_1 &= f_1(\alpha_1, \ldots, \alpha_6) \\ &\vdots \\ \sigma_9 &= f_9(\alpha_1, \ldots, \alpha_6) \end{aligned}$$

$$\sigma = F(\alpha)$$

Stima dei parametri

La teoria della dispersione vuole che le **cross sections** $(\sigma_1, \ldots, \sigma_9)$ siano funzione di certi parametri teorici incogniti (**phase shifts** $\alpha_1, \ldots, \alpha_6$).

$$
\begin{aligned}
\sigma_1 &= f_1(\alpha_1, \ldots, \alpha_6) \\
\vdots \quad & \quad \vdots \\
\sigma_9 &= f_9(\alpha_1, \ldots, \alpha_6)
\end{aligned}
$$

$$\sigma = F(\alpha)$$

Stima dei parametri

La teoria della dispersione vuole che le **cross sections** $(\sigma_1, \ldots, \sigma_9)$ siano funzione di certi parametri teorici incogniti (**phase shifts** $\alpha_1, \ldots, \alpha_6$).

$$
\begin{array}{rcl}
\sigma_1 & = & f_1(\alpha_1, \ldots, \alpha_6) \\
\vdots & & \vdots \\
\sigma_9 & = & f_9(\alpha_1, \ldots, \alpha_6)
\end{array}
$$

$$\sigma = F(\alpha)$$

Più in dettaglio (1)

In particolare, se indichiamo

$$e_j = e^{2\alpha_j i} - 1, \qquad j = 1, \dots, 6$$

possiamo definire i seguenti coefficienti

$$b_p = \frac{e_1 + 2e_2}{3};\;\; a_{\beta p} = \frac{\sqrt{2}}{9}(e_3 - e_4 + 2e_5 - 2e_6);\;\; a_{\alpha p} = \frac{2e_3 + e_4 + 4e_5 + 2e_6}{9}$$

$$b_N = \frac{\sqrt{2}}{3}(e_1 - e_2);\;\; a_{\beta N} = \frac{2}{9}(e_3 - e_4 - e_5 + e_6);\;\; a_{\alpha N} = \frac{\sqrt{2}}{9}(2e_3 + e_4 - 2e_5 - e_6)$$

$$b = e_1;\;\; a_\beta = \frac{\sqrt{2}}{3}(e_3 - e_4);\;\; a_\alpha = \frac{2e_3 + e_4}{3}$$

Più in dettaglio (1)

In particolare, se indichiamo

$$e_j = e^{2\alpha_j i} - 1, \qquad j = 1, \dots, 6$$

possiamo definire i seguenti coefficienti

$$b_p = \frac{e_1 + 2e_2}{3};\ a_{\beta p} = \frac{\sqrt{2}}{9}(e_3 - e_4 + 2e_5 - 2e_6);\ a_{\alpha p} = \frac{2e_3 + e_4 + 4e_5 + 2e_6}{9}$$

$$b_N = \frac{\sqrt{2}}{3}(e_1 - e_2);\ a_{\beta N} = \frac{2}{9}(e_3 - e_4 - e_5 + e_6);\ a_{\alpha N} = \frac{\sqrt{2}}{9}(2e_3 + e_4 - 2e_5 - e_6)$$

$$b = e_1;\ a_\beta = \frac{\sqrt{2}}{3}(e_3 - e_4);\ a_\alpha = \frac{2e_3 + e_4}{3}$$

Più in dettaglio (2)

e quindi

$$A_- = \frac{2|b_p|^2 + 9|a_{\beta p}|^2}{8};\ B_- = \frac{3}{2}R(b_p a^*_{\alpha p});\ C_- = \frac{18|a_{\alpha p}|^2 - 9|a_{\beta p}|^2}{8}$$
$$A_0 = \frac{2|b_N|^2 + 9|a_{\beta N}|^2}{8};\ B_0 = \frac{3}{2}R(b_N a^*_{\alpha N});\ C_0 = \frac{18|a_{\alpha N}|^2 - 9|a_{\beta N}|^2}{8}$$
$$A_+ = \frac{2|b|^2 + 9|a_{\beta}|^2}{8};\ B_0 = \frac{3}{2}R(b a^*_{\alpha});\ C_+ = \frac{18|a_{\alpha}|^2 - 9|a_{\beta}|^2}{8}$$

Stima dei parametri

per cui otteniamo:

$$
\begin{aligned}
\sigma_1 &= A_- + B_- \cos\chi_1 + C_- \cos^2\chi_1 \\
\sigma_2 &= A_- + B_- \cos\chi_2 + C_- \cos^2\chi_2 \\
\sigma_3 &= A_- + B_- \cos\chi_3 + C_- \cos^2\chi_3 \\
\sigma_4 &= A_+ + B_+ \cos\chi_1 + C_+ \cos^2\chi_1 \\
\sigma_5 &= A_+ + B_+ \cos\chi_2 + C_+ \cos^2\chi_2 \\
\sigma_6 &= A_+ + B_+ \cos\chi_3 + C_+ \cos^2\chi_3 \\
\sigma_7 &= 2A_o + \frac{2-q}{3}C_o + pB_o \cos\chi_1' + qC_o \cos^2\chi_1' \\
\sigma_8 &= 2A_o + \frac{2-q}{3}C_o + pB_o \cos\chi_2' + qC_o \cos^2\chi_2' \\
\sigma_9 &= 2A_o + \frac{2-q}{3}C_o + pB_o \cos\chi_3' + qC_o \cos^2\chi_3'
\end{aligned}
$$

Stima dei parametri

Quindi, possiamo scrivere il sistema non-lineare seguente

$$\sigma = F(\alpha),$$

definire la funzione di errore:

$$M(\alpha) = \sum_{i=1}^{9} \left(\frac{\sigma_i - F_i(\alpha)}{\epsilon_i} \right)^2$$

risolvere il problema

$$\min_{\alpha} M(\alpha)$$

Stima dei parametri

Quindi, possiamo scrivere il sistema non-lineare seguente

$$\sigma = F(\alpha),$$

definire la funzione di errore:

$$M(\alpha) = \sum_{i=1}^{9} \left(\frac{\sigma_i - F_i(\alpha)}{\epsilon_i} \right)^2$$

risolvere il problema

$$\min_{\alpha} M(\alpha)$$

Stima dei parametri

Quindi, possiamo scrivere il sistema non-lineare seguente

$$\sigma = F(\alpha),$$

definire la funzione di errore:

$$M(\alpha) = \sum_{i=1}^{9} \left(\frac{\sigma_i - F_i(\alpha)}{\epsilon_i} \right)^2$$

risolvere il problema

$$\min_{\alpha} M(\alpha)$$

Soluzione del problema

Dal rapporto LA-1492, pp. 12-13

"The problem that has been here discussed is an example of a minimum problem for a function of many variables."

"In principle, problems of this type could be handled in two ways."

"One involves standard mathematical procedure of equating to 0 all the partial derivatives of the function and obtaining thereby a system of n equations with n unknowns."

"The second procedure is the one chosen in the present example: to search for the minimum value by computing the function at very many points until the minimum is attained"

Soluzione del problema

Dal rapporto LA-1492, pp. 12-13

"The problem that has been here discussed is an example of a minimum problem for a function of many variables."

"In principle, problems of this type could be handled in two ways."

"One involves standard mathematical procedure of equating to 0 all the partial derivatives of the function and obtaining thereby a system of n equations with n unknowns."

"The second procedure is the one chosen in the present example: to search for the minimum value by computing the function at very many points until the minimum is attained"

Soluzione del problema

Dal rapporto LA-1492, pp. 12-13

"The problem that has been here discussed is an example of a minimum problem for a function of many variables."

"In principle, problems of this type could be handled in two ways."

"One involves standard mathematical procedure of equating to 0 all the partial derivatives of the function and obtaining thereby a system of n equations with n unknowns."

"The second procedure is the one chosen in the present example: to search for the minimum value by computing the function at very many points until the minimum is attained"

Soluzione del problema

Dal rapporto LA-1492, pp. 12-13

"The problem that has been here discussed is an example of a minimum problem for a function of many variables."

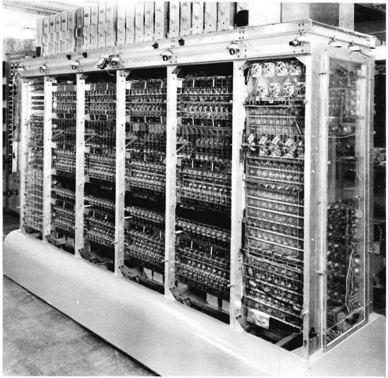
"In principle, problems of this type could be handled in two ways."

"One involves standard mathematical procedure of equating to 0 all the partial derivatives of the function and obtaining thereby a system of n equations with n unknowns."

"The second procedure is the one chosen in the present example: to search for the minimum value by computing the function at very many points until the minimum is attained"

MANIAC I

Fermi e Metropolis usarono il computer MANIAC (**M**athematical **A**nalyzer, **N**umerical **I**ntegrator, **A**nd **C**omputer, 1952-1958) del Los Alamos Laboratory per calcolare la funzione $M(x)$



MANIAC I

"The coding of this part of the problem $[M(x + \Delta x)]$ requires approximately 150 memory positions ... the machine computes its value in approximately **4/10 of a second**, whereas a hand computation of the same function takes about **20 minutes**"

Algoritmo "Fermi-Metropolis"

1. calcola il valore di $M(\alpha)$ per gli angoli iniziali
2. aumenta α_1 con passi di $1/2°$ ($\alpha_1 + 1/2°$, $\alpha_1 + 1°$ ...) finché il valore di M diminuisce
3. se $\alpha_1 + 1/2°$ produce un aumento di M, diminuisci α_1 con passi di $-1/2°$ finché M diminuisce
4. ripeti i passi 2 e 3 con $\alpha_2, \alpha_3, \ldots, \alpha_6$ al posto di α_1
5. ripeti 2, 3 e 4 finché per due cicli consecutivi il valore di M non si riduce

Al termine, ripeti lo stesso procedimento con passi di $\pm 1/16°$

Algoritmo "Fermi-Metropolis"

1. calcola il valore di $M(\alpha)$ per gli angoli iniziali
2. aumenta α_1 con passi di $1/2^\circ$ $(\alpha_1 + 1/2^\circ, \alpha_1 + 1^\circ \ldots)$ finché il valore di M diminuisce
3. se $\alpha_1 + 1/2^\circ$ produce un aumento di M, diminuisci α_1 con passi di $-1/2^\circ$ finché M diminuisce
4. ripeti i passi 2 e 3 con $\alpha_2, \alpha_3, \ldots, \alpha_6$ al posto di α_1
5. ripeti 2, 3 e 4 finché per due cicli consecutivi il valore di M non si riduce

Al termine, ripeti lo stesso procedimento con passi di $\pm 1/16^\circ$

Algoritmo "Fermi-Metropolis"

1. calcola il valore di $M(\alpha)$ per gli angoli iniziali
2. aumenta α_1 con passi di $1/2°$ $(\alpha_1 + 1/2°,\ \alpha_1 + 1° \ldots)$ finché il valore di M diminuisce
3. se $\alpha_1 + 1/2°$ produce un aumento di M, diminuisci α_1 con passi di $-1/2°$ finché M diminuisce
4. ripeti i passi 2 e 3 con $\alpha_2, \alpha_3, \ldots, \alpha_6$ al posto di α_1
5. ripeti 2, 3 e 4 finché per due cicli consecutivi il valore di M non si riduce

Al termine, ripeti lo stesso procedimento con passi di $\pm 1/16°$

Algoritmo "Fermi-Metropolis"

1. calcola il valore di $M(\alpha)$ per gli angoli iniziali
2. aumenta α_1 con passi di $1/2°$ $(\alpha_1 + 1/2°,\ \alpha_1 + 1° \ \ldots)$ finché il valore di M diminuisce
3. se $\alpha_1 + 1/2°$ produce un aumento di M, diminuisci α_1 con passi di $-1/2°$ finché M diminuisce
4. ripeti i passi 2 e 3 con $\alpha_2, \alpha_3, \ldots, \alpha_6$ al posto di α_1
5. ripeti 2, 3 e 4 finché per due cicli consecutivi il valore di M non si riduce

Al termine, ripeti lo stesso procedimento con passi di $\pm 1/16°$

Algoritmo "Fermi-Metropolis"

1. calcola il valore di $M(\alpha)$ per gli angoli iniziali
2. aumenta α_1 con passi di $1/2°$ ($\alpha_1 + 1/2°$, $\alpha_1 + 1°$ ...) finché il valore di M diminuisce
3. se $\alpha_1 + 1/2°$ produce un aumento di M, diminuisci α_1 con passi di $-1/2°$ finché M diminuisce
4. ripeti i passi 2 e 3 con $\alpha_2, \alpha_3, \dots, \alpha_6$ al posto di α_1
5. ripeti 2, 3 e 4 finché per due cicli consecutivi il valore di M non si riduce

Al termine, ripeti lo stesso procedimento con passi di $\pm 1/16°$

Il problema

Obiettivo: minimizzare una funzione $f(x)$ di n variabili reali per la quale non è data una procedura che restituisce le derivate di f in un punto x

- Metodi che usano il gradiente (naturalmente **approssimato**)
- Metodi che **NON** usano approssimazioni del gradiente

Il problema

Obiettivo: minimizzare una funzione $f(x)$ di n variabili reali per la quale non è data una procedura che restituisce le derivate di f in un punto x

- Metodi che usano il gradiente (naturalmente **approssimato**)
- Metodi che **NON** usano approssimazioni del gradiente

Il problema

Obiettivo: minimizzare una funzione $f(x)$ di n variabili reali per la quale non è data una procedura che restituisce le derivate di f in un punto x

- Metodi che usano il gradiente (naturalmente **approssimato**)
- Metodi che **NON** usano approssimazioni del gradiente

Formule di approssimazione delle derivate

- differenze finite in avanti (FFD):

$$\frac{\partial f}{\partial x_i} \simeq \frac{f(x+\epsilon e_i) - f(x)}{\epsilon} \quad \text{con } e_i \ i\text{-esima colonna di } I$$

- differenze finite all'indietro (BFD):

$$\frac{\partial f}{\partial x_i} \simeq \frac{f(x) - f(x-\epsilon e_i)}{\epsilon}$$

- differenze finite centrali (CFD):

$$\frac{\partial f}{\partial x_i} \simeq \frac{f(x+\epsilon e_i) - f(x-\epsilon e_i)}{2\epsilon}$$

Formule di approssimazione delle derivate

- differenze finite in avanti (FFD):

$$\frac{\partial f}{\partial x_i} \simeq \frac{f(x+\epsilon e_i) - f(x)}{\epsilon} \quad \text{con } e_i \ i\text{-esima colonna di } I$$

- differenze finite all'indietro (BFD):

$$\frac{\partial f}{\partial x_i} \simeq \frac{f(x) - f(x-\epsilon e_i)}{\epsilon}$$

- differenze finite centrali (CFD):

$$\frac{\partial f}{\partial x_i} \simeq \frac{f(x+\epsilon e_i) - f(x-\epsilon e_i)}{2\epsilon}$$

Formule di approssimazione delle derivate

- differenze finite in avanti (FFD):

$$\frac{\partial f}{\partial x_i} \simeq \frac{f(x+\epsilon e_i) - f(x)}{\epsilon} \quad \text{con } e_i \ i\text{-esima colonna di } I$$

- differenze finite all'indietro (BFD):

$$\frac{\partial f}{\partial x_i} \simeq \frac{f(x) - f(x-\epsilon e_i)}{\epsilon}$$

- differenze finite centrali (CFD):

$$\frac{\partial f}{\partial x_i} \simeq \frac{f(x+\epsilon e_i) - f(x-\epsilon e_i)}{2\epsilon}$$

Metodo che usa il gradiente approssimato

- Calcolo $d = -\nabla_\epsilon f(x)$
- Determino un passo α t.c.

$$\begin{aligned} f(x+\alpha d) &\leq f(x) + \gamma\alpha\nabla_\epsilon f(x)^\top d \quad \text{(suff. riduzione)} \\ f(x+\delta\alpha d) &> f(x) + \gamma\delta\alpha\nabla_\epsilon f(x)^\top d \quad \text{(suff. spostamento)} \end{aligned}$$

  con $\gamma \in (0, 1/2)$ e $\delta > 1$
- Pongo $x \leftarrow x + \alpha d$

Metodo che usa il gradiente approssimato

- Calcolo $d = -\nabla_\epsilon f(x)$
- Determino un passo α t.c.

$$\begin{aligned} f(x+\alpha d) &\le f(x) + \gamma\alpha\nabla_\epsilon f(x)^\top d \quad \text{(suff. riduzione)} \\ f(x+\delta\alpha d) &> f(x) + \gamma\delta\alpha\nabla_\epsilon f(x)^\top d \quad \text{(suff. spostamento)} \end{aligned}$$

con $\gamma \in (0, 1/2)$ e $\delta > 1$

- Pongo $x \leftarrow x + \alpha d$

Metodo che usa il gradiente approssimato

- Calcolo $d = -\nabla_\epsilon f(x)$
- Determino un passo α t.c.

$$
\begin{aligned}
f(x+\alpha d) &\leq f(x) + \gamma\alpha\nabla_\epsilon f(x)^\top d \quad \text{(suff. riduzione)} \\
f(x+\delta\alpha d) &> f(x) + \gamma\delta\alpha\nabla_\epsilon f(x)^\top d \quad \text{(suff. spostamento)}
\end{aligned}
$$

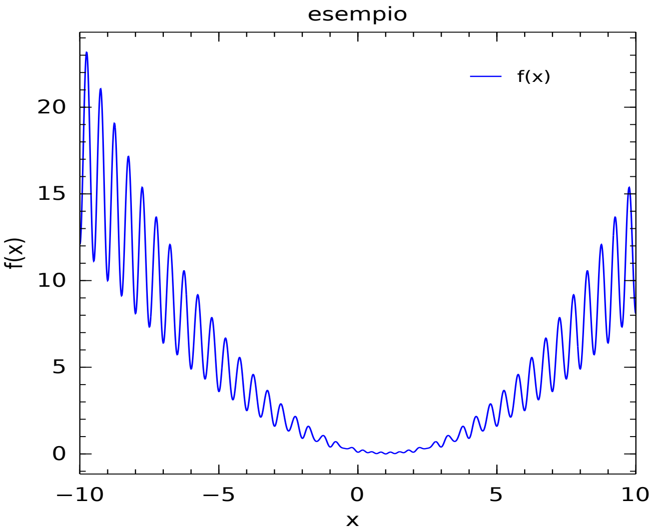
  con $\gamma \in (0, 1/2)$ e $\delta > 1$

- Pongo $x \leftarrow x + \alpha d$

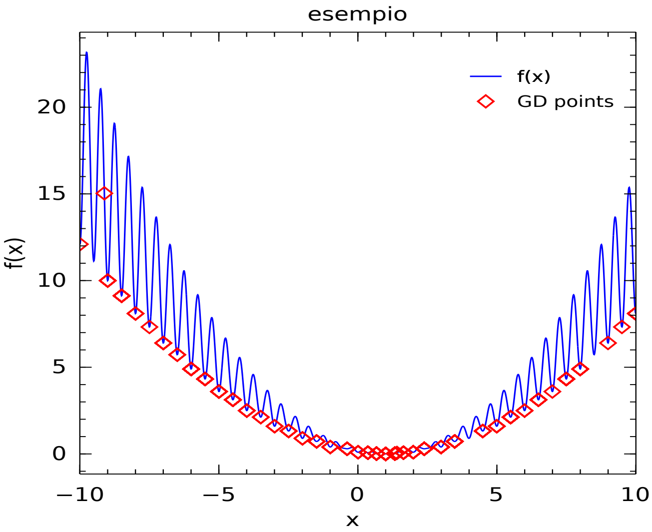
Esempio

Confronto su un problema in Julia

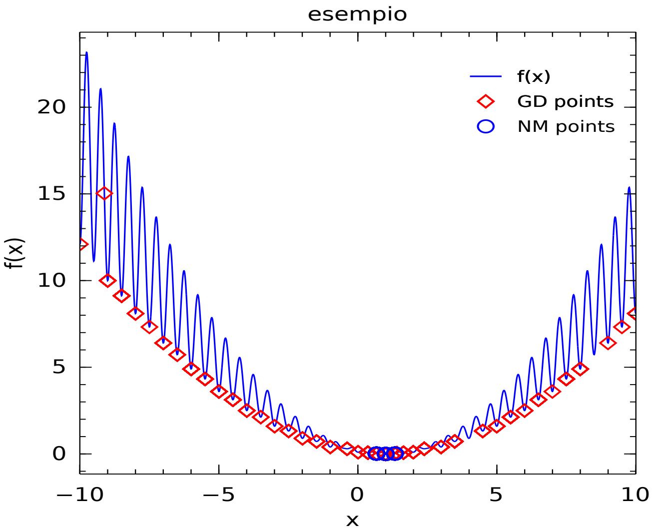
Esempio, molti minimi locali



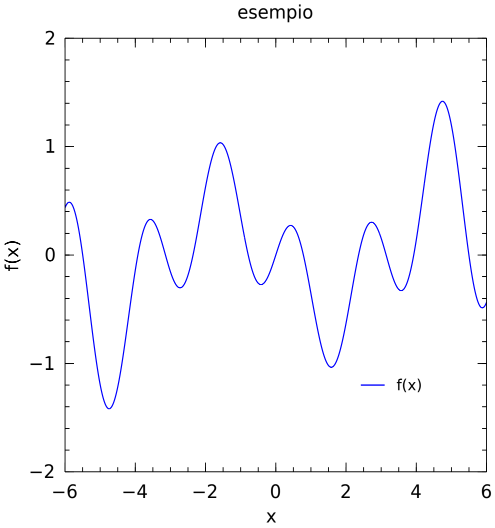
Esempio, molti minimi locali



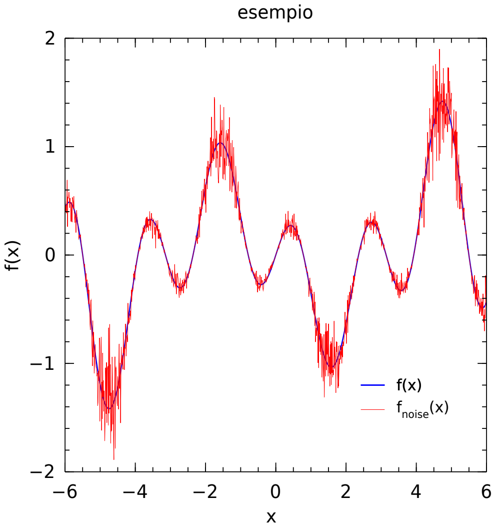
Esempio, molti minimi locali



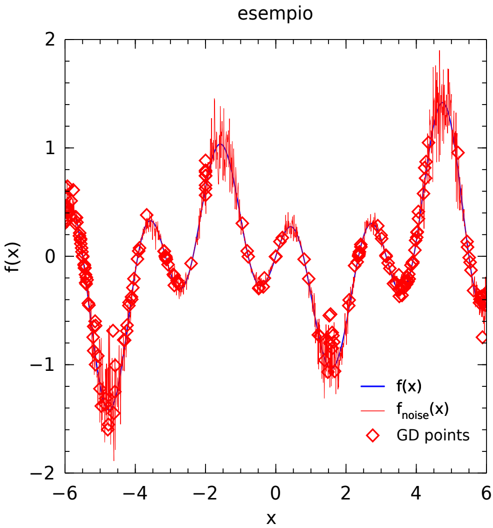
Esempio



Esempio

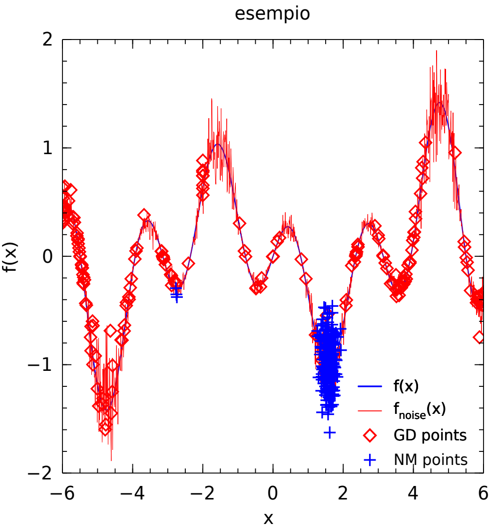


Esempio



200 minimizzazioni a partire da 200 p.ti iniziali random

Esempio



200 minimizzazioni a partire da 200 p.ti iniziali random

Perché?

Supponiamo $f(x)$ sia affetta da rumore additivo quindi

$$\tilde{f}(x) = f(x) + \epsilon$$

Differenze finite su $\tilde{f}$:

$$\begin{aligned} \frac{\partial \tilde{f}}{\partial x_i}(x) &\cong \frac{\tilde{f}(x+\Delta e_i) - \tilde{f}(x)}{\Delta} = \frac{f(x+\Delta e_i) + \epsilon_1 - f(x) - \epsilon_2}{\Delta} \\ &\cong \frac{\partial f}{\partial x_i}(x) + \frac{\epsilon_1 - \epsilon_2}{\Delta} \end{aligned}$$

Problema: per avere una buona approssimazione devo usare $\Delta \ll 1$ **ma** in questo caso

$$\frac{\epsilon_1 - \epsilon_2}{\Delta} \gg 1!!!$$

Perché?

Supponiamo $f(x)$ sia affetta da rumore additivo quindi

$$\tilde{f}(x) = f(x) + \epsilon$$

Differenze finite su $\tilde{f}$:

$$\begin{aligned}\frac{\partial \tilde{f}}{\partial x_i}(x) \quad &\cong \quad \frac{\tilde{f}(x+\Delta e_i) - \tilde{f}(x)}{\Delta} = \frac{f(x+\Delta e_i) + \epsilon_1 - f(x) - \epsilon_2}{\Delta} \\ &\cong \quad \frac{\partial f}{\partial x_i}(x) + \frac{\epsilon_1 - \epsilon_2}{\Delta}\end{aligned}$$

Problema: per avere una buona approssimazione devo usare $\Delta \ll 1$ **ma** in questo caso

$$\frac{\epsilon_1 - \epsilon_2}{\Delta} \gg 1!!!$$

Perché?

Supponiamo $f(x)$ sia affetta da rumore additivo quindi

$$\tilde{f}(x) = f(x) + \epsilon$$

Differenze finite su $\tilde{f}$:

$$\frac{\partial \tilde{f}}{\partial x_i}(x) \cong \frac{\tilde{f}(x+\Delta e_i) - \tilde{f}(x)}{\Delta} = \frac{f(x+\Delta e_i) + \epsilon_1 - f(x) - \epsilon_2}{\Delta}$$

$$\cong \frac{\partial f}{\partial x_i}(x) + \frac{\epsilon_1 - \epsilon_2}{\Delta}$$

Problema: per avere una buona approssimazione devo usare $\Delta \ll 1$ **ma** in questo caso

$$\frac{\epsilon_1 - \epsilon_2}{\Delta} \gg 1!!!$$

Perché?

Supponiamo $f(x)$ sia affetta da rumore additivo quindi

$$\tilde{f}(x) = f(x) + \epsilon$$

Differenze finite su $\tilde{f}$:

$$\begin{aligned} \frac{\partial \tilde{f}}{\partial x_i}(x) &\cong \frac{\tilde{f}(x + \Delta e_i) - \tilde{f}(x)}{\Delta} = \frac{f(x + \Delta e_i) + \epsilon_1 - f(x) - \epsilon_2}{\Delta} \\ &\cong \frac{\partial f}{\partial x_i}(x) + \frac{\epsilon_1 - \epsilon_2}{\Delta} \end{aligned}$$

Problema: per avere una buona approssimazione devo usare $\Delta \ll 1$ **ma** in questo caso

$$\frac{\epsilon_1 - \epsilon_2}{\Delta} \gg 1!!!$$

Perché?

Supponiamo $f(x)$ sia affetta da rumore additivo quindi

$$\tilde{f}(x) = f(x) + \epsilon$$

Differenze finite su $\tilde{f}$:

$$\begin{aligned}\frac{\partial \tilde{f}}{\partial x_i}(x) &\cong \frac{\tilde{f}(x+\Delta e_i) - \tilde{f}(x)}{\Delta} = \frac{f(x+\Delta e_i) + \epsilon_1 - f(x) - \epsilon_2}{\Delta} \\ &\cong \frac{\partial f}{\partial x_i}(x) + \frac{\epsilon_1 - \epsilon_2}{\Delta}\end{aligned}$$

Problema: per avere una buona approssimazione devo usare $\Delta \ll 1$ **ma** in questo caso

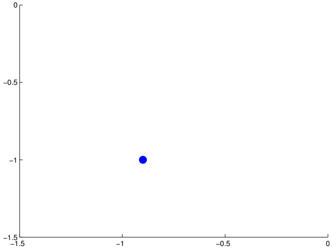
$$\frac{\epsilon_1 - \epsilon_2}{\Delta} \gg 1!!!$$

Compass Search

Consideriamo il problema:

$$\min_{x \in \mathbb{R}^2} f(x)$$

Punto iniziale: $x_0 = (-0.9; -1.0)^\top$
$f(x)$ iniziale: $f(x_0) = 11.3524$
Passo iniziale: $\Delta = 0.3$

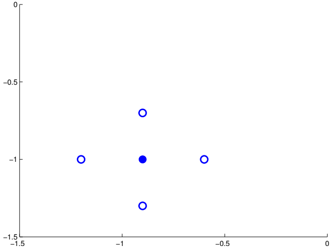


Compass Search

Consideriamo il problema:

$$\min_{x \in \mathbb{R}^2} f(x)$$

Punto iniziale: $x_0 = (-0.9; -1.0)^\top$
$f(x)$ iniziale: $f(x_0) = 11.3524$
Passo iniziale: $\Delta = 0.3$



Calcoliamo $f(x)$ nei punti

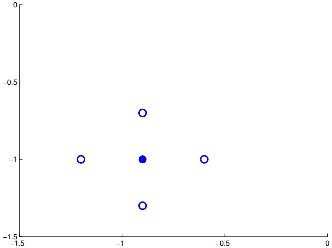
Est	11.7904
Ovest	19.9504
Nord	
Sud	29.4628

Compass Search

Consideriamo il problema:

$$\min_{x \in \mathbb{R}^2} f(x)$$

Punto iniziale: $x_0 = (-0.9; -1.0)^\top$
$f(x)$ iniziale: $f(x_0) = 11.3524$
Passo iniziale: $\Delta = 0.3$



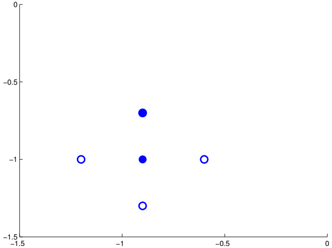
Calcoliamo $f(x)$ nei punti

Est	11.7904
Ovest	19.9504
Nord	5.0788
Sud	29.4628

Compass Search

Consideriamo il problema:

$$\min_{x \in \mathbb{R}^2} f(x)$$



Punto corrente: $x_k = (-0.9; -0.7)^\top$
$f(x)$ corrente: $f(x_k) = 5.0788$
Passo corrente: $\Delta = 0.3$

Calcoliamo $f(x)$ nei punti

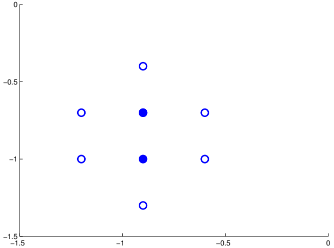
Est	11.7904
Ovest	19.9504
Nord	**5.0788**
Sud	29.4628

Compass Search

Consideriamo il problema:

$$\min_{x \in \mathbb{R}^2} f(x)$$

Punto corrente: $x_k = (-0.9; -0.7)^\top$
$f(x)$ corrente: $f(x_k) = 5.0788$
Passo corrente: $\Delta = 0.3$



Calcoliamo $f(x)$ nei punti

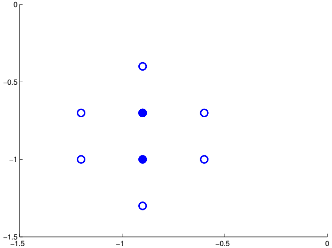
Est	
Ovest	17.4208
Nord	6.4948
Sud	11.3524

Compass Search

Consideriamo il problema:

$$\min_{x\in\mathbb{R}^2} f(x)$$

Punto corrente: $x_k = (-0.9; -0.7)^\top$
$f(x)$ corrente: $f(x_k) = 5.0788$
Passo corrente: $\Delta = 0.3$



Calcoliamo $f(x)$ nei punti

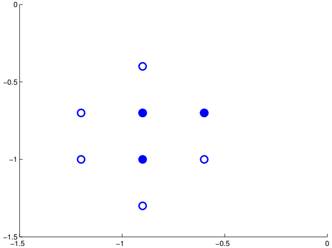
Est	2.2048
Ovest	17.4208
Nord	6.4948
Sud	11.3524

Compass Search

Consideriamo il problema:

$$\min_{x \in \mathbb{R}^2} f(x)$$

Punto corrente: $x_k = (-0.6; -0.7)^\top$
$f(x)$ corrente: $f(x_k) = 2.2048$
Passo corrente: $\Delta = 0.3$



Calcoliamo $f(x)$ nei punti

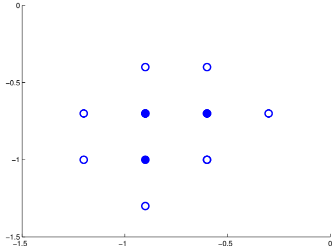
Est	**2.2048**
Ovest	17.4208
Nord	6.4948
Sud	11.3524

Compass Search

Consideriamo il problema:

$$\min_{x \in \mathbb{R}^2} f(x)$$

Punto corrente: $x_k = (-0.6; -0.7)^\top$
$f(x)$ corrente: $f(x_k) = 2.2048$
Passo corrente: $\Delta = 0.3$



Calcoliamo $f(x)$ nei punti

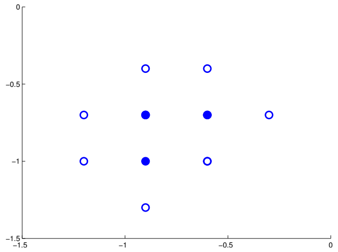
Est	4.9108
Ovest	5.0788
Nord	
Sud	11.7904

Compass Search

Consideriamo il problema:

$$\min_{x \in \mathbb{R}^2} f(x)$$

Punto corrente: $x_k = (-0.6; -0.7)^\top$
$f(x)$ corrente: $f(x_k) = 2.2048$
Passo corrente: $\Delta = 0.3$



Calcoliamo $f(x)$ nei punti

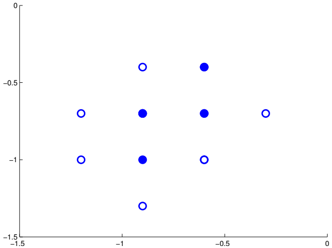
Est	4.9108
Ovest	5.0788
Nord	0.5248
Sud	11.7904

Compass Search

Consideriamo il problema:

$$\min_{x \in \mathbb{R}^2} f(x)$$

Punto corrente: $x_k = (-0.6; -0.4)^\top$
$f(x)$ corrente: $f(x_k) = 0.5248$
Passo corrente: $\Delta = 0.3$



Calcoliamo $f(x)$ nei punti

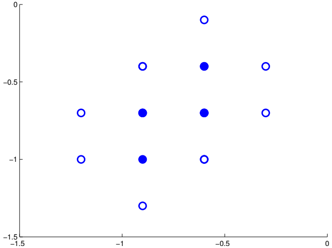
Est	4.9108
Ovest	5.0788
Nord	**0.5248**
Sud	11.7904

Compass Search

Consideriamo il problema:

$$\min_{x \in \mathbb{R}^2} f(x)$$

Punto corrente: $x_k = (-0.6; -0.4)^\top$
$f(x)$ corrente: $f(x_k) = 0.5248$
Passo corrente: $\Delta = 0.3$



Calcoliamo $f(x)$ nei punti

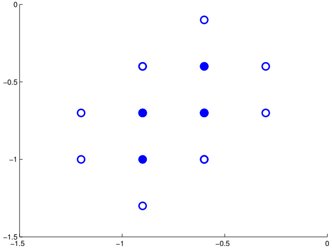
Est	0.5668
Ovest	6.4948
Nord	3.3808
Sud	2.2048

Compass Search

Consideriamo il problema:

$$\min_{x \in \mathbb{R}^2} f(x)$$

Punto corrente: $x_k = (-0.6; -0.4)^\top$
$f(x)$ corrente: $f(x_k) = 0.5248$
Passo corrente: $\Delta = 0.3$



Calcoliamo $f(x)$ nei punti

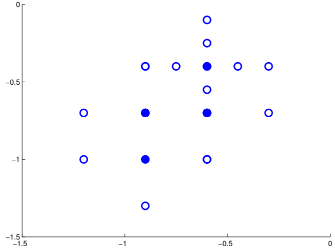
Est	0.5668
Ovest	6.4948
Nord	3.3808
Sud	2.2048

Compass Search

Consideriamo il problema:

$$\min_{x \in \mathbb{R}^2} f(x)$$

Punto corrente: $x_k = (-0.6; -0.4)^\top$
$f(x)$ corrente: $f(x_k) = 0.5248$
Passo corrente: $\Delta = 0.15$



Calcoliamo $f(x)$ nei punti

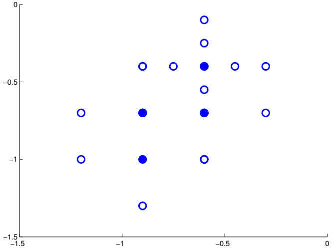
Est	
Ovest	2.5335
Nord	1.5660
Sud	0.6054

Compass Search

Consideriamo il problema:

$$\min_{x \in \mathbb{R}^2} f(x)$$

Punto corrente: $x_k = (-0.6; -0.4)^\top$
$f(x)$ corrente: $f(x_k) = 0.5248$
Passo corrente: $\Delta = 0.15$



Calcoliamo $f(x)$ nei punti

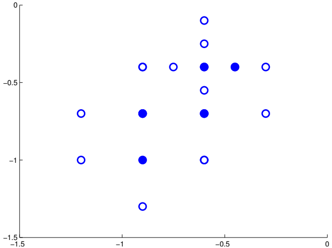
Est	0.0069
Ovest	2.5335
Nord	1.5660
Sud	0.6054

Compass Search

Consideriamo il problema:

$$\min_{x \in \mathbb{R}^2} f(x)$$

Punto corrente: $x_k = (-0.45; -0.4)^\top$
$f(x)$ corrente: $f(x_k) = 0.0069$
Passo corrente: $\Delta = 0.15$
Calcoliamo $f(x)$ nei punti



Est	**0.0069**
Ovest	2.5335
Nord	1.5660
Sud	0.6054

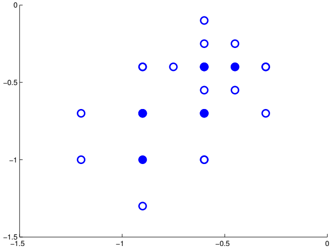
Compass Search

Consideriamo il problema:

$$\min_{x \in \mathbb{R}^2} f(x)$$

Punto corrente: $x_k = (-0.45; -0.4)^\top$
$f(x)$ corrente: $f(x_k) = 0.0069$
Passo corrente: $\Delta = 0.15$

Calcoliamo $f(x)$ nei punti

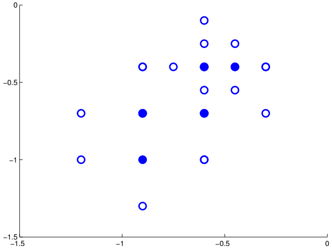


Est	0.5668
Ovest	0.5248
Nord	0.3957
Sud	0.7670

Compass Search

Consideriamo il problema:

$$\min_{x \in \mathbb{R}^2} f(x)$$



Punto corrente: $x_k = (-0.45; -0.4)^\top$
$f(x)$ corrente: $f(x_k) = 0.0069$
Passo corrente: $\Delta = 0.15$
Calcoliamo $f(x)$ nei punti

Est	0.5668
Ovest	0.5248
Nord	0.3957
Sud	0.7670

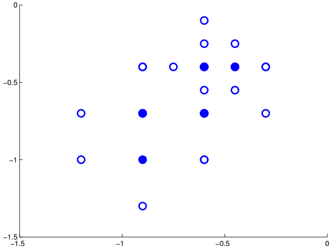
Compass Search

Consideriamo il problema:

$$\min_{x \in \mathbb{R}^2} f(x)$$

Punto corrente: $x_k = (-0.45; -0.4)^\top$
$f(x)$ corrente: $f(x_k) = 0.0069$
Passo corrente: $\Delta = 0.075$

Calcoliamo $f(x)$ nei punti



Est	0.5668
Ovest	0.5248
Nord	0.3957
Sud	0.7670

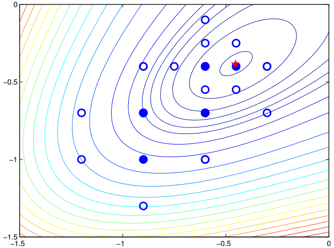
Compass Search

Consideriamo il problema:

$$\min_{x \in \mathbb{R}^2} f(x)$$

Punto corrente: $x_k = (-0.45; -0.4)^\top$
$f(x)$ corrente: $f(x_k) = 0.0069$
Passo corrente: $\Delta = 0.075$

Calcoliamo $f(x)$ nei punti



Est	0.5668
Ovest	0.5248
Nord	0.3957
Sud	0.7670

Ricapitolando

$f(x_k)$	Δ_k
11.352400	0.300000
5.078800	0.300000
2.204800	0.300000
0.524800	0.300000
0.524800	0.150000
0.006925	0.150000
0.006925	0.075000
0.006925	0.037500
0.006925	0.018750
0.000298	0.018750
0.000298	0.009375
0.000298	0.004687
0.000298	0.002344
0.000173	0.002344
0.000054	0.002344
0.000043	0.002344
0.000033	0.002344

Ricapitolando

$f(x_k)$	Δ_k
11.352400	0.300000
5.078800	0.300000
2.204800	0.300000
0.524800	0.300000
0.524800	0.150000
0.006925	0.150000
0.006925	0.075000
0.006925	0.037500
0.006925	0.018750
0.000298	0.018750
0.000298	0.009375
0.000298	0.004687
0.000298	0.002344
0.000173	0.002344
0.000054	0.002344
0.000043	0.002344
0.000033	0.002344

Ricapitolando

$f(x_k)$	Δ_k
11.352400	0.300000
5.078800	0.300000
2.204800	0.300000
0.524800	0.300000
0.524800	0.150000
0.006925	0.150000
0.006925	0.075000
0.006925	0.037500
0.006925	0.018750
0.000298	0.018750
0.000298	0.009375
0.000298	0.004687
0.000298	0.002344
0.000173	0.002344
0.000054	0.002344
0.000043	0.002344
0.000033	0.002344

Ricapitolando

$f(x_k)$	Δ_k
11.352400	0.300000
5.078800	0.300000
2.204800	0.300000
0.524800	0.300000
0.524800	0.150000
0.006925	0.150000
0.006925	0.075000
0.006925	0.037500
0.006925	0.018750
0.000298	0.018750
0.000298	0.009375
0.000298	0.004687
0.000298	0.002344
0.000173	0.002344
0.000054	0.002344
0.000043	0.002344
0.000033	0.002344

Ricapitolando

$f(x_k)$	Δ_k
11.352400	0.300000
5.078800	0.300000
2.204800	0.300000
0.524800	0.300000
0.524800	0.150000
0.006925	0.150000
0.006925	0.075000
0.006925	0.037500
0.006925	0.018750
0.000298	0.018750
0.000298	0.009375
0.000298	0.004687
0.000298	0.002344
0.000173	0.002344
0.000054	0.002344
0.000043	0.002344
0.000033	0.002344

Ricapitolando

$f(x_k)$	Δ_k
11.352400	0.300000
5.078800	0.300000
2.204800	0.300000
0.524800	0.300000
0.524800	0.150000
0.006925	0.150000
0.006925	0.075000
0.006925	0.037500
0.006925	0.018750
0.000298	0.018750
0.000298	0.009375
0.000298	0.004687
0.000298	0.002344
0.000173	0.002344
0.000054	0.002344
0.000043	0.002344
0.000033	0.002344

Ricapitolando

$f(x_k)$	Δ_k
11.352400	0.300000
5.078800	0.300000
2.204800	0.300000
0.524800	0.300000
0.524800	0.150000
0.006925	0.150000
0.006925	0.075000
0.006925	0.037500
0.006925	0.018750
0.000298	0.018750
0.000298	0.009375
0.000298	0.004687
0.000298	0.002344
0.000173	0.002344
0.000054	0.002344
0.000043	0.002344
0.000033	0.002344