

Otimização

Estrutura

- Introdução
- Métodos baseados no gradiente
 - Método de Newton
 - Método de Gauss-Newton
 - Método de *Steepest Decent*
 - Método de Levenberg-Marquardt
- Métodos Heurísticos
 - *Simulated Annealing*
 - Método das Formigas

Introdução

$$\bar{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_M \end{bmatrix}_{M \times 1}$$

parâmetros

$$\bar{d} = \begin{bmatrix} d_1 \\ \vdots \\ d_N \end{bmatrix}_{N \times 1}$$

dados
observados

$$\bar{g}(\bar{p}) = \begin{bmatrix} g_1(\bar{p}) \\ \vdots \\ g_N(\bar{p}) \end{bmatrix}_{N \times 1}$$

dados
preditos

Introdução

$$\bar{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_M \end{bmatrix}_{M \times 1}$$

parâmetros

$$\bar{d} = \begin{bmatrix} d_1 \\ \vdots \\ d_N \end{bmatrix}_{N \times 1}$$

dados
observados

$$\bar{g}(\bar{p}) = \begin{bmatrix} g_1(\bar{p}) \\ \vdots \\ g_N(\bar{p}) \end{bmatrix}_{N \times 1}$$

dados
preditos

$$\phi(\bar{p}) = [\bar{d} - \bar{g}(\bar{p})]^T [\bar{d} - \bar{g}(\bar{p})]$$

norma L2
(função escalar)

Introdução

$$\bar{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_M \end{bmatrix}_{M \times 1}$$

parâmetros

$$\bar{\nabla} \phi(\bar{p}^*) = \bar{0}_{M \times 1}$$

$$\bar{d} = \begin{bmatrix} d_1 \\ \vdots \\ d_N \end{bmatrix}_{N \times 1}$$

dados
observados

$$\bar{g}(\bar{p}) = \begin{bmatrix} g_1(\bar{p}) \\ \vdots \\ g_N(\bar{p}) \end{bmatrix}_{N \times 1}$$

dados
preditos

$$\phi(\bar{p}) = [\bar{d} - \bar{g}(\bar{p})]^T [\bar{d} - \bar{g}(\bar{p})]$$

norma L2
(função escalar)

Introdução

$$\bar{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_M \end{bmatrix}_{M \times 1}$$

parâmetros

$$\bar{d} = \begin{bmatrix} d_1 \\ \vdots \\ d_N \end{bmatrix}_{N \times 1}$$

dados
observados

$$\bar{g}(\bar{p}) = \begin{bmatrix} g_1(\bar{p}) \\ \vdots \\ g_N(\bar{p}) \end{bmatrix}_{N \times 1}$$

dados
preditos

$$\phi(\bar{p}) = [\bar{d} - \bar{g}(\bar{p})]^T [\bar{d} - \bar{g}(\bar{p})]$$

norma L2
(função escalar)

$$\bar{\nabla} \phi(\bar{p}^*) = \bar{0}_{M \times 1}$$

Problema linear

Problema não-linear

Introdução

$$\bar{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_M \end{bmatrix}_{M \times 1}$$

parâmetros

$$\bar{d} = \begin{bmatrix} d_1 \\ \vdots \\ d_N \end{bmatrix}_{N \times 1}$$

dados
observados

$$\bar{g}(\bar{p}) = \begin{bmatrix} g_1(\bar{p}) \\ \vdots \\ g_N(\bar{p}) \end{bmatrix}_{N \times 1}$$

dados
preditos

$$\phi(\bar{p}) = [\bar{d} - \bar{g}(\bar{p})]^T [\bar{d} - \bar{g}(\bar{p})]$$

norma L2
(função escalar)

$$\bar{\nabla} \phi(\bar{p}^*) = \bar{0}_{M \times 1}$$

Problema linear

Problema não-linear

Introdução

$$\bar{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_M \end{bmatrix}_{M \times 1}$$

parâmetros

$$\bar{d} = \begin{bmatrix} d_1 \\ \vdots \\ d_N \end{bmatrix}_{N \times 1}$$

dados
observados

$$\bar{g}(\bar{p}) = \begin{bmatrix} g_1(\bar{p}) \\ \vdots \\ g_N(\bar{p}) \end{bmatrix}_{N \times 1}$$

dados
preditos

$$\phi(\bar{p}) = [\bar{d} - \bar{g}(\bar{p})]^T [\bar{d} - \bar{g}(\bar{p})]$$

norma L2
(função escalar)

$$\bar{\nabla} \phi(\bar{p}^*) = \bar{0}_{M \times 1}$$

Problema linear

$$\bar{g}(\bar{p}) = \bar{B}\bar{p} + \bar{b}$$

$$\bar{p}^* = \left(\begin{bmatrix} \bar{B} & \bar{B} \end{bmatrix}^T \begin{bmatrix} \bar{B} & \bar{B} \end{bmatrix} \right)^{-1} \begin{bmatrix} \bar{B} & \bar{B} \end{bmatrix}^T [\bar{d} - \bar{b}]$$

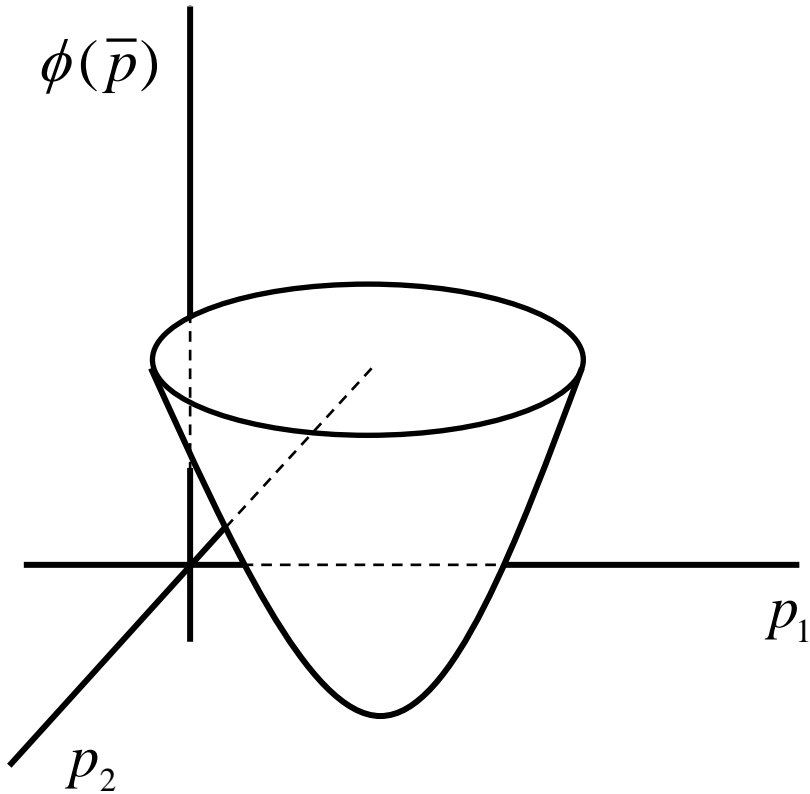
Problema não-linear

$$\bar{g}(\bar{p}) \neq \bar{B}\bar{p} + \bar{b}$$

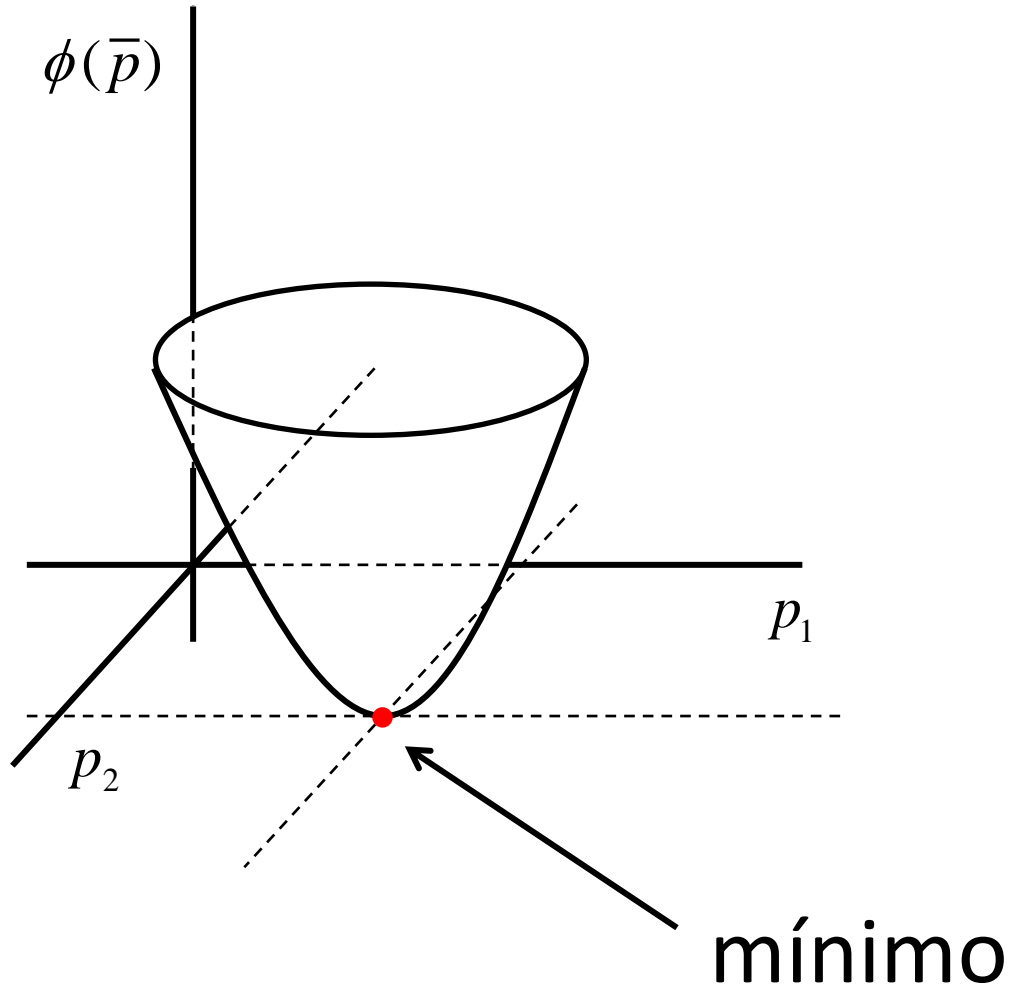
$$\Delta \bar{p} = \left(\bar{G}(\bar{p}_0)^T \bar{G}(\bar{p}_0) \right)^{-1} \bar{G}(\bar{p}_0)^T [\bar{d} - \bar{g}(\bar{p}_0)]$$

Introdução

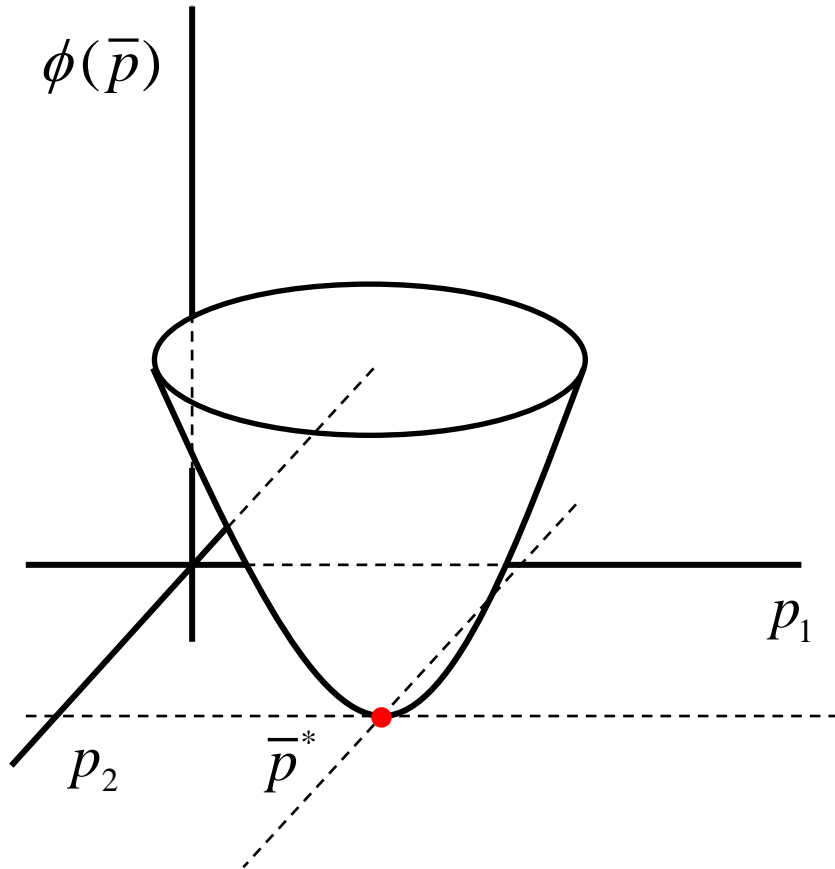
Problema linear



Introdução

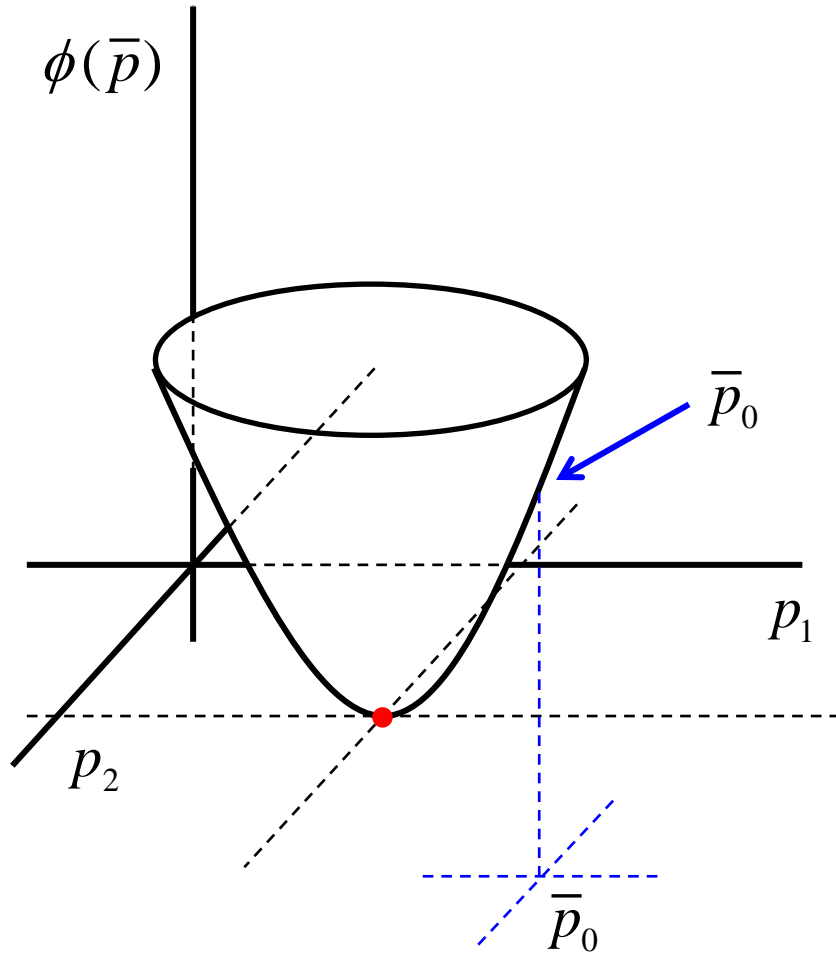


Introdução



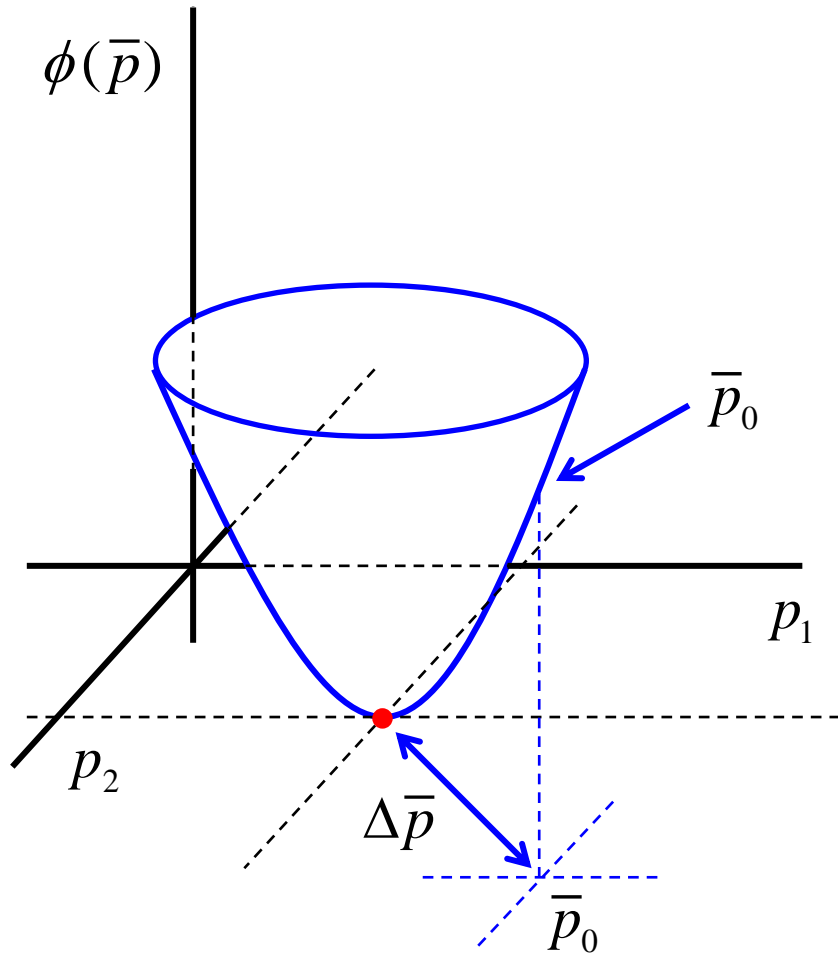
O mínimo pode
ser calculado
diretamente

Introdução

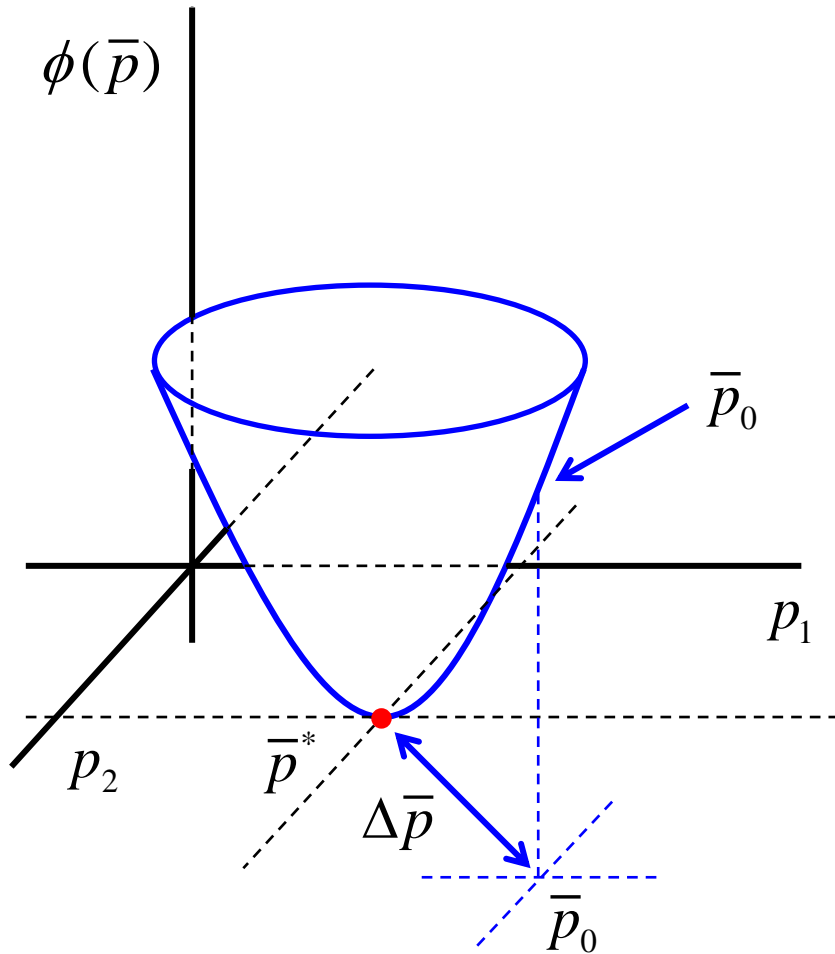


Ou a partir de
uma aproximação
inicial

Introdução



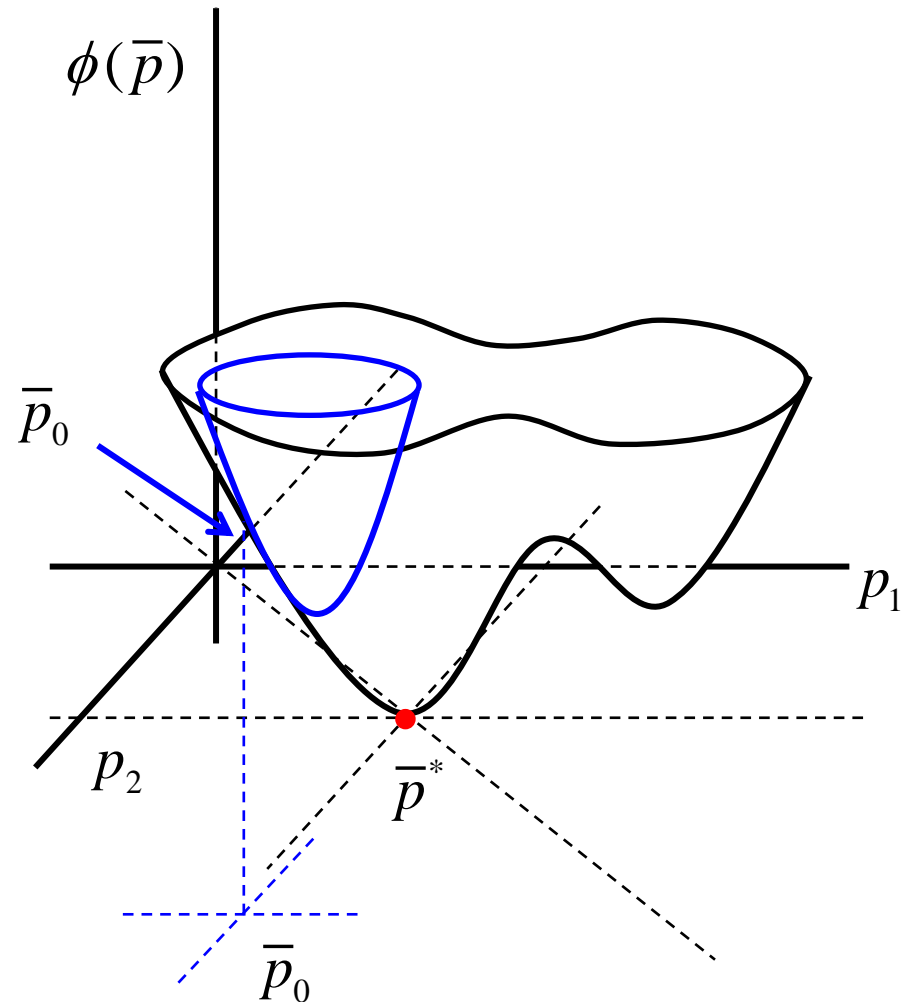
Introdução



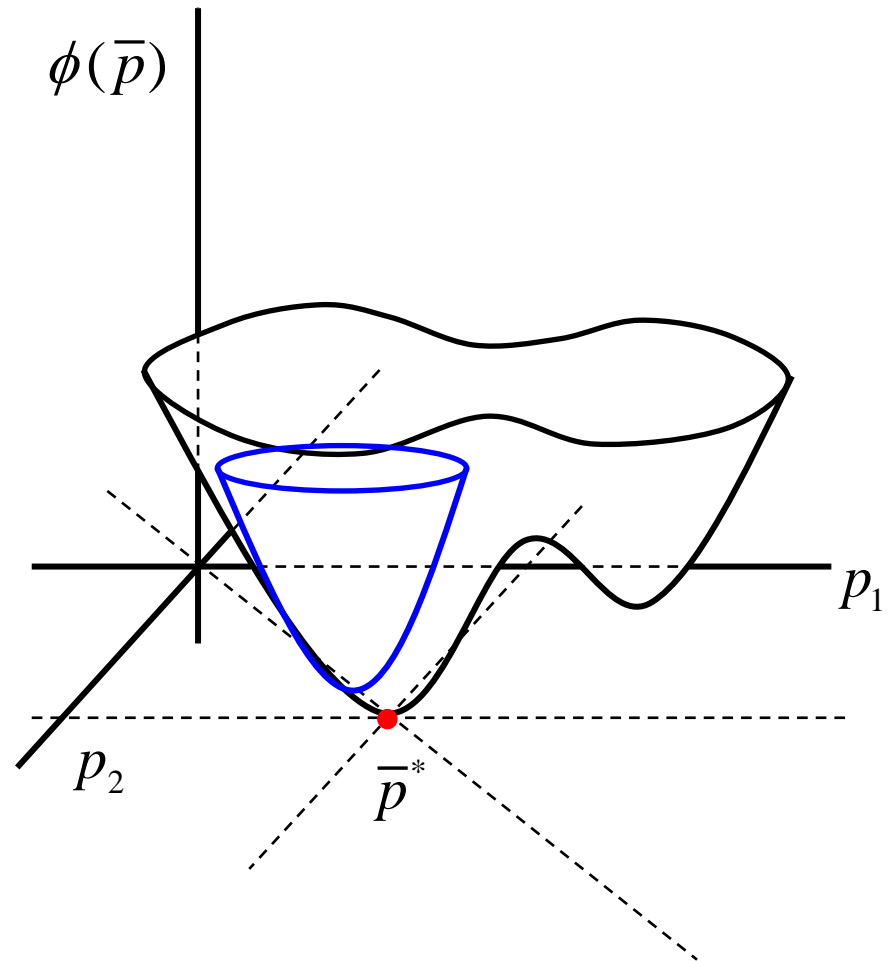
Nesse caso, o mínimo é encontrado com apenas um “passo” a partir da aproximação inicial

Introdução

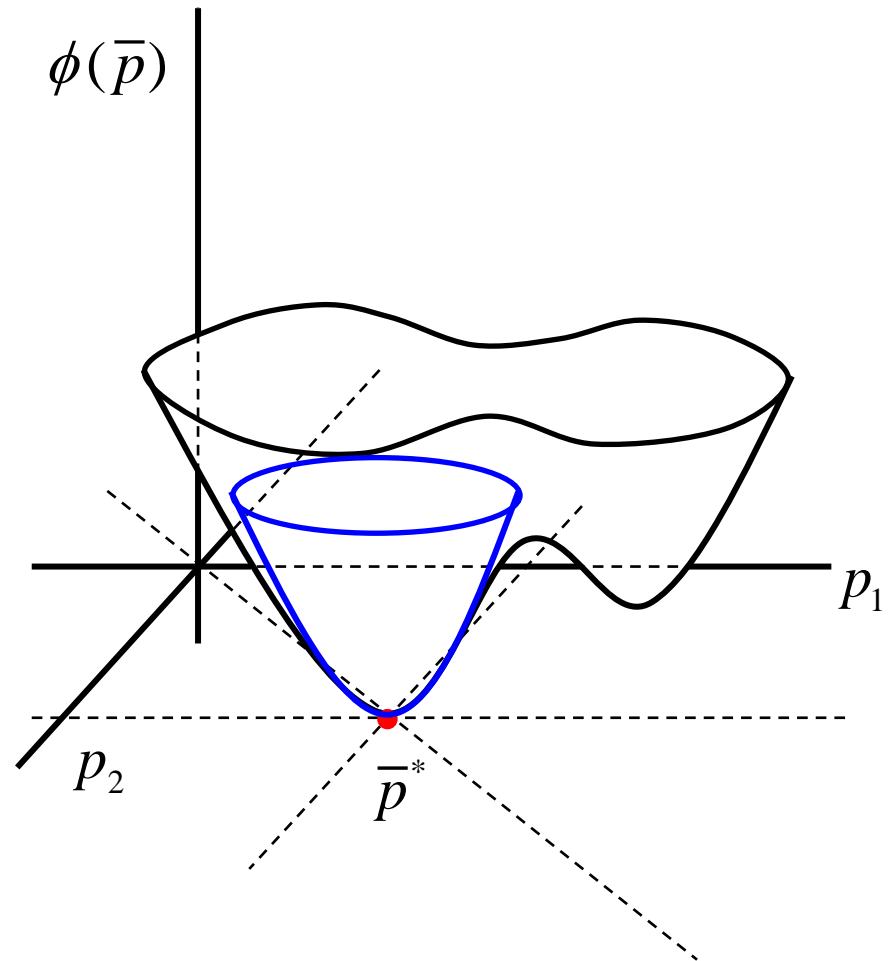
Por outro lado, em um problema não-linear, o mínimo é encontrado após sucessivos “passos” a partir de uma aproximação inicial



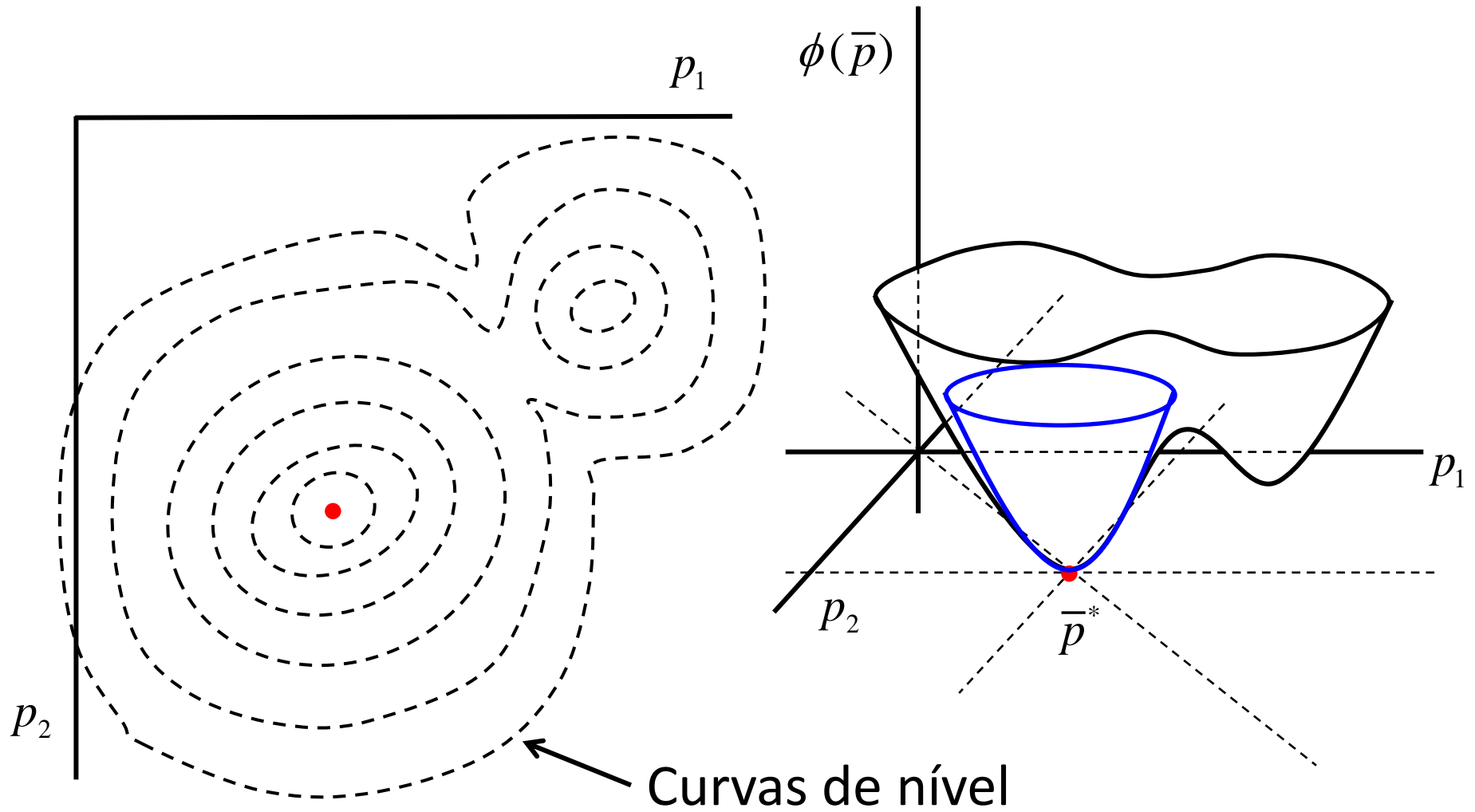
Introdução



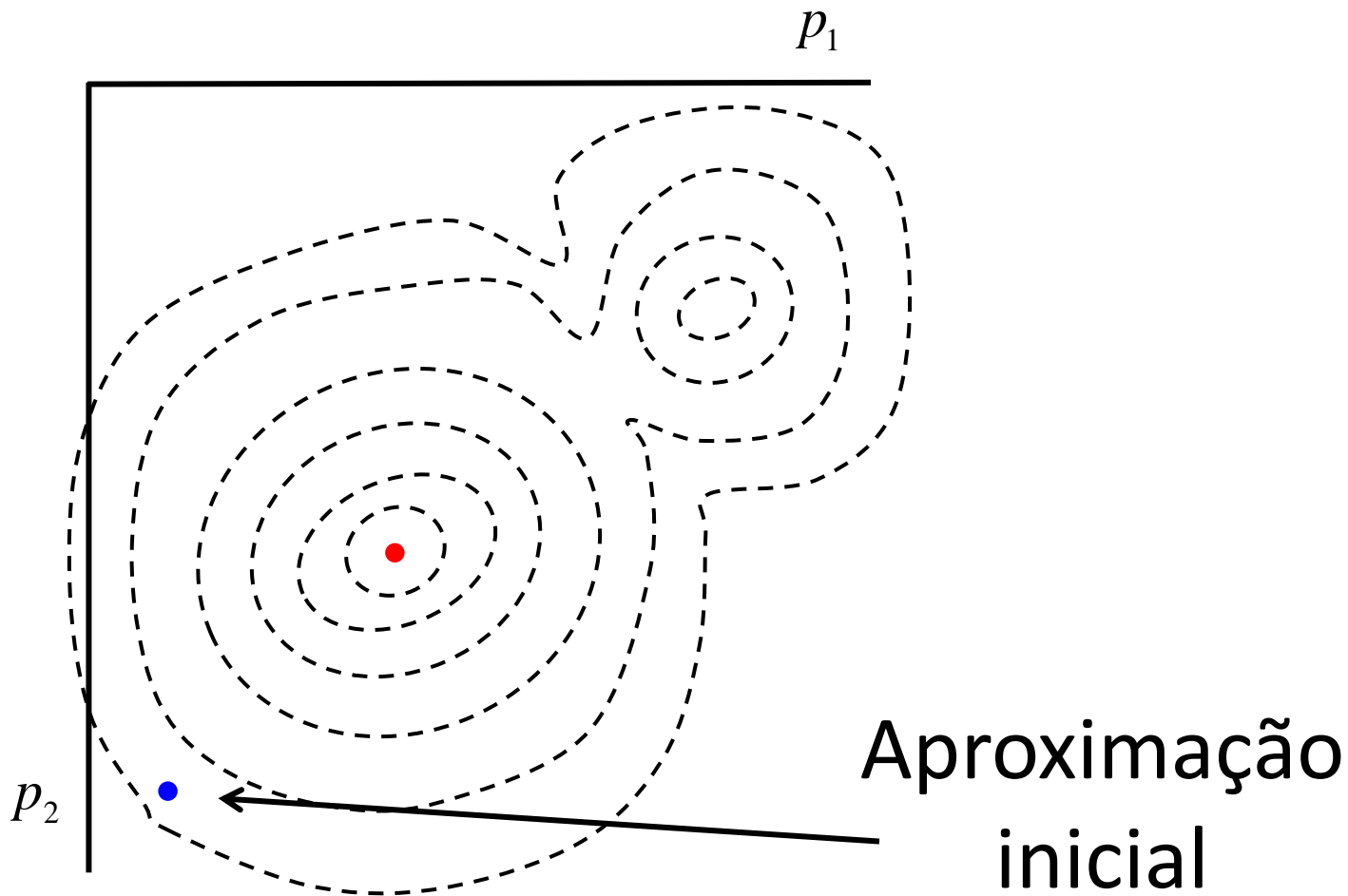
Introdução



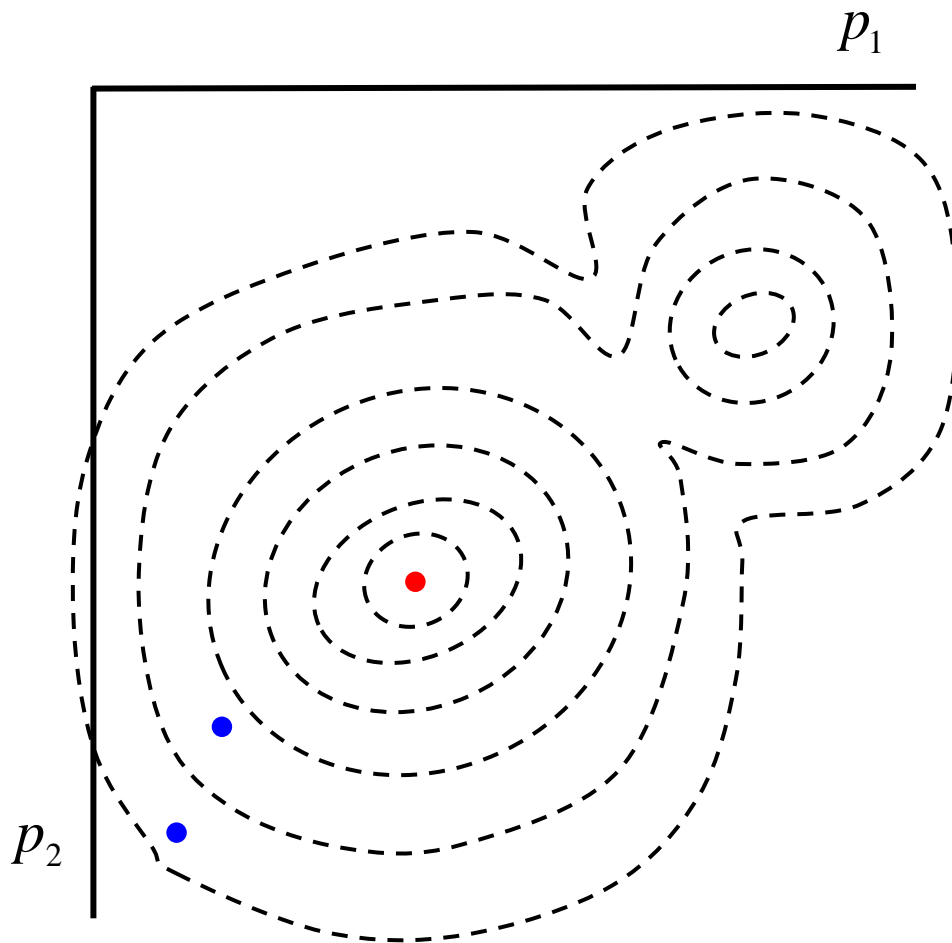
Introdução



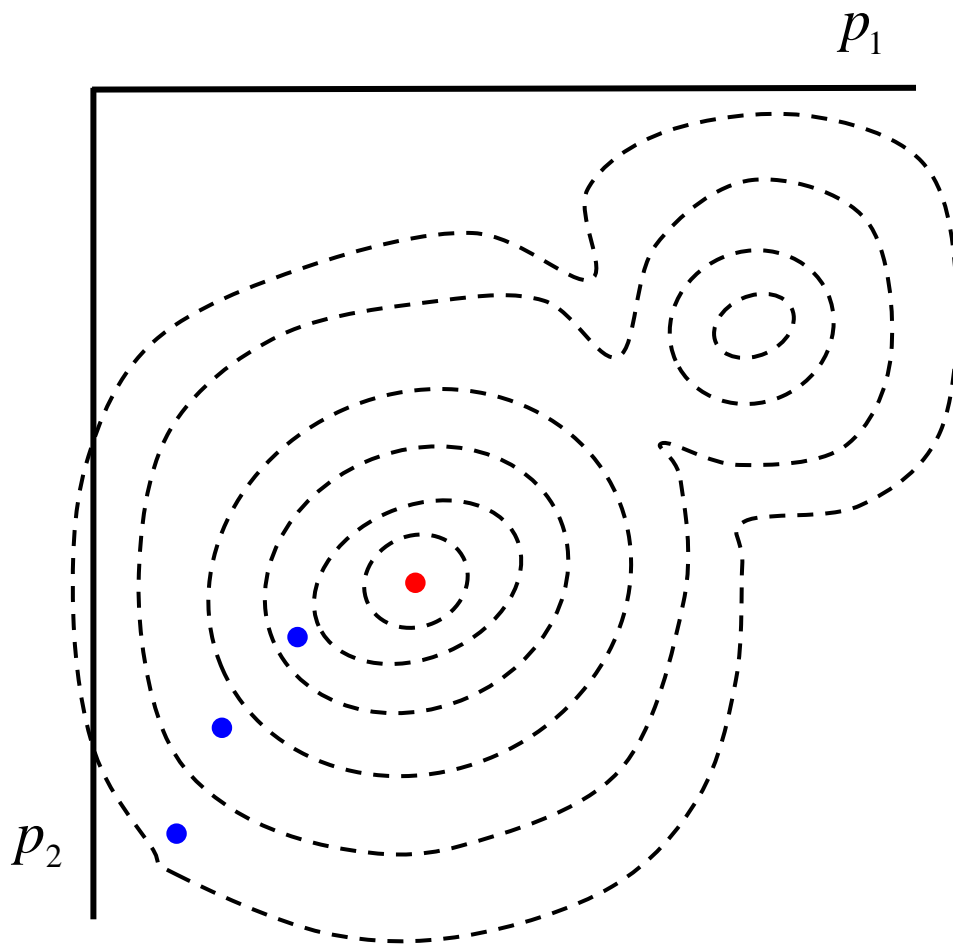
Introdução



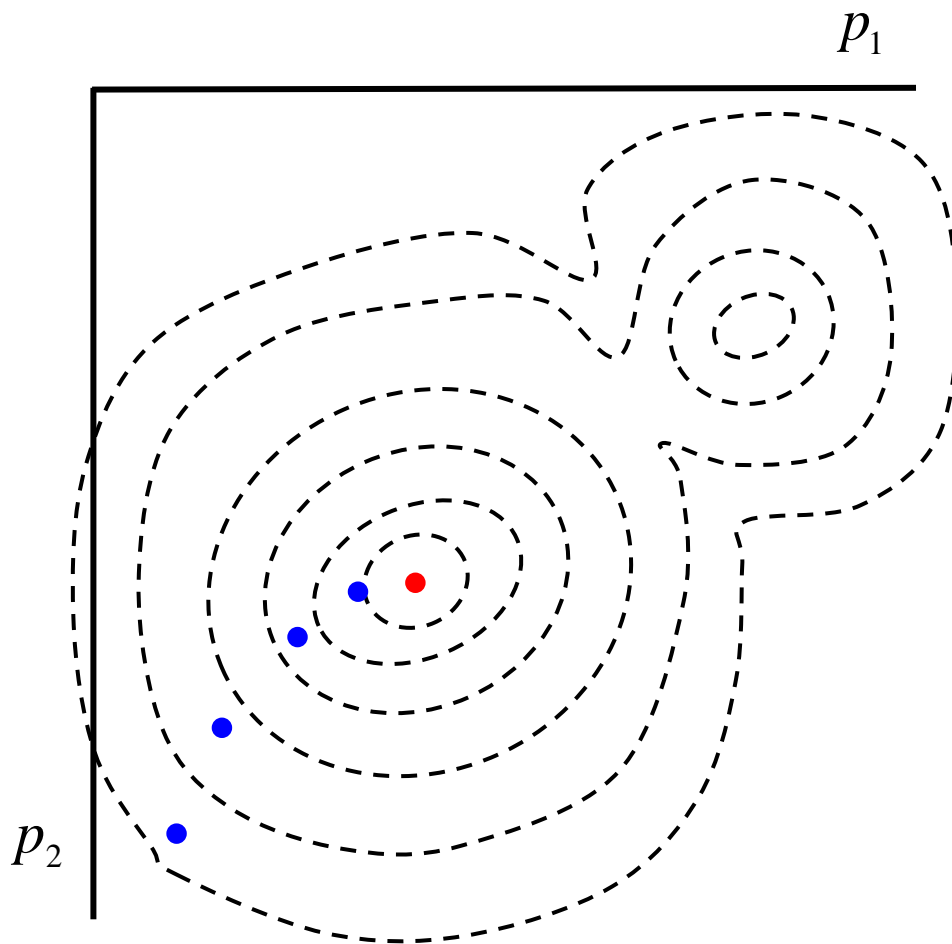
Introdução



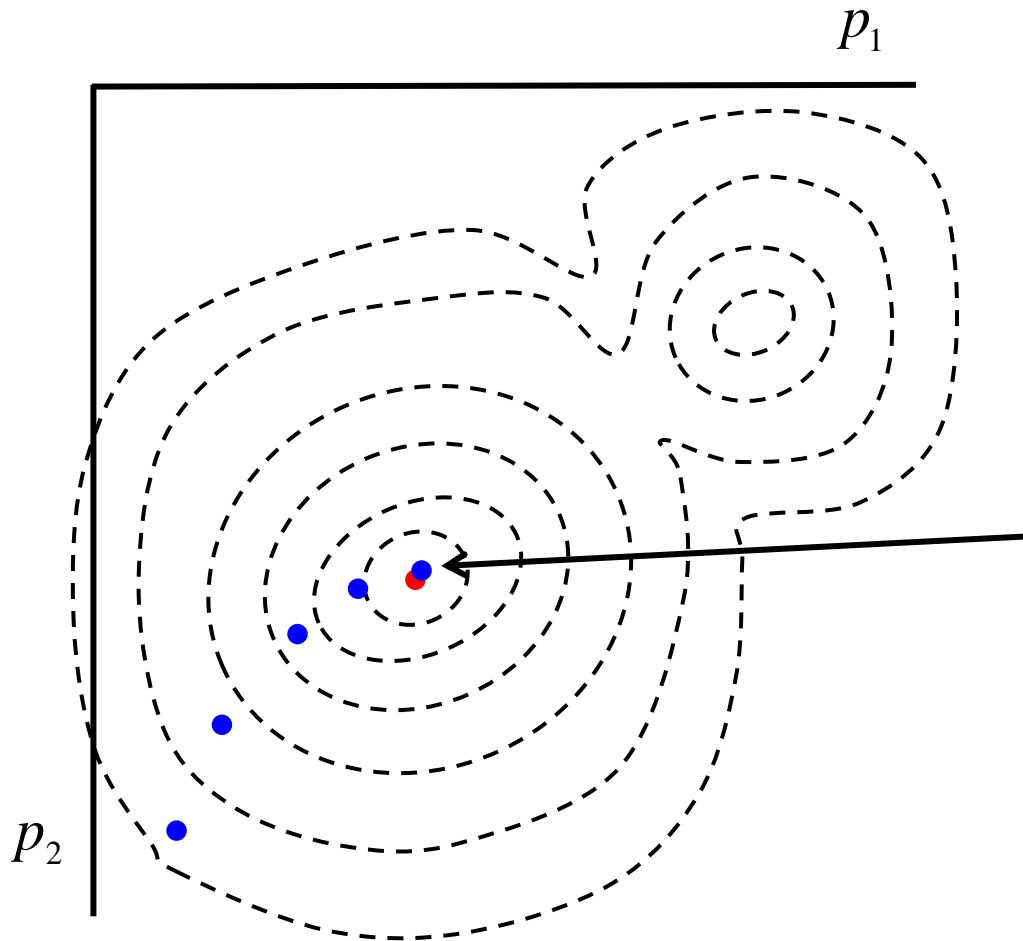
Introdução



Introdução

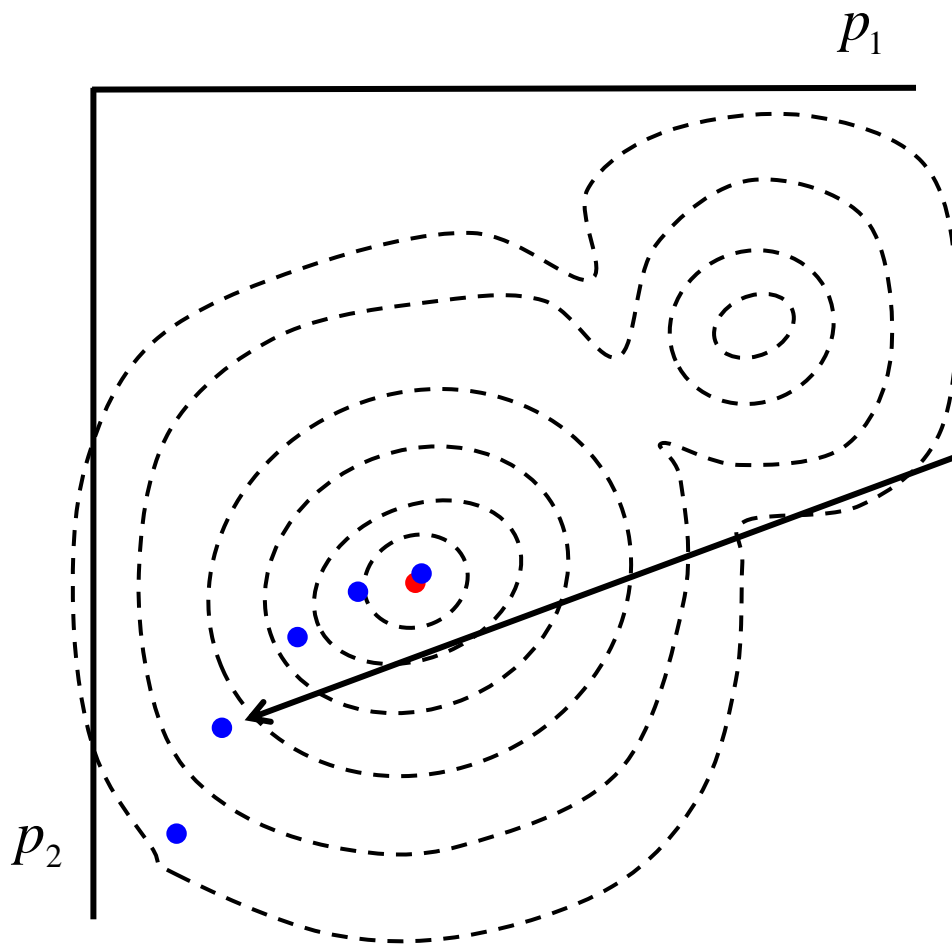


Introdução



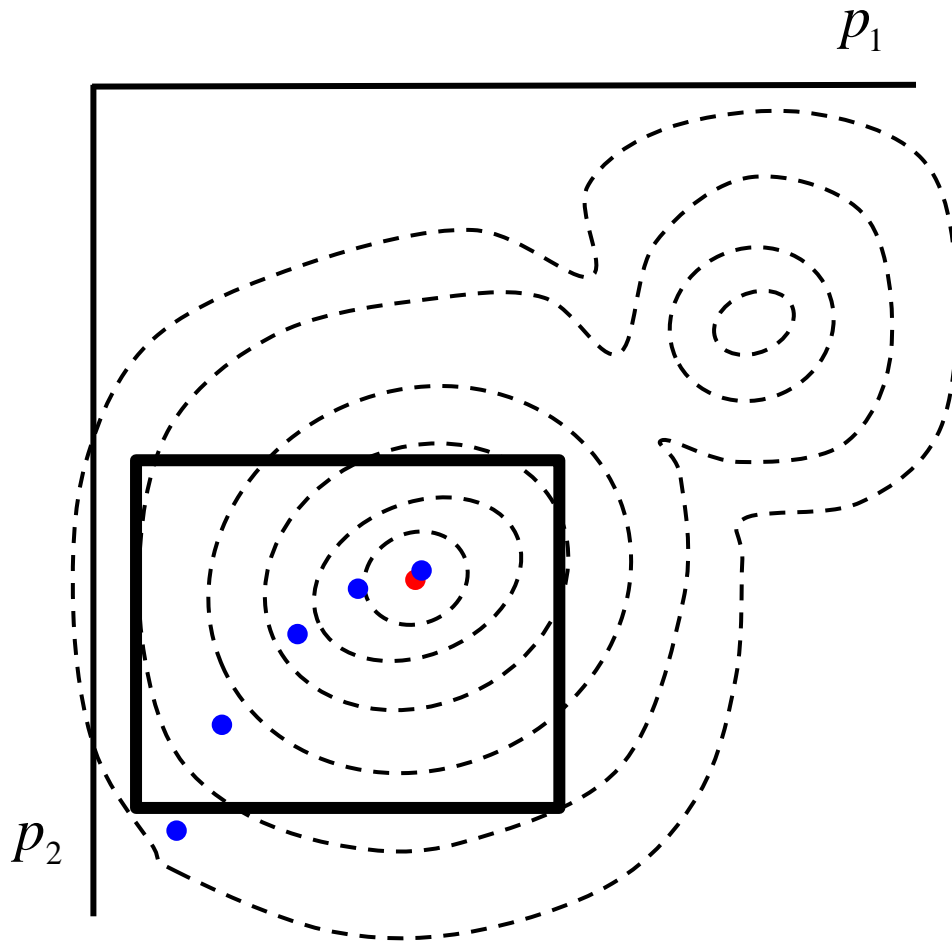
Estimativa do
ponto mínimo
obtida após a
otimização

Introdução

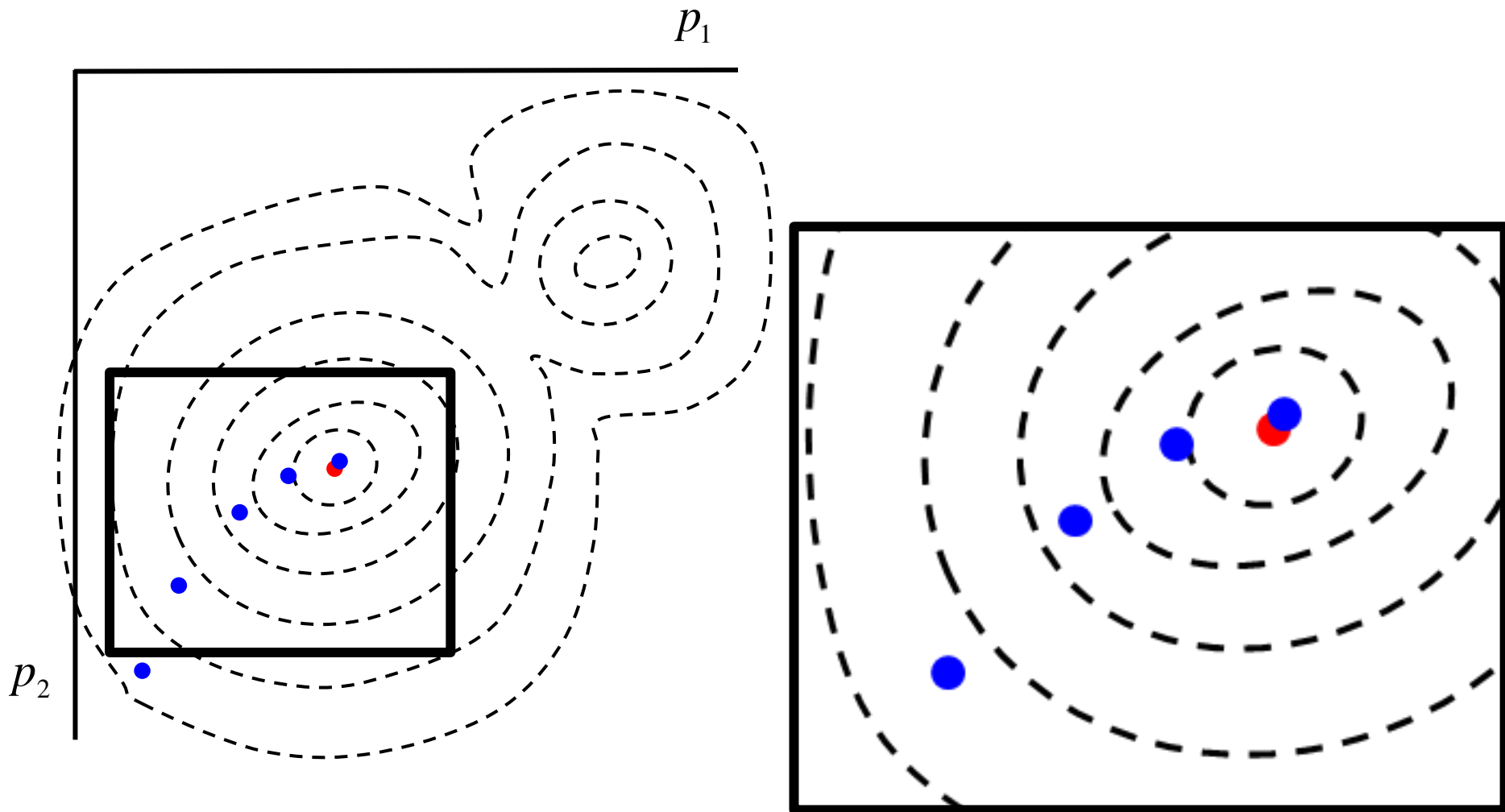


A partir de um determinado ponto, é necessário estabelecer uma direção e o comprimento do passo que será dado

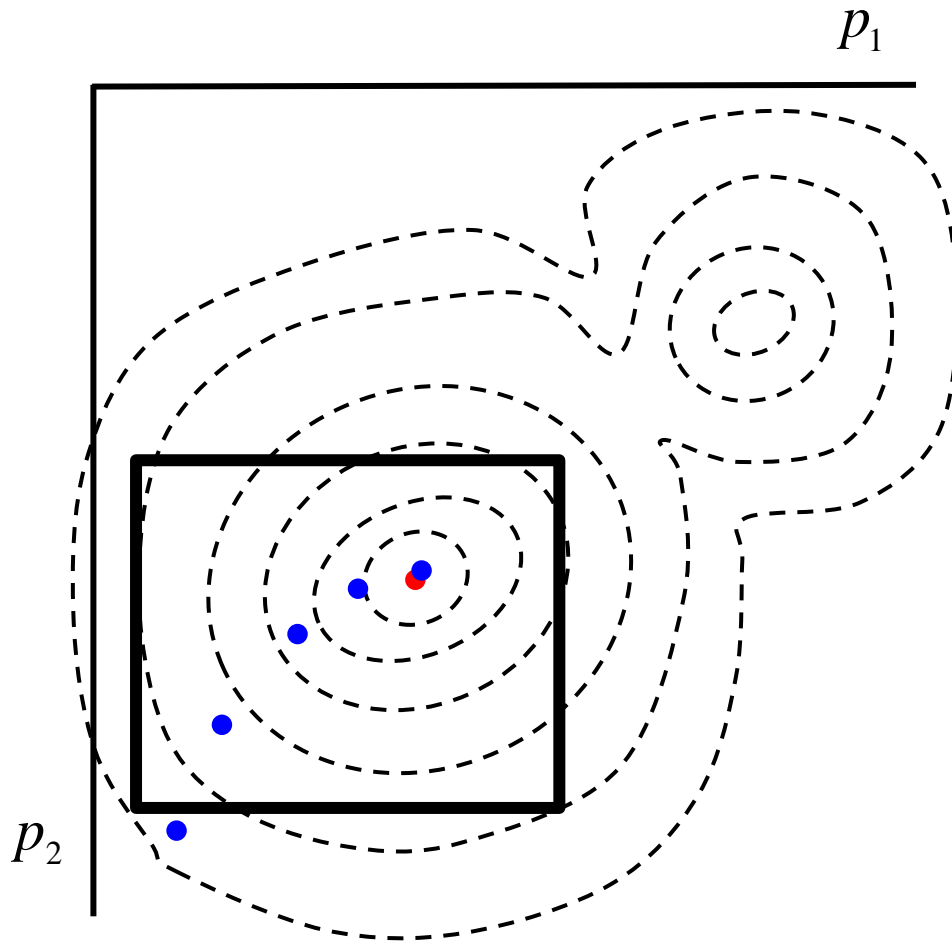
Introdução



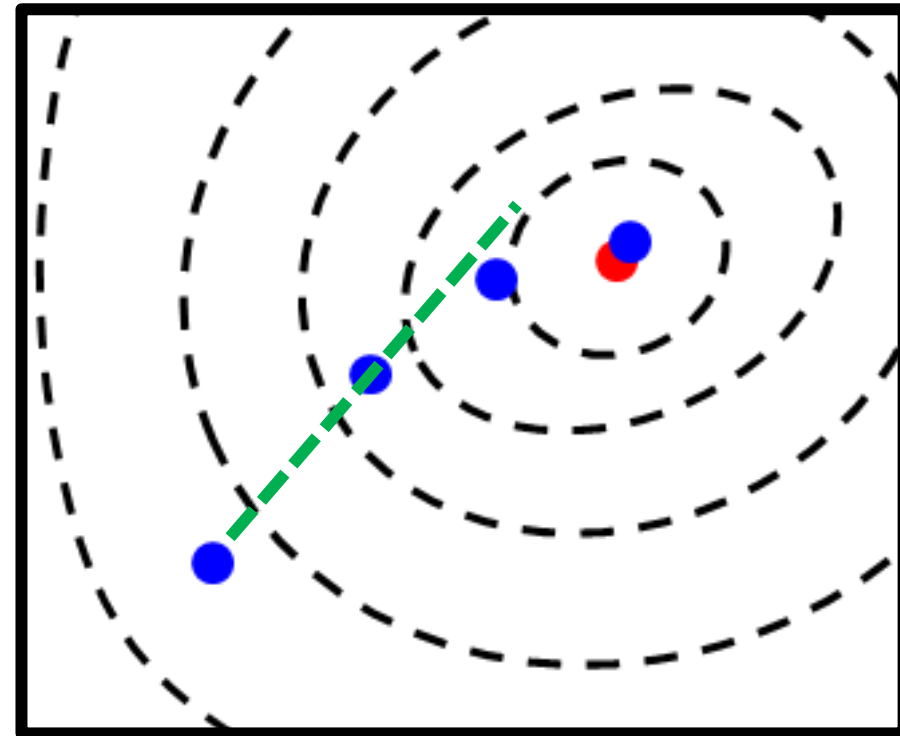
Introdução



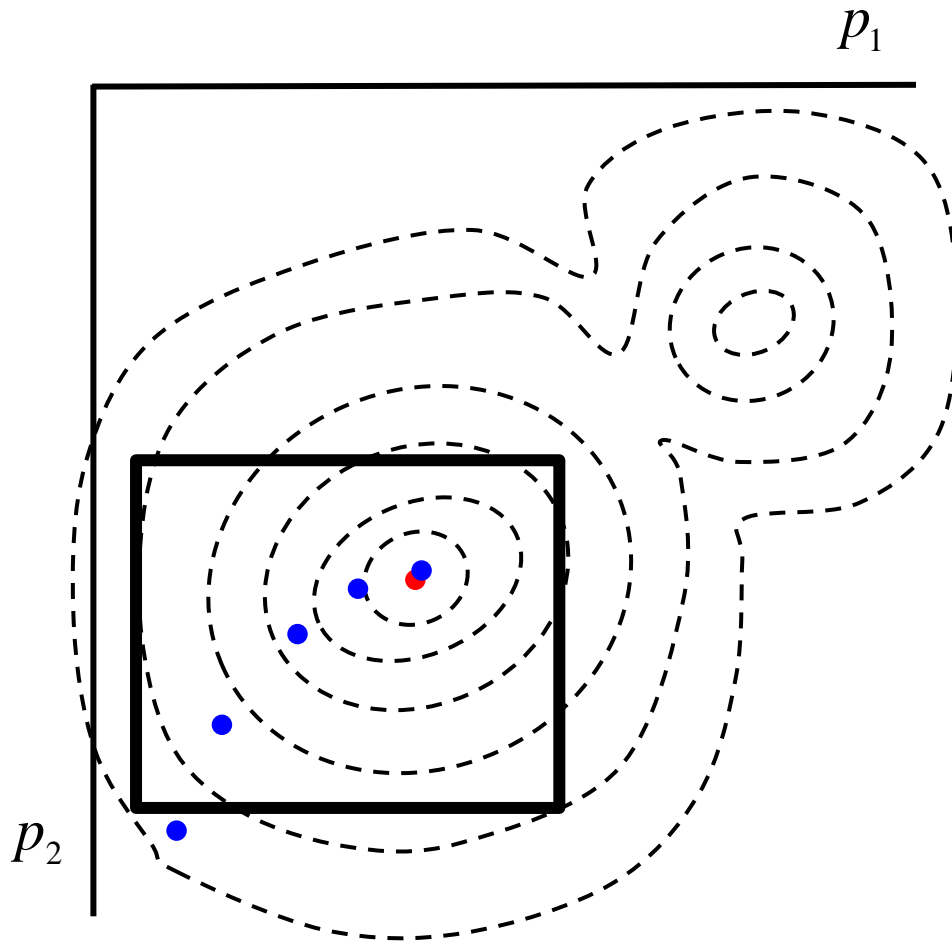
Introdução



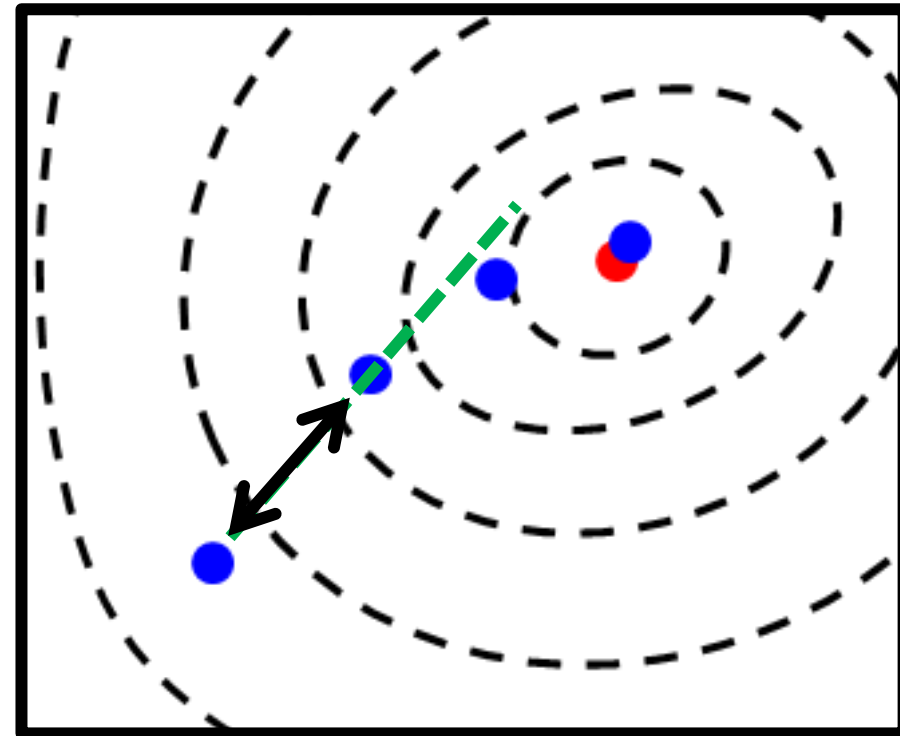
Direção



Introdução

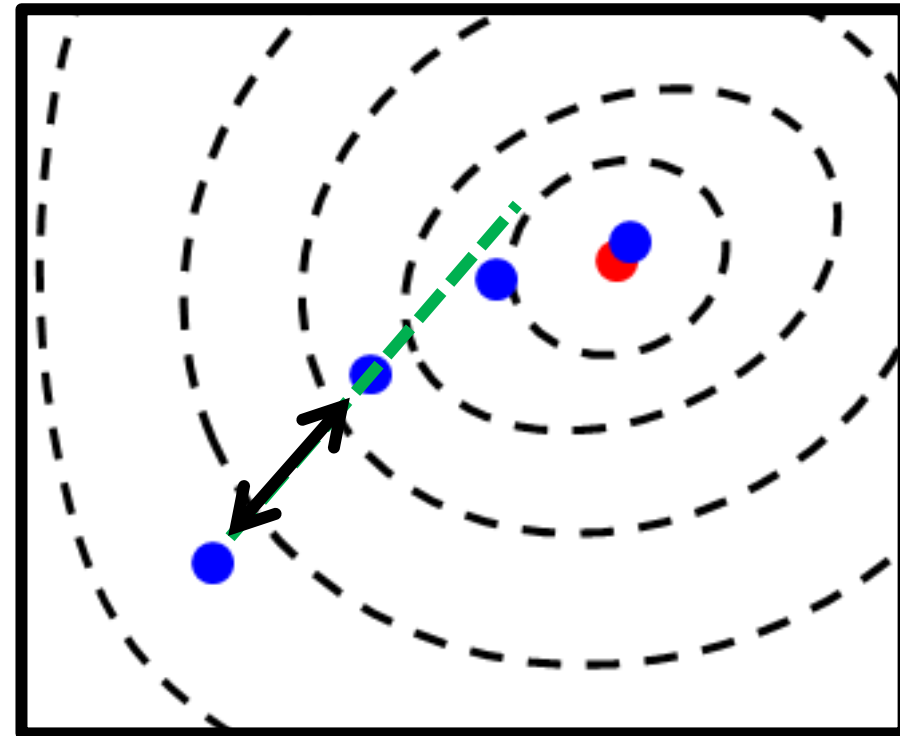


Tamanho do passo



Introdução

Existem dois grupos principais de métodos para estimar o mínimo de uma função:

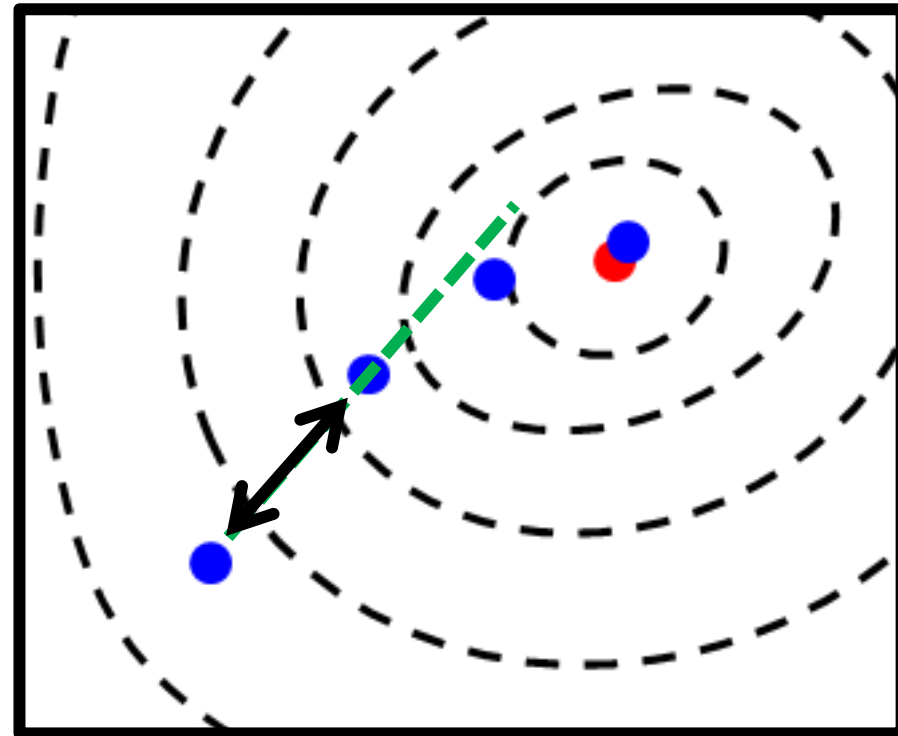


Introdução

Existem dois grupos principais de métodos para estimar o mínimo de uma função:

Métodos que se baseiam no gradiente da função

Métodos Heurísticos



Introdução

- Métodos baseados no gradiente
 - Método de Newton
 - Método de Gauss-Newton
 - Método de *Steepest Decent*
 - Método de Levenberg-Marquardt
- Métodos Heurísticos
 - *Simulated Annealing*
 - Método das Formigas

Métodos baseados no gradiente

$$\Omega(\bar{p})$$

Métodos baseados no gradiente

$$\Omega(\bar{p})$$

$$\Omega(\bar{p}_0 + \Delta\bar{p}) \approx \Omega(\bar{p}_0) + \bar{\nabla}\Omega(\bar{p}_0)^T \Delta\bar{p} + \frac{1}{2} \Delta\bar{p}^T \bar{\bar{\nabla}}\Omega(\bar{p}_0) \Delta\bar{p}$$

Métodos baseados no gradiente

$$\Omega(\bar{p})$$

$$\Omega(\bar{p}_0 + \Delta\bar{p}) \approx \Omega(\bar{p}_0) + \bar{\nabla}\Omega(\bar{p}_0)^T \Delta\bar{p} + \frac{1}{2} \Delta\bar{p}^T \bar{\bar{\nabla}}\Omega(\bar{p}_0) \Delta\bar{p}$$

$$\bar{\bar{\nabla}}\Omega(\bar{p}_0) \Delta\bar{p} = -\bar{\nabla}\Omega(\bar{p}_0)$$

Métodos baseados no gradiente

$$\Omega(\bar{p})$$

$$\Omega(\bar{p}_0 + \Delta\bar{p}) \approx \Omega(\bar{p}_0) + \bar{\nabla}\Omega(\bar{p}_0)^T \Delta\bar{p} + \frac{1}{2} \Delta\bar{p}^T \bar{\nabla}^2\Omega(\bar{p}_0) \Delta\bar{p}$$

$$\bar{\nabla}^2\Omega(\bar{p}_0) \Delta\bar{p} = -\bar{\nabla}\Omega(\bar{p}_0)$$

Diferença entre os métodos

Métodos baseados no gradiente

$$\Omega(\bar{p})$$

$$\Omega(\bar{p}_0 + \Delta\bar{p}) \approx \Omega(\bar{p}_0) + \bar{\nabla}\Omega(\bar{p}_0)^T \Delta\bar{p} + \frac{1}{2} \Delta\bar{p}^T \bar{\nabla}^2\Omega(\bar{p}_0) \Delta\bar{p}$$

$$\bar{\nabla}^2\Omega(\bar{p}_0) \Delta\bar{p} = -\bar{\nabla}\Omega(\bar{p}_0)$$

Newton

$$\bar{\nabla}^2\Omega(\bar{p}_0)$$

Gauss - Newton

$$\bar{M}(\bar{p}_0)$$

Steepest Decent

$$1/\eta$$

Levenberg - Marquardt

$$\bar{M}(\bar{p}_0) + \lambda \bar{I}$$

Diferença entre os métodos

Métodos baseados no gradiente

Vamos às contas!

Métodos baseados no gradiente

Método	Convergência
Steepest Decent	0
Levenberg - Marquardt	1
Gauss - Newton	2
Newton	3

0 – lento



3 – rápido

Métodos baseados no gradiente

Método	Aproximação inicial
Steepest Decent	Pode ser distante
Levenberg - Marquardt	Pode ser distante
Gauss - Newton	Deve ser próxima
Newton	Deve ser próxima

Métodos baseados no gradiente

Método	Direção
Steepest Decent	É a do gradiente
Levenberg - Marquardt	Entre a do gradiente e a do produto hessiana-gradiente
Gauss - Newton	Predita pelo produto hessiana-gradiente
Newton	Predita pelo produto hessiana-gradiente

Métodos baseados no gradiente

Método	Tamanho do passo
Steepest Decent	Depende de um parâmetro e do gradiente
Levenberg - Marquardt	Depende de um parâmetro, do gradiente e da hessiana
Gauss - Newton	Depende do gradiente e da hessiana
Newton	Depende do gradiente e da hessiana

Métodos baseados no gradiente

Método	Custo computacional
Steepest Decent	0
Levenberg - Marquardt	2
Gauss - Newton	1
Newton	3

0 – baixo



3 – alto

Métodos Heurísticos

(Simulated Annealing)

temp_inicial, temp_final, kappa, itmax, p_{inferior} , p_{superior} , p_{inicial}

iteracao = 0, $p_1 = p_{\text{inicial}}$, *calcula* $f(p_1)$, $f_{\text{minimo}} = f(p_1)$, $p_{\text{minimo}} = p_1$
temperatura = temp_inicial

FAÇA {

- iteracao = iteracao + 1
- *sorteia* p_2 , *calcula* $f(p_2)$
- SE $[f(p_2) < f_{\text{minimo}}] \rightarrow f_{\text{minimo}} = f(p_2)$, $p_{\text{minimo}} = p_2$
- *probabilidade de sobrevivência* ps
 - SE $[f(p_2) < f(p_1)] \rightarrow ps = 1$
 - SE $[f(p_2) \geq f(p_1)] \rightarrow ps = \exp \{[f(p_1) - f(p_2)]/temperatura\}$
- SE $\{(ps = 1) \text{ OU } [(ps \neq 1) \text{ E } (moeda \leq ps)]\} \rightarrow p_1 = p_2$
- temperatura = kappa*temperatura

} ENQUANTO $[(iteracao \leq itmax) \text{ E } (temperatura \geq temp_final)]$

Métodos Heurísticos

(Simulated Annealing)

temp_inicial, temp_final, kappa, itmax, p_{inferior} , p_{superior} , p_{inicial}

iteracao = 0, $p_1 = p_{\text{inicial}}$, *calcula* $f(p_1)$, $f_{\text{minimo}} = f(p_1)$, $p_{\text{minimo}} = p_1$
temperatura = temp_inicial

FAÇA {

- iteracao = iteracao + 1
- *sorteia* p_2 , *calcula* $f(p_2)$
- SE $[f(p_2) < f_{\text{minimo}}]$ $\rightarrow f_{\text{minimo}} = f(p_2)$, $p_{\text{minimo}} = p_2$
- *probabilidade de sobrevivência* ps
 - SE $[f(p_2) < f(p_1)] \rightarrow ps = 1$
 - SE $[f(p_2) \geq f(p_1)] \rightarrow ps = \exp \{[f(p_1) - f(p_2)]/temperatura\}$
- SE $\{(ps = 1) \text{ OU } [(ps \neq 1) \text{ E } (moeda \leq ps)]\} \rightarrow p_1 = p_2$
- temperatura = kappa*temperatura

} ENQUANTO $[(iteracao \leq itmax) \text{ E } (temperatura \geq temp_final)]$

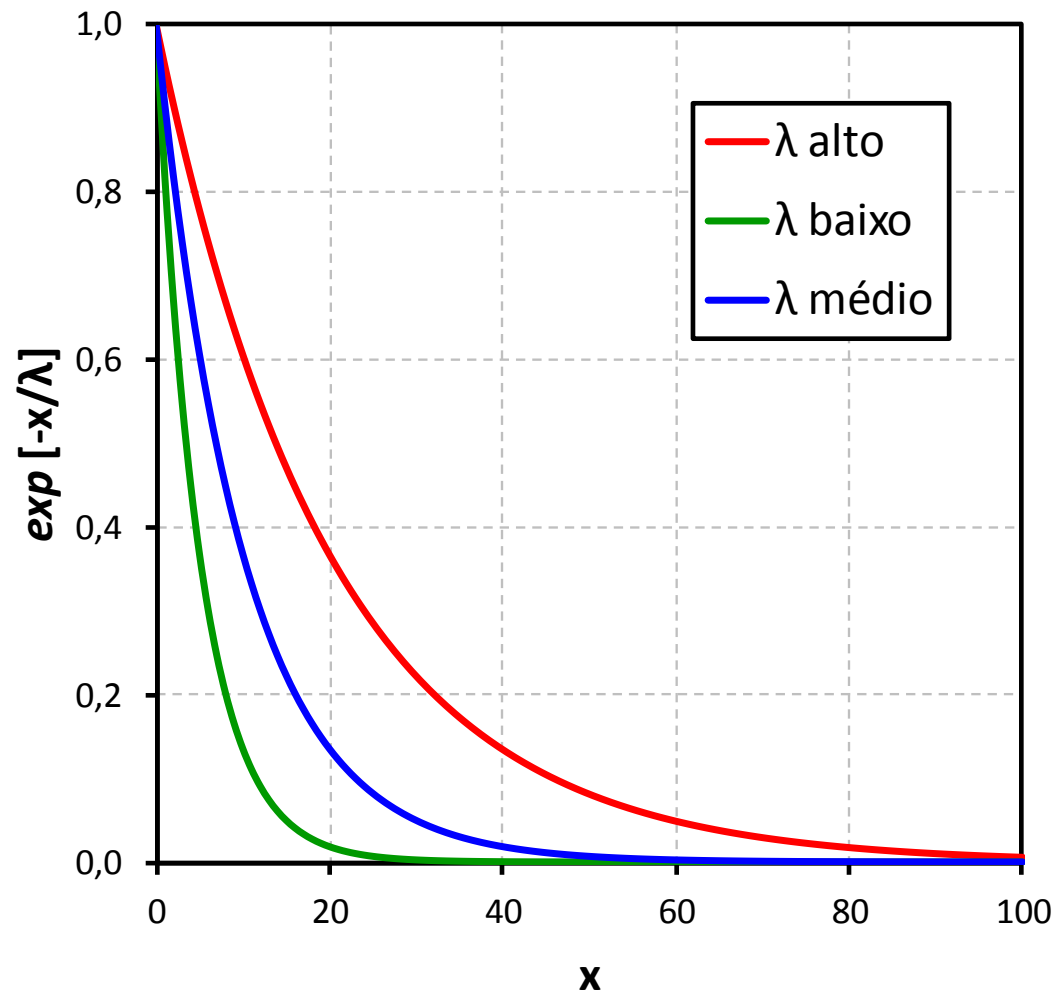
Métodos Heurísticos

(Simulated Annealing)

$$f(p_1) \leq f(p_2)$$

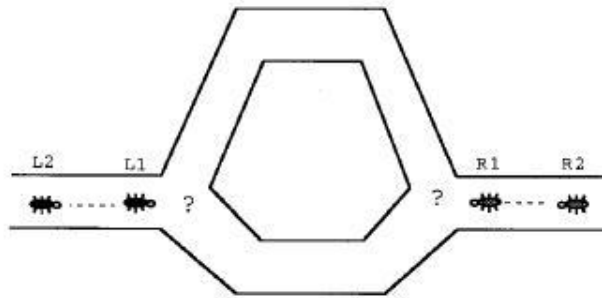
$$f(p_2) - f(p_1) = x$$

temperatura = λ

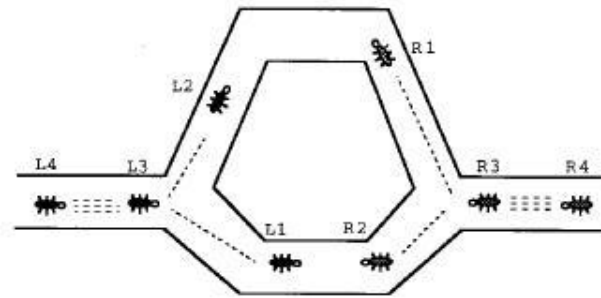


Métodos Heurísticos

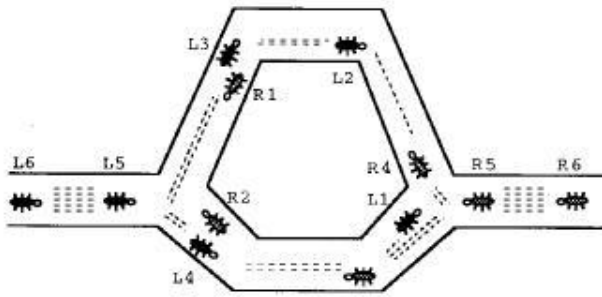
(Método das Formigas)



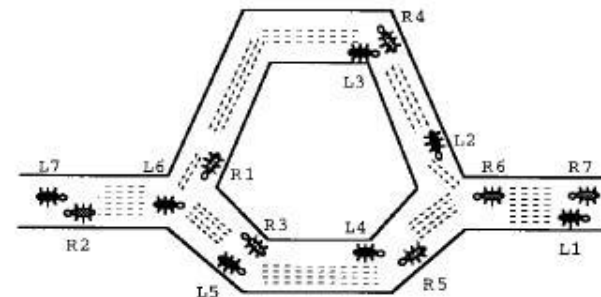
(a)



(b)



(c)



(d)

Como as formigas escolhem o caminho mais curto. Formigas liberam uma substância enquanto caminham e essa substância fica acumulada. (a) Formigas chegam em um ponto de decisão. (b) Algumas escolhem o caminho mais longo e outras escolhem o mais curto. A escolha é aleatória. (c) Como diferentes formigas andam praticamente com a mesma velocidade constante, aquelas que escolheram o caminho mais curto chegam do outro lado primeiro que as outras. Logo, em um dado período de tempo, o número de formigas que percorreram o caminho mais curto é maior. (d) O número de linhas pontilhadas representa a substância acumulada e esse acúmulo influencia o caminho que uma formiga irá seguir. Quanto mais substância acumulada, maior é a probabilidade de uma formiga escolher aquele caminho.