Otimização Contínua: Aspectos teóricos e computacionais

Ademir Alves Ribeiro
Elizabeth Wegner Karas

Capítulo 5 - Otimização irrestrita

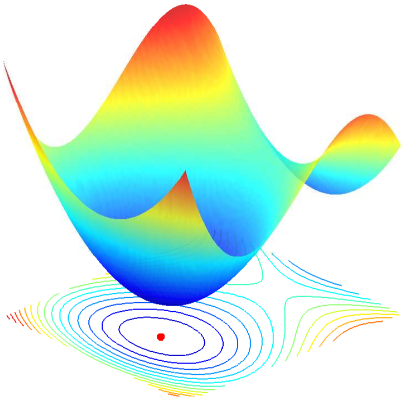
O problema geral de otimização irrestrito

O problema

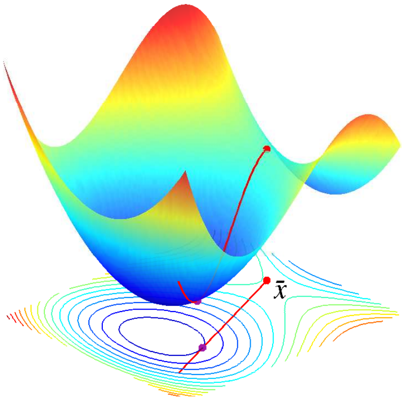
$$\begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeito a} & x \in \mathbb{R}^n \end{array}$$

- $f : \mathbb{R}^n \to \mathbb{R}$ uma função de classe $\mathcal{C}^2$

O problema geral de otimização irrestrito



Busca linear



O método do gradiente

Dado: $x^0 \in \mathbb{R}^n$

$k = 0$

REPITA enquanto $\nabla f(x^k) \neq 0$

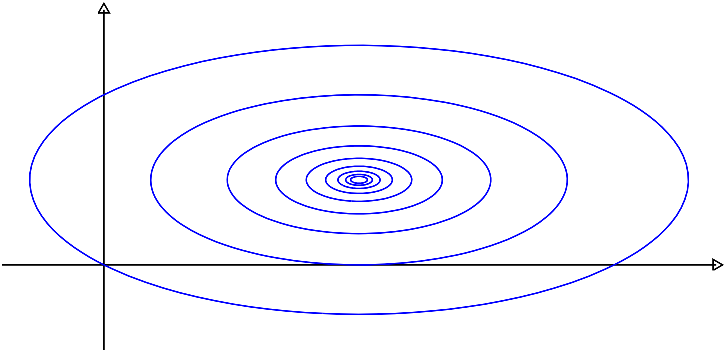
Defina $d^k = -\nabla f(x^k)$

Obtenha $t_k > 0$ tal que $f(x^k + t_k d^k) < f(x^k)$

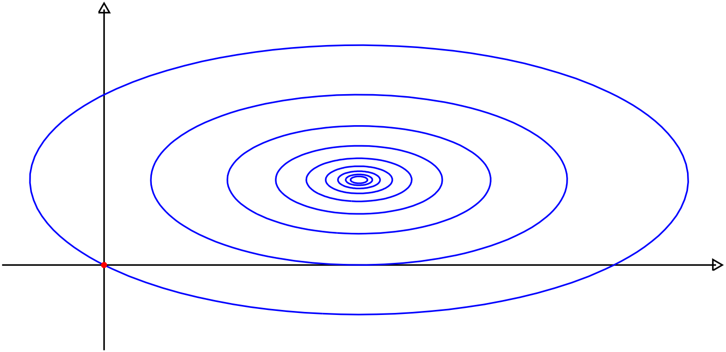
Faça $x^{k+1} = x^k + t_k d^k$

$k = k + 1$

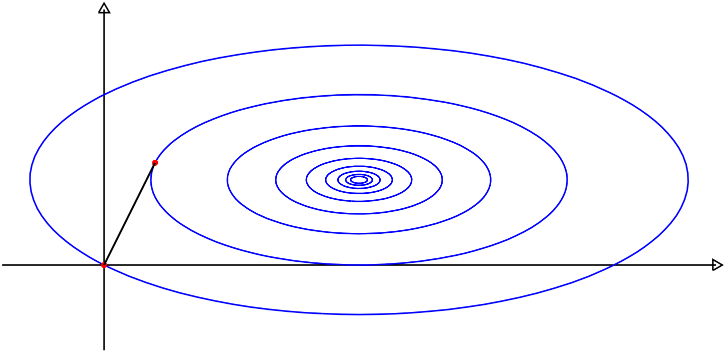
O método do gradiente



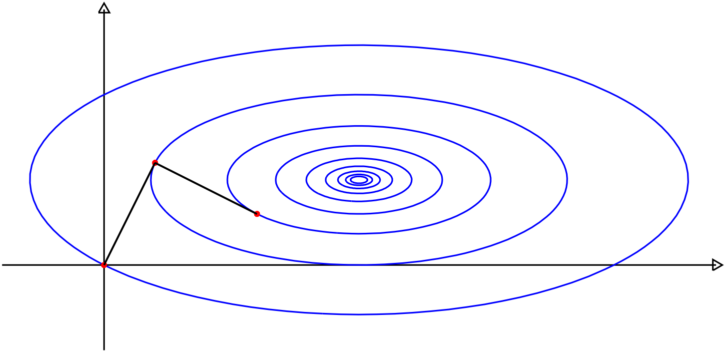
O método do gradiente



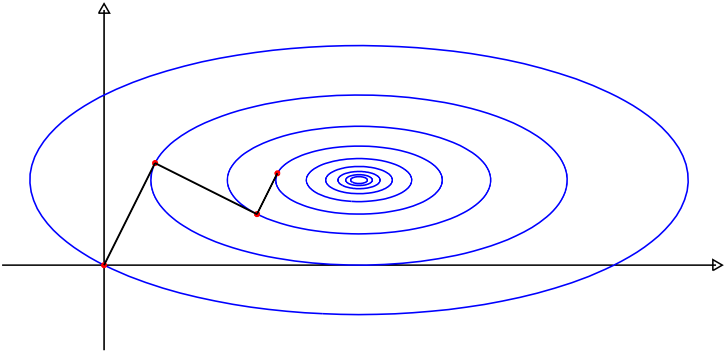
O método do gradiente



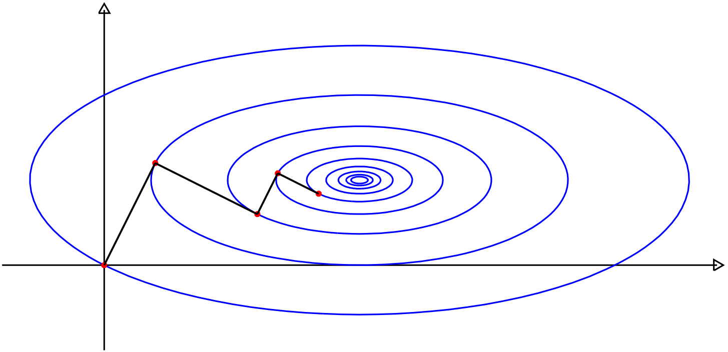
O método do gradiente



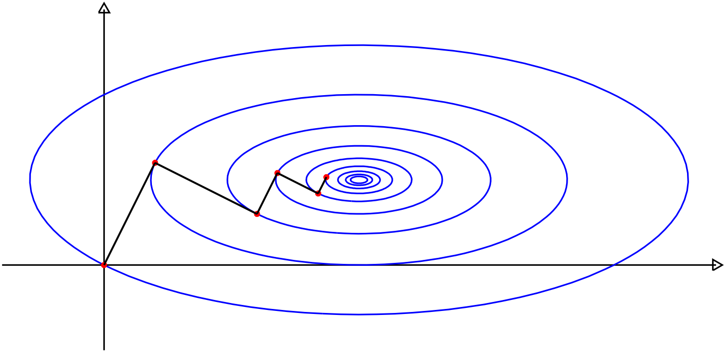
O método do gradiente



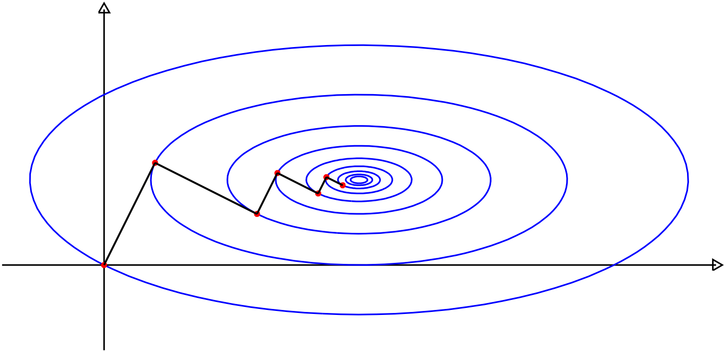
O método do gradiente



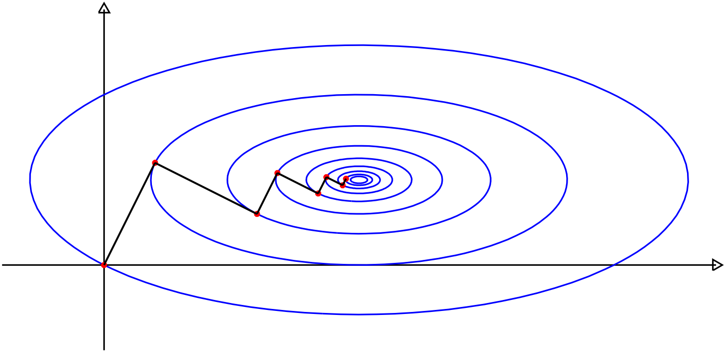
O método do gradiente



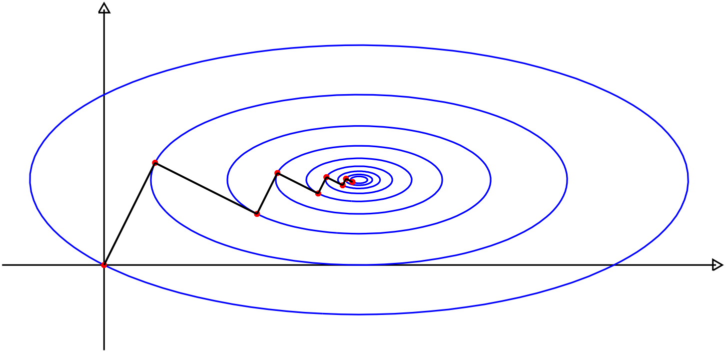
O método do gradiente



O método do gradiente



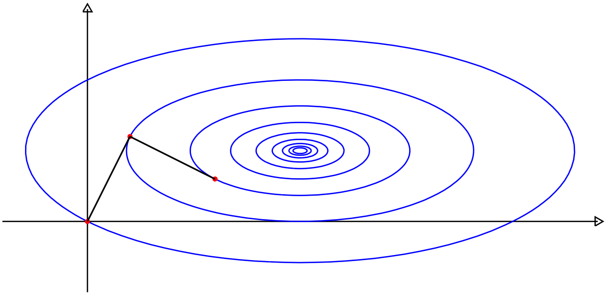
O método do gradiente



Lema

Se no método do gradiente o passo t_k é obtido por busca exata então

$$(d^{k+1})^T d^k = 0.$$



Método do gradiente

Convergência global

O algoritmo do gradiente com o tamanho do passo t_k calculado pela busca exata, é globalmente convergente, ou seja, todo ponto de acumulação da sequência gerada pelo algoritmo é um ponto estacionário.

O mesmo resultado vale se utilizarmos a busca de Armijo para calcular t_k.

Método do gradiente

Convergência global

O algoritmo do gradiente com o tamanho do passo t_k calculado pela busca exata, é globalmente convergente, ou seja, todo ponto de acumulação da sequência gerada pelo algoritmo é um ponto estacionário.
O mesmo resultado vale se utilizarmos a busca de Armijo para calcular t_k.

Dados $A \in \mathbb{R}^{n\times n}$ uma matriz definida positiva, $b \in \mathbb{R}^n$, $c \in \mathbb{R}$.

Função quadrática

$$f(x) = \frac{1}{2}x^T Ax + b^T x + c$$

Como $A \in \mathbb{R}^{n\times n}$ é definida positiva, f é convexa e tem um único minimizador x^*, que é global e satisfaz

$$Ax^* + b = \nabla f(x^*) = 0.$$

Velocidade de convergência

Teorema

$$\|x^{k+1} - x^*\|_2 \leq \sqrt{1 - \frac{\lambda_1}{\lambda_n}}\,\|x^k - x^*\|_2$$

Conjectura

$$\|x^{k+1} - x^*\|_2 \leq \left(1 - \frac{\lambda_1}{\lambda_n}\right)\|x^k - x^*\|_2$$

onde λ_1 é o menor e λ_n o maior autovalor de A.

Velocidade de convergência

Teorema

$$\|x^{k+1} - x^*\|_2 \leq \sqrt{1 - \frac{\lambda_1}{\lambda_n}} \|x^k - x^*\|_2$$

Conjectura

$$\|x^{k+1} - x^*\|_2 \leq \left(1 - \frac{\lambda_1}{\lambda_n}\right) \|x^k - x^*\|_2$$

onde λ_1 é o menor e λ_n o maior autovalor de A.

Velocidade de convergência

Teorema

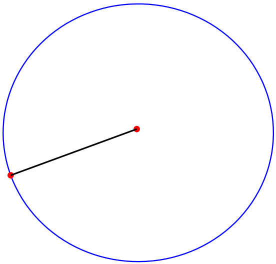
$$\|x^{k+1} - x^*\|_2 \leq \sqrt{1 - \frac{\lambda_1}{\lambda_n}} \|x^k - x^*\|_2$$

Conjectura verdadeira para $n = 2$

$$\|x^{k+1} - x^*\|_2 \leq \left(1 - \frac{\lambda_1}{\lambda_n}\right) \|x^k - x^*\|_2$$

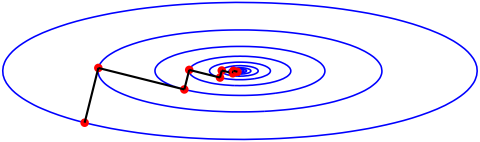
onde λ_1 é o menor e λ_n o maior autovalor de A.

Velocidade de convergência



$\lambda_1 = \lambda_n$

convergência em um único passo



$\lambda_1 \ll \lambda_n$

convergência mais lenta

Considere $x^* \in \mathbb{R}^n$ um minimizador local de f, com $\nabla^2 f(x^*)$ definida positiva. Suponha que a sequência (x^k) gerada pelo algoritmo do gradiente, com busca exata, converge para x^*. Então a sequência $\left(f(x^k)\right)$ converge linearmente para $f(x^*)$ com taxa não superior a

$$\left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^2,$$

onde λ_1 é o menor e λ_n o maior autovalor de $\nabla^2 f(x^*)$.

Método de Newton - Motivação

Considere $F : \mathbb{R}^n \to \mathbb{R}^n$ de classe $\mathcal{C}^1$ e o problema de resolver o sistema

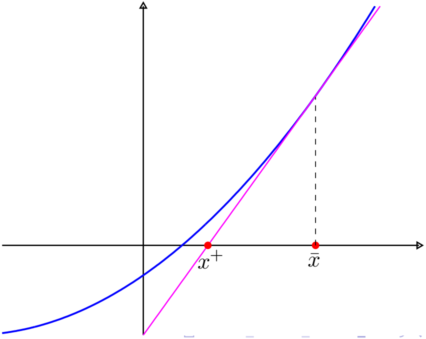
$$F(x) = 0.$$

A ideia é aproximar F por seu polinômio de Taylor de primeira ordem. Dado $\bar{x}$, considere o sistema linear

$$F(\bar{x}) + J_F(\bar{x})(x - \bar{x}) = 0,$$

onde J_F é a matriz Jacobiana de F.
Caso $J_F(\bar{x})$ seja inversível, a solução do sistema linear é

$$x^+ = \bar{x} - \left(J_F(\bar{x})\right)^{-1} F(\bar{x}).$$



Método de Newton - Motivação

Considere $F : \mathbb{R}^n \to \mathbb{R}^n$ de classe $\mathcal{C}^1$ e o problema de resolver o sistema

$$F(x) = 0.$$

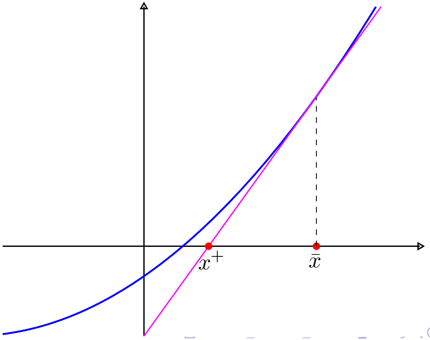
A ideia é aproximar F por seu polinômio de Taylor de primeira ordem. Dado $\bar{x}$, considere o sistema linear

$$F(\bar{x}) + J_F(\bar{x})(x - \bar{x}) = 0,$$

onde J_F é a matriz Jacobiana de F.

Caso $J_F(\bar{x})$ seja inversível, a solução do sistema linear é

$$x^+ = \bar{x} - \left(J_F(\bar{x})\right)^{-1} F(\bar{x}).$$



Método de Newton - Motivação

Considere $F : \mathbb{R}^n \to \mathbb{R}^n$ de classe $\mathcal{C}^1$ e o problema de resolver o sistema

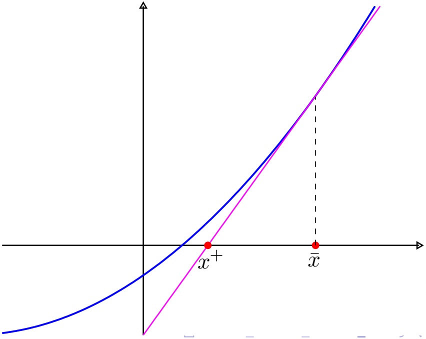
$$F(x) = 0.$$

A ideia é aproximar F por seu polinômio de Taylor de primeira ordem. Dado $\bar{x}$, considere o sistema linear

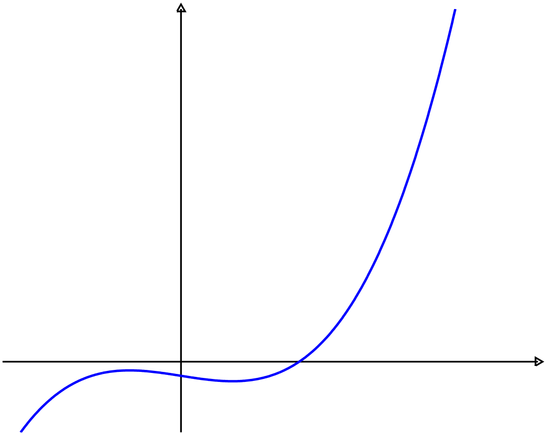
$$F(\bar{x}) + J_F(\bar{x})(x - \bar{x}) = 0,$$

onde J_F é a matriz Jacobiana de F.
Caso $J_F(\bar{x})$ seja inversível, a solução do sistema linear é

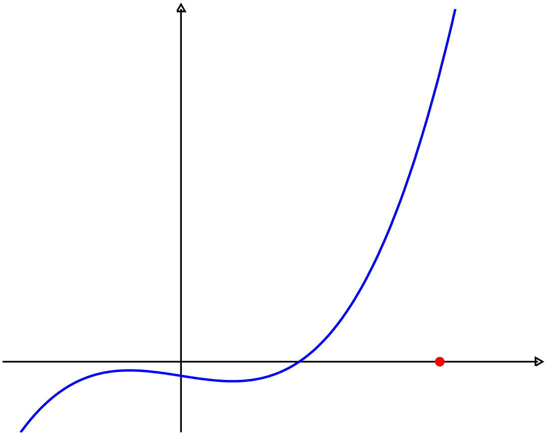
$$x^+ = \bar{x} - \left(J_F(\bar{x})\right)^{-1} F(\bar{x}).$$



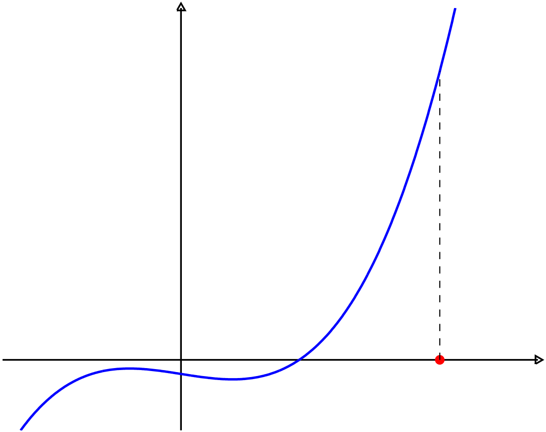
O Método de Newton para resolver $F(x) = 0$



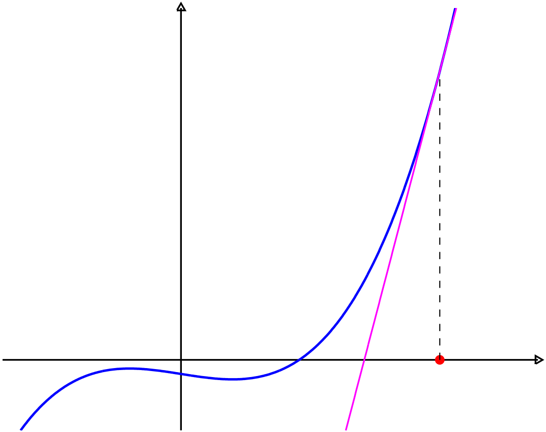
O Método de Newton para resolver $F(x) = 0$



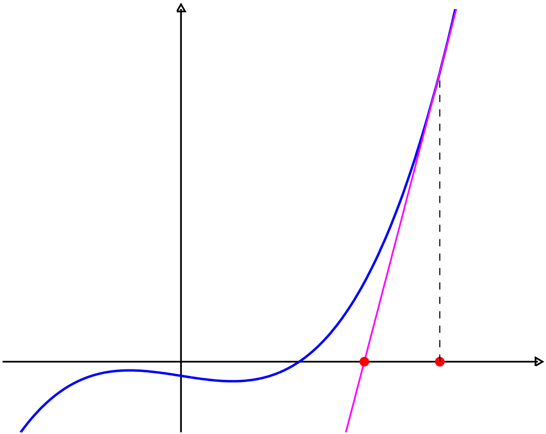
O Método de Newton para resolver $F(x) = 0$



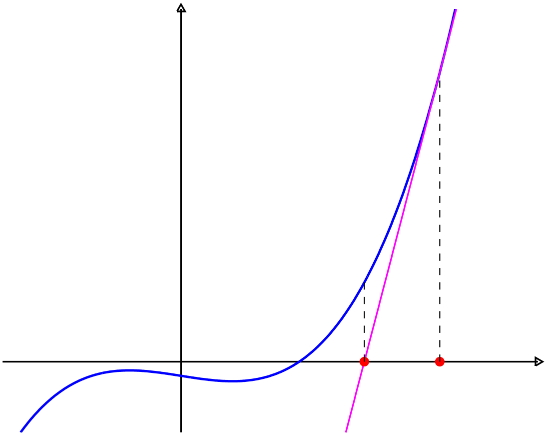
O Método de Newton para resolver $F(x) = 0$



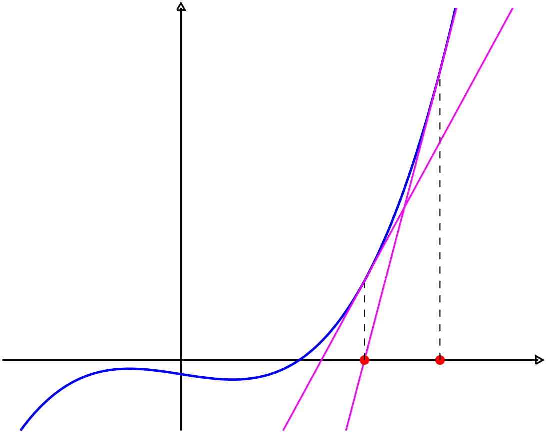
O Método de Newton para resolver $F(x) = 0$



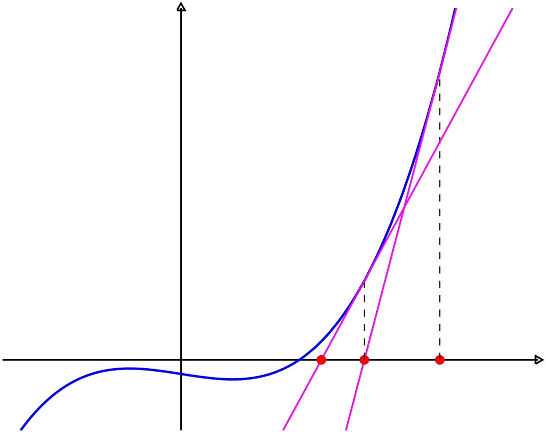
O Método de Newton para resolver $F(x) = 0$



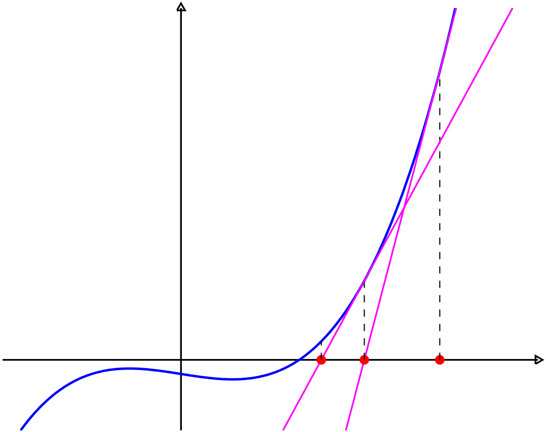
O Método de Newton para resolver $F(x) = 0$



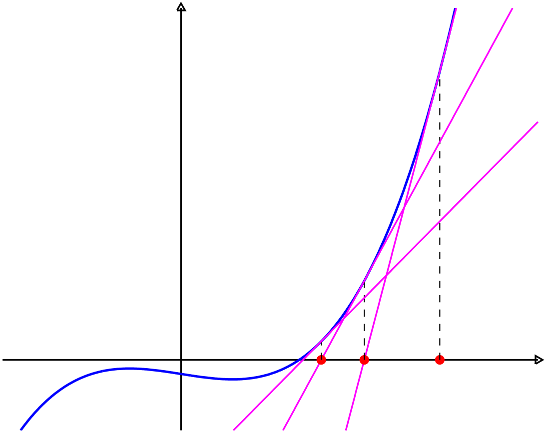
O Método de Newton para resolver $F(x) = 0$



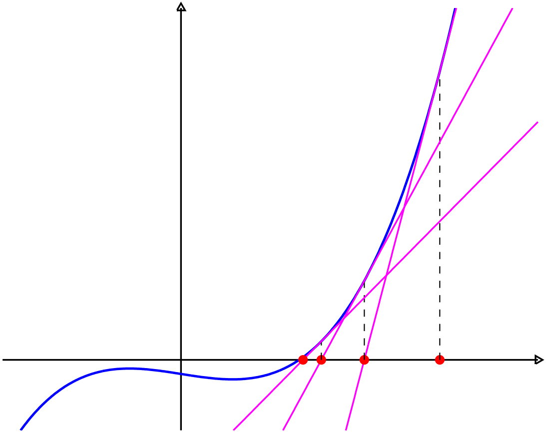
O Método de Newton para resolver $F(x) = 0$



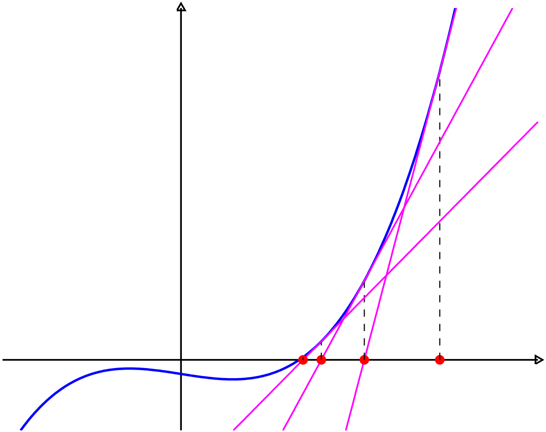
O Método de Newton para resolver $F(x) = 0$



O Método de Newton para resolver $F(x) = 0$



O Método de Newton para resolver $F(x) = 0$



- Para sistema de equações $F(x) = 0$:

$$x^+ = \bar{x} - \left(J_F(\bar{x})\right)^{-1} F(\bar{x}).$$

- Para minimizar $f(x)$:
  - Condição de otimalidade: $\nabla f(x) = 0 \qquad (F = \nabla f).$

$$x^+ = \bar{x} - \left(\nabla f^2(\bar{x})\right)^{-1} \nabla f(\bar{x}).$$

- Para sistema de equações $F(x) = 0$:

$$x^+ = \bar{x} - \big(J_F(\bar{x})\big)^{-1} F(\bar{x}).$$

- Para minimizar $f(x)$:
  - Condição de otimalidade: $\nabla f(x) = 0$ $\qquad (F = \nabla f)$.

$$x^+ = \bar{x} - \big(\nabla f^2(\bar{x})\big)^{-1} \nabla f(\bar{x}).$$

Método de Newton

- Para sistema de equações $F(x) = 0$:

$$x^+ = \bar{x} - \left(J_F(\bar{x})\right)^{-1} F(\bar{x}).$$

- Para minimizar $f(x)$:
  - Condição de otimalidade: $\nabla f(x) = 0 \qquad (F = \nabla f)$.

$$x^+ = \bar{x} - \left(\nabla f^2(\bar{x})\right)^{-1} \nabla f(\bar{x}).$$

Método de Newton

- Para sistema de equações $F(x) = 0$:

$$x^+ = \bar{x} - \left(J_F(\bar{x})\right)^{-1} F(\bar{x}).$$

- Para minimizar $f(x)$:
  - Condição de otimalidade: $\nabla f(x) = 0$ $\quad (F = \nabla f).$

$$x^+ = \bar{x} - \left(\nabla f^2(\bar{x})\right)^{-1} \nabla f(\bar{x}).$$

Método de Newton

- Para sistema de equações $F(x) = 0$:

$$x^+ = \bar{x} - \left(J_F(\bar{x})\right)^{-1} F(\bar{x}).$$

- Para minimizar $f(x)$:
  - Condição de otimalidade: $\nabla f(x) = 0 \qquad (F = \nabla f)$.

$$x^+ = \bar{x} - \left(\nabla f^2(\bar{x})\right)^{-1} \nabla f(\bar{x}).$$

Método de Newton

- Para sistema de equações $F(x) = 0$:

$$x^+ = \bar{x} - \left(J_F(\bar{x})\right)^{-1} F(\bar{x}).$$

- Para minimizar $f(x)$:
  - Condição de otimalidade: $\nabla f(x) = 0 \qquad (F = \nabla f)$.

$$x^+ = \bar{x} - \left(\nabla f^2(\bar{x})\right)^{-1} \nabla f(\bar{x}).$$

Algoritmo do método de Newton para minimização

Dado: $x^0 \in \mathbb{R}^n$
$k = 0$
REPITA enquanto $\nabla f(x^k) \neq 0$
    Defina $d^k = -\left(\nabla^2 f(x^k)\right)^{-1} \nabla f(x^k)$
    Determine o tamanho do passo $t_k > 0$
    Faça $x^{k+1} = x^k + t_k d^k$
    $k = k + 1$

Método de Newton

- O cálculo da direção d^k é feito resolvendo-se o sistema de equações lineares
$$\nabla^2 f(x^k)d = -\nabla f(x^k),$$
que tem um custo computacional menor do que o gasto para inverter uma matriz.
- A direção de Newton d^k nem sempre é uma direção de descida. Se $\nabla^2 f(x^k)$ é definida positiva, então d^k é uma direção de descida.

Interpretação geométrica do passo de Newton

Dado $x^k \in \mathbb{R}^n$, considere a aproximação de Taylor de segunda ordem de f, dada por

$$p(x) = f(x^k) + \nabla f(x^k)^T (x - x^k) + \frac{1}{2}(x - x^k)^T \nabla^2 f(x^k)(x - x^k).$$

Com o objetivo de minimizar p, fazemos

$$\nabla f(x^k) + \nabla^2 f(x^k)(x - x^k) = \nabla p(x) = 0,$$

obtendo exatamente o passo de Newton d^k.

Desta forma, se $\nabla^2 f(x^k)$ é definida positiva, então
o passo de Newton minimiza o modelo quadrático de f em torno de x^k.

Dado $x^k \in \mathbb{R}^n$, considere a aproximação de Taylor de segunda ordem de f, dada por

$$p(x) = f(x^k) + \nabla f(x^k)^T (x - x^k) + \frac{1}{2}(x - x^k)^T \nabla^2 f(x^k)(x - x^k).$$

Com o objetivo de minimizar p, fazemos

$$\nabla f(x^k) + \nabla^2 f(x^k)(x - x^k) = \nabla p(x) = 0,$$

obtendo exatamente o passo de Newton d^k.

Desta forma, se $\nabla^2 f(x^k)$ é definida positiva, então
o passo de Newton minimiza o modelo quadrático de f em torno de x^k.

Interpretação geométrica do passo de Newton

Dado $x^k \in \mathbb{R}^n$, considere a aproximação de Taylor de segunda ordem de f, dada por

$$p(x) = f(x^k) + \nabla f(x^k)^T (x - x^k) + \frac{1}{2}(x - x^k)^T \nabla^2 f(x^k)(x - x^k).$$

Com o objetivo de minimizar p, fazemos

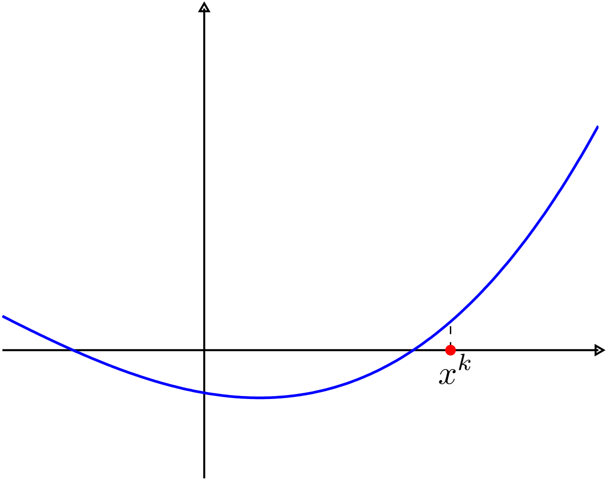
$$\nabla f(x^k) + \nabla^2 f(x^k)(x - x^k) = \nabla p(x) = 0,$$

obtendo exatamente o passo de Newton d^k.

Desta forma, se $\nabla^2 f(x^k)$ é definida positiva, então
o passo de Newton minimiza o modelo quadrático de f em torno de x^k.

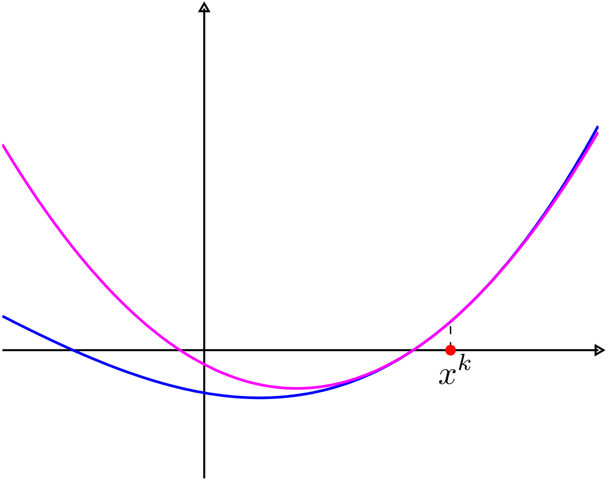
Interpretação geométrica do passo de Newton

O passo de Newton minimiza o modelo quadrático de f em torno de x^k.



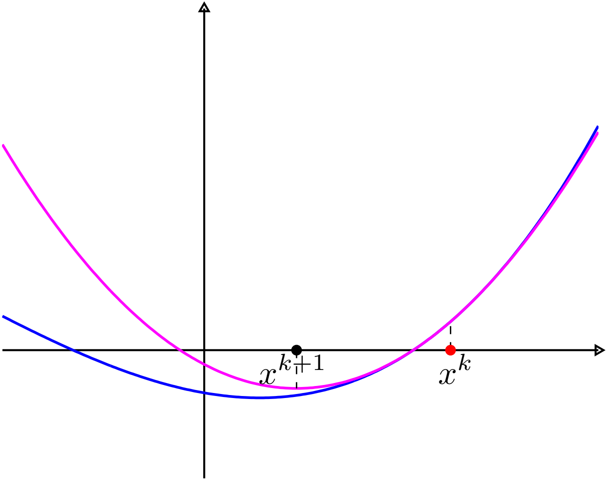
Interpretação geométrica do passo de Newton

O passo de Newton minimiza o modelo quadrático de f em torno de x^k.



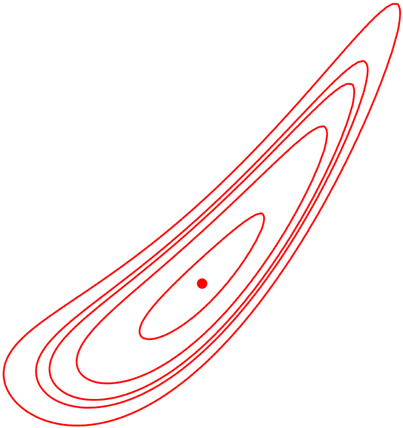
Interpretação geométrica do passo de Newton

O passo de Newton minimiza o modelo quadrático de f em torno de x^k.



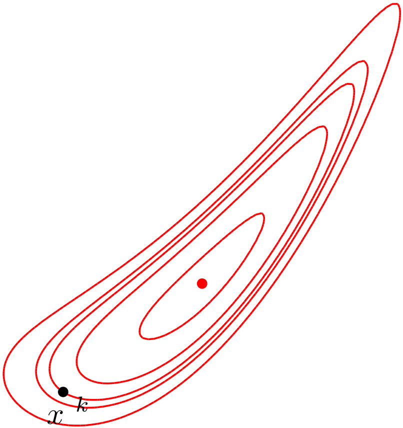
Interpretação geométrica do passo de Newton

O passo de Newton minimiza o modelo quadrático de f em torno de x^k.



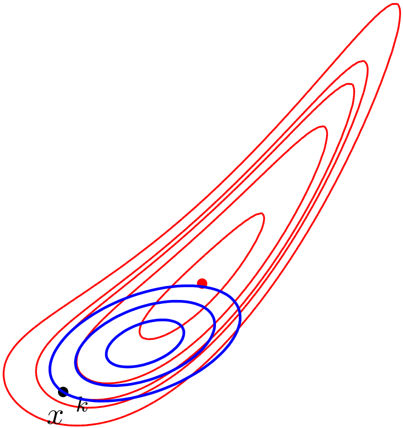
Interpretação geométrica do passo de Newton

O passo de Newton minimiza o modelo quadrático de f em torno de x^k.



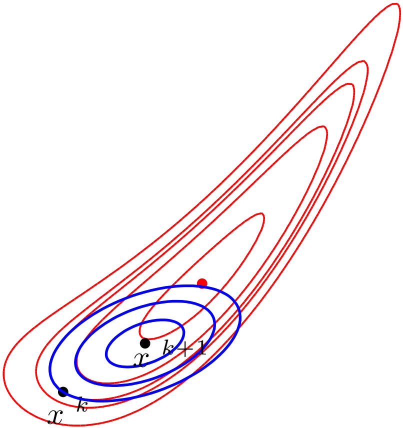
Interpretação geométrica do passo de Newton

O passo de Newton minimiza o modelo quadrático de f em torno de x^k.



Interpretação geométrica do passo de Newton

O passo de Newton minimiza o modelo quadrático de f em torno de x^k.



Propriedade do método de Newton

O método de Newton minimiza uma função quadrática em um único passo.

Considere

$$f(x) = \frac{1}{2}x^T Ax + b^T x + c,$$

com $A \in \mathbb{R}^{n\times n}$ definida positiva, $b \in \mathbb{R}^n$ e $c \in \mathbb{R}$.
Dado $x^0 \in \mathbb{R}^n$, o passo de Newton é

$$d^0 = -\left(\nabla^2 f(x^0)\right)^{-1} \nabla f(x^0) = -A^{-1}(Ax^0 + b) = -x^0 - A^{-1}b.$$

Assim,

$$x^1 = x^0 + d^0 = -A^{-1}b = x^*.$$

Propriedade do método de Newton

O método de Newton minimiza uma função quadrática em um único passo.

Considere

$$f(x) = \frac{1}{2}x^T Ax + b^T x + c,$$

com $A \in \mathbb{R}^{n\times n}$ definida positiva, $b \in \mathbb{R}^n$ e $c \in \mathbb{R}$.
Dado $x^0 \in \mathbb{R}^n$, o passo de Newton é

$$d^0 = -\left(\nabla^2 f(x^0)\right)^{-1}\nabla f(x^0) = -A^{-1}(Ax^0 + b) = -x^0 - A^{-1}b.$$

Assim,

$$x^1 = x^0 + d^0 = -A^{-1}b = x^*.$$

Propriedade do método de Newton

O método de Newton minimiza uma função quadrática em um único passo.

Considere

$$f(x) = \frac{1}{2}x^T Ax + b^T x + c,$$

com $A \in \mathbb{R}^{n\times n}$ definida positiva, $b \in \mathbb{R}^n$ e $c \in \mathbb{R}$.
Dado $x^0 \in \mathbb{R}^n$, o passo de Newton é

$$d^0 = -\left(\nabla^2 f(x^0)\right)^{-1} \nabla f(x^0) = -A^{-1}(Ax^0 + b) = -x^0 - A^{-1}b.$$

Assim,

$$x^1 = x^0 + d^0 = -A^{-1}b = x^*.$$

CENGAGE Learning

Teorema

Suponha que $\nabla^2 f(x)$ é definida positiva, para todo $x \in \mathbb{R}^n$. Então o Algoritmo de Newton, com o tamanho do passo t_k calculado pela busca exata, é globalmente convergente. O mesmo resultado vale se utilizarmos a busca de Armijo para calcular t_k.

Convergência local do método de Newton

Teorema

Seja $f : \mathbb{R}^n \to \mathbb{R}$ de classe $\mathcal{C}^2$. Suponha que $x^* \in \mathbb{R}^n$ seja um minimizador local de f, com $\nabla^2 f(x^*)$ definida positiva. Então existe $\delta > 0$ tal que se $x^0 \in B(x^*, \delta)$, o Algoritmo de Newton, aplicado com $t_k = 1$ para todo $k \in \mathbb{N}$, gera uma sequência (x^k) tal que:

(i) $\nabla^2 f(x^k)$ é definida positiva, para todo $k \in \mathbb{N}$;

(ii) (x^k) converge superlinearmente para x^*;

(iii) Se $\nabla^2 f$ é Lipschitz, então a convergência é quadrática.

Convergência local do método de Newton

Teorema

Seja $f : \mathbb{R}^n \to \mathbb{R}$ de classe $\mathcal{C}^2$. Suponha que $x^* \in \mathbb{R}^n$ seja um minimizador local de f, com $\nabla^2 f(x^*)$ definida positiva. Então existe $\delta > 0$ tal que se $x^0 \in B(x^*, \delta)$, o Algoritmo de Newton, aplicado com $t_k = 1$ para todo $k \in \mathbb{N}$, gera uma sequência (x^k) tal que:

(i) $\nabla^2 f(x^k)$ é definida positiva, para todo $k \in \mathbb{N}$;

(ii) (x^k) converge superlinearmente para x^*;

(iii) Se $\nabla^2 f$ é Lipschitz, então a convergência é quadrática.

Convergência local do método de Newton

Teorema

Seja $f : \mathbb{R}^n \to \mathbb{R}$ de classe $\mathcal{C}^2$. Suponha que $x^* \in \mathbb{R}^n$ seja um minimizador local de f, com $\nabla^2 f(x^*)$ definida positiva. Então existe $\delta > 0$ tal que se $x^0 \in B(x^*, \delta)$, o Algoritmo de Newton, aplicado com $t_k = 1$ para todo $k \in \mathbb{N}$, gera uma sequência (x^k) tal que:

(i) $\nabla^2 f(x^k)$ é definida positiva, para todo $k \in \mathbb{N}$;

(ii) (x^k) converge superlinearmente para x^*;

(iii) Se $\nabla^2 f$ é Lipschitz, então a convergência é quadrática.

Convergência local do método de Newton

Teorema

Seja $f : \mathbb{R}^n \to \mathbb{R}$ de classe $\mathcal{C}^2$. Suponha que $x^* \in \mathbb{R}^n$ seja um minimizador local de f, com $\nabla^2 f(x^*)$ definida positiva. Então existe $\delta > 0$ tal que se $x^0 \in B(x^*, \delta)$, o Algoritmo de Newton, aplicado com $t_k = 1$ para todo $k \in \mathbb{N}$, gera uma sequência (x^k) tal que:

(i) $\nabla^2 f(x^k)$ é definida positiva, para todo $k \in \mathbb{N}$;

(ii) (x^k) converge superlinearmente para x^*;

(iii) Se $\nabla^2 f$ é Lipschitz, então a convergência é quadrática.

Comparação

Teorema

- Método de Newton
  - convergência quadrática
  - alto custo computacional, pois faz uso de derivadas de segunda ordem.
- Método do gradiente
  - convergência linear
  - menor custo computacional

Comparação

Teorema

- Método de Newton
  - convergência quadrática
  - alto custo computacional, pois faz uso de derivadas de segunda ordem.
- Método do gradiente
  - convergência linear
  - menor custo computacional

Comparação

Teorema

- Método de Newton
  - convergência quadrática
  - alto custo computacional, pois faz uso de derivadas de segunda ordem.
- Método do gradiente
  - convergência linear
  - menor custo computacional

Comparação

Teorema

- Método de Newton
  - convergência quadrática
  - alto custo computacional, pois faz uso de derivadas de segunda ordem.
- Método do gradiente
  - convergência linear
  - menor custo computacional

Comparação

Teorema

- Método de Newton
  - convergência quadrática
  - alto custo computacional, pois faz uso de derivadas de segunda ordem.
- Método do gradiente
  - convergência linear
  - menor custo computacional

Comparação

Teorema

- Método de Newton
  - convergência quadrática
  - alto custo computacional, pois faz uso de derivadas de segunda ordem.
- Método do gradiente
  - convergência linear
  - menor custo computacional

Considere $A \in \mathbb{R}^{n \times n}$ uma matriz definida positiva.

Definição

Os vetores $d^0, d^1, \ldots, d^k \in \mathbb{R}^n \setminus \{0\}$ são A-conjugados se

$$(d^i)^T A d^j = 0,$$

para todos $i, j = 0, 1, \ldots, k$, com $i \neq j$.

Em particular, se A é a matriz identidade, vetores A-conjugados são ortogonais no sentido usual.

Um conjunto qualquer de vetores A-conjugados é linearmente independente.

Vetores A-conjugados

Considere $A \in \mathbb{R}^{n \times n}$ uma matriz definida positiva.

Definição

Os vetores $d^0, d^1, \ldots, d^k \in \mathbb{R}^n \setminus \{0\}$ são A-conjugados se

$$(d^i)^T A d^j = 0,$$

para todos $i, j = 0, 1, \ldots, k$, com $i \neq j$.

Em particular, se A é a matriz identidade, vetores A-conjugados são ortogonais no sentido usual.

Um conjunto qualquer de vetores A-conjugados é linearmente independente.

Considere $A \in \mathbb{R}^{n\times n}$ uma matriz definida positiva.

Definição

Os vetores $d^0, d^1, \ldots, d^k \in \mathbb{R}^n \setminus \{0\}$ são A-conjugados se

$$(d^i)^T A d^j = 0,$$

para todos $i, j = 0, 1, \ldots, k$, com $i \neq j$.

Em particular, se A é a matriz identidade, vetores A-conjugados são ortogonais no sentido usual.

Um conjunto qualquer de vetores A-conjugados é linearmente independente.

Busca exata para funções quadráticas

- Função quadrática $f(x) = \frac{1}{2}x^T Ax + b^T x + c.$
- Queremos uma expressão para o tamanho do passo da busca exata:

$$t_k = \text{argmin}_{t\in\mathbb{R}} \left\{ f(x^k + td^k) \right\}.$$

Defina $\varphi : \mathbb{R} \to \mathbb{R}$ por $\varphi(t) = f(x^k + td^k)$. Assim,

$$0 = \varphi'(t_k) = \nabla f(x^k + t_k d^k)^T d^k = \nabla f(x^{k+1})^T d^k.$$

Mas

$$\nabla f(x^{k+1}) = A(x^k + t_k d^k) + b = \nabla f(x^k) + t_k Ad^k.$$

Logo,

$$t_k = -\frac{\nabla f(x^k)^T d^k}{(d^k)^T Ad^k}.$$

Busca exata para funções quadráticas

- Função quadrática $f(x) = \frac{1}{2}x^T Ax + b^T x + c$.
- Queremos uma expressão para o tamanho do passo da busca exata:

$$t_k = \text{argmin}_{t \in \mathbb{R}} \left\{ f(x^k + td^k) \right\}.$$

Defina $\varphi : \mathbb{R} \to \mathbb{R}$ por $\varphi(t) = f(x^k + td^k)$. Assim,

$$0 = \varphi'(t_k) = \nabla f(x^k + t_k d^k)^T d^k = \nabla f(x^{k+1})^T d^k.$$

Mas

$$\nabla f(x^{k+1}) = A(x^k + t_k d^k) + b = \nabla f(x^k) + t_k Ad^k.$$

Logo,

$$t_k = -\frac{\nabla f(x^k)^T d^k}{(d^k)^T Ad^k}.$$

CENGAGE Learning

Busca exata para funções quadráticas

- Função quadrática $f(x) = \frac{1}{2}x^T A x + b^T x + c.$
- Queremos uma expressão para o tamanho do passo da busca exata:

$$t_k = \text{argmin}_{t \in \mathbb{R}} \left\{ f(x^k + t d^k) \right\}.$$

Defina $\varphi : \mathbb{R} \to \mathbb{R}$ por $\varphi(t) = f(x^k + t d^k)$. Assim,

$$0 = \varphi'(t_k) = \nabla f(x^k + t_k d^k)^T d^k = \nabla f(x^{k+1})^T d^k.$$

Mas

$$\nabla f(x^{k+1}) = A(x^k + t_k d^k) + b = \nabla f(x^k) + t_k A d^k.$$

Logo,

$$t_k = -\frac{\nabla f(x^k)^T d^k}{(d^k)^T A d^k}.$$

Busca exata para funções quadráticas

- Função quadrática $f(x) = \frac{1}{2}x^T A x + b^T x + c.$
- Queremos uma expressão para o tamanho do passo da busca exata:

$$t_k = \text{argmin}_{t \in \mathbb{R}} \left\{ f(x^k + t d^k) \right\}.$$

Defina $\varphi : \mathbb{R} \to \mathbb{R}$ por $\varphi(t) = f(x^k + t d^k)$. Assim,

$$0 = \varphi'(t_k) = \nabla f(x^k + t_k d^k)^T d^k = \nabla f(x^{k+1})^T d^k.$$

Mas

$$\nabla f(x^{k+1}) = A(x^k + t_k d^k) + b = \nabla f(x^k) + t_k A d^k.$$

Logo,

$$t_k = -\frac{\nabla f(x^k)^T d^k}{(d^k)^T A d^k}.$$

CENGAGE Learning

Dado $x^0 \in \mathbb{R}^n$, defina $d^0 = -\nabla f(x^0)$ e, para $k = 0, 1, \ldots, n-2$,

$$d^{k+1} = -\nabla f(x^{k+1}) + \beta_k d^k,$$

onde $x^{k+1} = x^k + t_k d^k$, $t_k = -\dfrac{\nabla f(x^k)^T d^k}{(d^k)^T A d^k}$ e

β_k calculado de modo que d^k e d^{k+1} sejam A-conjugadas:

$$(d^k)^T A\big(-\nabla f(x^{k+1}) + \beta_k d^k\big) = (d^k)^T A d^{k+1} = 0,$$

o que fornece

$$\beta_k = \frac{(d^k)^T A \nabla f(x^{k+1})}{(d^k)^T A d^k}.$$

Dado $x^0 \in \mathbb{R}^n$, faça $d^0 = -\nabla f(x^0)$

$k = 0$

REPITA enquanto $\nabla f(x^k) \neq 0$

$$t_k = -\frac{\nabla f(x^k)^T d^k}{(d^k)^T A d^k}$$

$$x^{k+1} = x^k + t_k d^k$$

$$\beta_k = \frac{(d^k)^T A \nabla f(x^{k+1})}{(d^k)^T A d^k}$$

$$d^{k+1} = -\nabla f(x^{k+1}) + \beta_k d^k$$

$$k = k + 1$$

Propriedades

Se x^k e d^k foram gerados pelo algoritmo de gradientes conjugadas, então

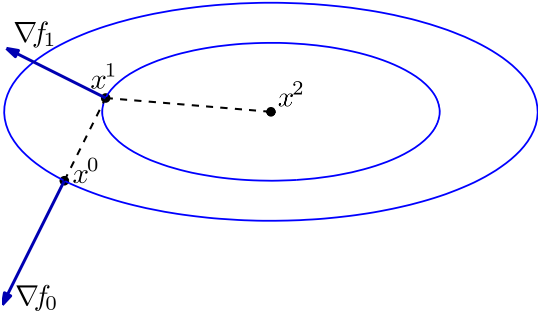
$$\nabla f(x^k)^T \nabla f(x^j) = 0 \quad \text{e} \quad (d^k)^T A d^j = 0,$$

para todo $j = 0, 1, \dots, k-1$.

Teorema

O método de gradientes conjugados quando aplicado para minimizar uma função quadrática estritamente convexa encontra o minimizador em no máximo n iterações.

Método de gradientes conjugados



Note:

- ortogonalidade dos gradientes nos iterandos.
- solução obtida em 2 passos de uma função definida em $\mathbb{R}^2$.

Outras formas de calcular o coeficiente β_k

- Expressão que obtivemos:

$$\beta_k = \frac{(d^k)^T A \nabla f(x^{k+1})}{(d^k)^T A d^k}$$

- Devida a Polak e Ribière:

$$\beta_k^{PR} = \frac{\nabla f(x^{k+1})^T \left(\nabla f(x^{k+1}) - \nabla f(x^k)\right)}{\nabla f(x^k)^T \nabla f(x^k)}$$

- Devida a Fletcher e Reeves:

$$\beta_k^{FR} = \frac{\nabla f(x^{k+1})^T \nabla f(x^{k+1})}{\nabla f(x^k)^T \nabla f(x^k)}.$$

Tais expressões coincidem no caso quadrático.
Para funções não quadráticas tais expressões podem não ser iguais, o que fornece variantes do método de gradientes conjugados.

Outras formas de calcular o coeficiente β_k

- Expressão que obtivemos:

$$\beta_k = \frac{(d^k)^T A \nabla f(x^{k+1})}{(d^k)^T A d^k}$$

- Devida a Polak e Ribière:

$$\beta_k^{PR} = \frac{\nabla f(x^{k+1})^T \left(\nabla f(x^{k+1}) - \nabla f(x^k)\right)}{\nabla f(x^k)^T \nabla f(x^k)}$$

- Devida a Fletcher e Reeves:

$$\beta_k^{FR} = \frac{\nabla f(x^{k+1})^T \nabla f(x^{k+1})}{\nabla f(x^k)^T \nabla f(x^k)}.$$

Tais expressões coincidem no caso quadrático.
Para funções não quadráticas tais expressões podem não ser iguais, o que fornece variantes do método de gradientes conjugados.

- Expressão que obtivemos:

$$\beta_k = \frac{(d^k)^T A \nabla f(x^{k+1})}{(d^k)^T A d^k}$$

- Devida a Polak e Ribière:

$$\beta_k^{\text{PR}} = \frac{\nabla f(x^{k+1})^T \left(\nabla f(x^{k+1}) - \nabla f(x^k)\right)}{\nabla f(x^k)^T \nabla f(x^k)}$$

- Devida a Fletcher e Reeves:

$$\beta_k^{\text{FR}} = \frac{\nabla f(x^{k+1})^T \nabla f(x^{k+1})}{\nabla f(x^k)^T \nabla f(x^k)}.$$

Tais expressões coincidem no caso quadrático.
Para funções não quadráticas tais expressões podem não ser iguais, o que fornece variantes do método de gradientes conjugados.

Outras formas de calcular o coeficiente β_k

- Expressão que obtivemos:

$$\beta_k = \frac{(d^k)^T A \nabla f(x^{k+1})}{(d^k)^T A d^k}$$

- Devida a Polak e Ribière:

$$\beta_k^{PR} = \frac{\nabla f(x^{k+1})^T \left(\nabla f(x^{k+1}) - \nabla f(x^k)\right)}{\nabla f(x^k)^T \nabla f(x^k)}$$

- Devida a Fletcher e Reeves:

$$\beta_k^{FR} = \frac{\nabla f(x^{k+1})^T \nabla f(x^{k+1})}{\nabla f(x^k)^T \nabla f(x^k)}.$$

Tais expressões coincidem no caso quadrático.

Para funções não quadráticas tais expressões podem não ser iguais, o que fornece variantes do método de gradientes conjugados.

Outras formas de calcular o coeficiente β_k

- Expressão que obtivemos:

$$\beta_k = \frac{(d^k)^T A \nabla f(x^{k+1})}{(d^k)^T A d^k}$$

- Devida a Polak e Ribière:

$$\beta_k^{\text{PR}} = \frac{\nabla f(x^{k+1})^T \left(\nabla f(x^{k+1}) - \nabla f(x^k)\right)}{\nabla f(x^k)^T \nabla f(x^k)}$$

- Devida a Fletcher e Reeves:

$$\beta_k^{\text{FR}} = \frac{\nabla f(x^{k+1})^T \nabla f(x^{k+1})}{\nabla f(x^k)^T \nabla f(x^k)}.$$

Tais expressões coincidem no caso quadrático.
Para funções não quadráticas tais expressões podem não ser iguais, o que fornece variantes do método de gradientes conjugados.

Dado $x^0 \in \mathbb{R}^n$, faça $d^0 = -\nabla f(x^0)$

$k = 0$

REPITA enquanto $\nabla f(x^k) \neq 0$

  Calcule o comprimento do passo t_k

  Faça $x^{k+1} = x^k + t_k d^k$

  SE $(k+1) \bmod n \neq 0$ Reinicialização a cada n iterações

    Calcule β_k por β_k^{PR} ou β_k^{PR}

  SENÃO

    $\beta_k = 0$

  Defina $d^{k+1} = -\nabla f(x^{k+1}) + \beta_k d^k$

  $k = k+1$

Resultado

O método de gradientes conjugados para minimização de funções quadráticas tem complexidade algorítmica da ordem $O\left(\frac{1}{k^2}\right)$.

Considere a função quadrática $f : \mathbb{R}^n \to \mathbb{R}$ dada por

$$f(x) = \frac{1}{2}x^T Ax + b^T x + c,$$

com $A \in \mathbb{R}^{n\times n}$ definida positiva, $b \in \mathbb{R}^n$ e $c \in \mathbb{R}$.

Definição

Dados $x^0 \in \mathbb{R}^n$ e $k \in \mathbb{N}$, o k-ésimo espaço de Krylov é dado por

$$\mathcal{K}_k = [A(x^0 - x^*), A^2(x^0 - x^*), \ldots, A^k(x^0 - x^*)].$$

Como $Ax^* + b = \nabla f(x^*) = 0$, tem-se que $A(x^0 - x^*) = Ax^0 + b = \nabla f(x^0)$, e o espaço de Krylov se escreve como

$$\mathcal{K}_k = [\nabla f(x^0), A\nabla f(x^0), \ldots, A^{k-1}\nabla f(x^0)].$$

Espaços de Krylov

Teorema

Considere as sequências (x^k) e (d^k), geradas pelo Algoritmo de Gradientes Conjugados. Se o método não termina em x^{k-1}, então

(i) $\mathcal{K}_k = [\nabla f(x^0), \nabla f(x^1), \ldots, \nabla f(x^{k-1})]$;

(ii) $\mathcal{K}_k = [d^0, d^1, \ldots, d^{k-1}]$.

Teorema

Considere as sequências (x^k) e (d^k), geradas pelo Algoritmo de Gradientes Conjugados. Se o método não termina em x^{k-1}, então

(i) $\mathcal{K}_k = [\nabla f(x^0), \nabla f(x^1), \ldots, \nabla f(x^{k-1})]$;

(ii) $\mathcal{K}_k = [d^0, d^1, \ldots, d^{k-1}]$.

Espaços de Krylov

Teorema

Considere as sequências (x^k) e (d^k), geradas pelo Algoritmo de Gradientes Conjugados. Se o método não termina em x^{k-1}, então

(i) $\mathcal{K}_k = [\nabla f(x^0), \nabla f(x^1), \ldots, \nabla f(x^{k-1})]$;

(ii) $\mathcal{K}_k = [d^0, d^1, \ldots, d^{k-1}]$.

Espaços de Krylov

$$\mathcal{P}_k = \left\{1 + a_1 t + a_2 t^2 + \cdots + a_k t^k \mid a_i \in \mathbb{R}, i = 1, \ldots, k\right\}.$$

Lema

Considere $V_k = x^0 + \mathcal{K}_k$. Temos $x \in V_k$ se, e somente se,

$$x - x^* = p(A)(x^0 - x^*),$$

para algum polinômio $p \in \mathcal{P}_k$.

Lema

Considere $x^k = \text{argmin}_{x \in V_k} \{f(x)\}$. Então, para todo polinômio $p \in \mathcal{P}_k$,

$$f(x^k) - f(x^*) \leq \frac{1}{2}(x^0 - x^*)^T A \big(p(A)\big)^2 (x^0 - x^*).$$

$$\mathcal{P}_k = \left\{1 + a_1 t + a_2 t^2 + \cdots + a_k t^k \mid a_i \in \mathbb{R}, i = 1, \ldots, k\right\}.$$

Lema

Considere $V_k = x^0 + \mathcal{K}_k$. Temos $x \in V_k$ se, e somente se,

$$x - x^* = p(A)(x^0 - x^*),$$

para algum polinômio $p \in \mathcal{P}_k$.

Lema

Considere $x^k = \operatorname{argmin}_{x \in V_k} \{f(x)\}$. Então, para todo polinômio $p \in \mathcal{P}_k$,

$$f(x^k) - f(x^*) \leq \frac{1}{2}(x^0 - x^*)^T A \big(p(A)\big)^2 (x^0 - x^*).$$

$$\mathcal{P}_k = \left\{1 + a_1 t + a_2 t^2 + \cdots + a_k t^k \mid a_i \in \mathbb{R}, i = 1, \ldots, k\right\}.$$

Lema

Considere $V_k = x^0 + \mathcal{K}_k$. Temos $x \in V_k$ se, e somente se,

$$x - x^* = p(A)(x^0 - x^*),$$

para algum polinômio $p \in \mathcal{P}_k$.

Lema

Considere $x^k = \text{argmin}_{x \in V_k} \{f(x)\}$. Então, para todo polinômio $p \in \mathcal{P}_k$,

$$f(x^k) - f(x^*) \leq \frac{1}{2}(x^0 - x^*)^T A \big(p(A)\big)^2 (x^0 - x^*).$$

Polinômios de Chebyshev

O polinômio de Chebyshev de grau k, $T_k : [-1,1] \to \mathbb{R}$, é definido por

$$T_k(t) = \cos\left(k \arccos(t)\right).$$

Para todo $t \in [-1,1]$,

- $T_0(t) = 1$
- $T_1(t) = t$
- $T_{k+1}(t) = 2tT_k(t) - T_{k-1}(t)$, para todo $k \geq 1$.

Polinômios de Chebyshev

O polinômio de Chebyshev de grau k, $T_k : [-1,1] \to \mathbb{R}$, é definido por

$$T_k(t) = \cos\big(k \arccos(t)\big).$$

Para todo $t \in [-1,1]$,

- $T_0(t) = 1$
- $T_1(t) = t$
- $T_{k+1}(t) = 2tT_k(t) - T_{k-1}(t)$, para todo $k \geq 1$.

O polinômio de Chebyshev de grau k, $T_k : [-1,1] \to \mathbb{R}$, é definido por

$$T_k(t) = \cos\big(k \arccos(t)\big).$$

Para todo $t \in [-1,1]$,

- $T_0(t) = 1$
- $T_1(t) = t$
- $T_{k+1}(t) = 2tT_k(t) - T_{k-1}(t)$, para todo $k \geq 1$.

CENGAGE Learning

O polinômio de Chebyshev de grau k, $T_k : [-1,1] \to \mathbb{R}$, é definido por

$$T_k(t) = \cos\big(k \arccos(t)\big).$$

Para todo $t \in [-1,1]$,

- $T_0(t) = 1$
- $T_1(t) = t$
- $T_{k+1}(t) = 2tT_k(t) - T_{k-1}(t)$, para todo $k \geq 1$.

O polinômio de Chebyshev de grau k, $T_k : [-1, 1] \to \mathbb{R}$, é definido por

$$T_k(t) = \cos\big(k \arccos(t)\big).$$

Para todo $t \in [-1, 1]$,

- $T_0(t) = 1$
- $T_1(t) = t$
- $T_{k+1}(t) = 2tT_k(t) - T_{k-1}(t)$, para todo $k \geq 1$.

Polinômios de Chebyshev

Exemplo

Determine os polinômios de Chebyshev T_k, com $k = 0, 1, \ldots, 6$ e faça o gráfico para $k = 1, \ldots, 4$.

Para todo $t \in [-1, 1]$,

- $T_0(t) = 1$
- $T_1(t) = t$
- $T_2(t) = 2t^2 - 1$
- $T_3(t) = 4t^3 - 3t$
- $T_4(t) = 8t^4 - 8t^2 + 1$
- $T_5(t) = 16t^5 - 20t^3 + 5t$
- $T_6(t) = 32t^6 - 48t^4 + 18t^2 - 1.$

Exemplo

Determine os polinômios de Chebyshev T_k, com $k = 0, 1, \dots, 6$ e faça o gráfico para $k = 1, \dots, 4$.

Para todo $t \in [-1, 1]$,

- $T_0(t) = 1$
- $T_1(t) = t$
- $T_2(t) = 2t^2 - 1$
- $T_3(t) = 4t^3 - 3t$
- $T_4(t) = 8t^4 - 8t^2 + 1$
- $T_5(t) = 16t^5 - 20t^3 + 5t$
- $T_6(t) = 32t^6 - 48t^4 + 18t^2 - 1.$

CENGAGE Learning

Exemplo

Determine os polinômios de Chebyshev T_k, com $k = 0, 1, \ldots, 6$ e faça o gráfico para $k = 1, \ldots, 4$.

Para todo $t \in [-1, 1]$,

- $T_0(t) = 1$
- $T_1(t) = t$
- $T_2(t) = 2t^2 - 1$
- $T_3(t) = 4t^3 - 3t$
- $T_4(t) = 8t^4 - 8t^2 + 1$
- $T_5(t) = 16t^5 - 20t^3 + 5t$
- $T_6(t) = 32t^6 - 48t^4 + 18t^2 - 1.$

Exemplo

Determine os polinômios de Chebyshev T_k, com $k = 0, 1, \ldots, 6$ e faça o gráfico para $k = 1, \ldots, 4$.

Para todo $t \in [-1, 1]$,

- $T_0(t) = 1$
- $T_1(t) = t$
- $T_2(t) = 2t^2 - 1$
- $T_3(t) = 4t^3 - 3t$
- $T_4(t) = 8t^4 - 8t^2 + 1$
- $T_5(t) = 16t^5 - 20t^3 + 5t$
- $T_6(t) = 32t^6 - 48t^4 + 18t^2 - 1.$

Exemplo

Determine os polinômios de Chebyshev T_k, com $k = 0, 1, \ldots, 6$ e faça o gráfico para $k = 1, \ldots, 4$.

Para todo $t \in [-1, 1]$,

- $T_0(t) = 1$
- $T_1(t) = t$
- $T_2(t) = 2t^2 - 1$
- $T_3(t) = 4t^3 - 3t$
- $T_4(t) = 8t^4 - 8t^2 + 1$
- $T_5(t) = 16t^5 - 20t^3 + 5t$
- $T_6(t) = 32t^6 - 48t^4 + 18t^2 - 1.$

Exemplo

Determine os polinômios de Chebyshev T_k, com $k = 0, 1, \ldots, 6$ e faça o gráfico para $k = 1, \ldots, 4$.

Para todo $t \in [-1, 1]$,

- $T_0(t) = 1$
- $T_1(t) = t$
- $T_2(t) = 2t^2 - 1$
- $T_3(t) = 4t^3 - 3t$
- $T_4(t) = 8t^4 - 8t^2 + 1$
- $T_5(t) = 16t^5 - 20t^3 + 5t$
- $T_6(t) = 32t^6 - 48t^4 + 18t^2 - 1.$

Exemplo

Determine os polinômios de Chebyshev T_k, com $k = 0, 1, \ldots, 6$ e faça o gráfico para $k = 1, \ldots, 4$.

Para todo $t \in [-1, 1]$,

- $T_0(t) = 1$
- $T_1(t) = t$
- $T_2(t) = 2t^2 - 1$
- $T_3(t) = 4t^3 - 3t$
- $T_4(t) = 8t^4 - 8t^2 + 1$
- $T_5(t) = 16t^5 - 20t^3 + 5t$
- $T_6(t) = 32t^6 - 48t^4 + 18t^2 - 1.$

Exemplo

Determine os polinômios de Chebyshev T_k, com $k = 0, 1, \ldots, 6$ e faça o gráfico para $k = 1, \ldots, 4$.

Para todo $t \in [-1, 1]$,

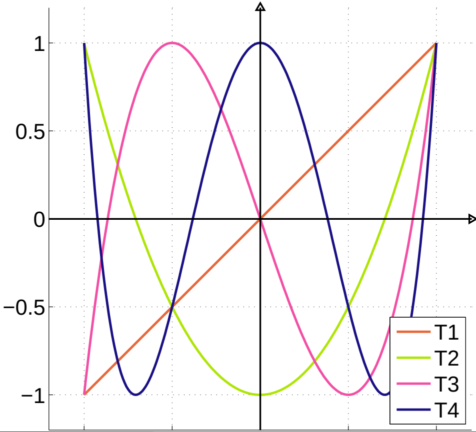
- $T_0(t) = 1$
- $T_1(t) = t$
- $T_2(t) = 2t^2 - 1$
- $T_3(t) = 4t^3 - 3t$
- $T_4(t) = 8t^4 - 8t^2 + 1$
- $T_5(t) = 16t^5 - 20t^3 + 5t$
- $T_6(t) = 32t^6 - 48t^4 + 18t^2 - 1.$

Exemplo

Determine os polinômios de Chebyshev T_k, com $k = 0, 1, \ldots, 6$ e faça o gráfico para $k = 1, \ldots, 4$.

Para todo $t \in [-1, 1]$,

- $T_0(t) = 1$
- $T_1(t) = t$
- $T_2(t) = 2t^2 - 1$
- $T_3(t) = 4t^3 - 3t$
- $T_4(t) = 8t^4 - 8t^2 + 1$
- $T_5(t) = 16t^5 - 20t^3 + 5t$
- $T_6(t) = 32t^6 - 48t^4 + 18t^2 - 1.$



Polinômios de Chebyshev

Propriedades

- $\|T_k\| = \sup\{|T_k(t)| \mid t \in [-1,1]\}$
- T_k é uma função par (ímpar) quando k é par (ímpar).

Lema

Sejam $L > 0$ e $k \in \mathbb{N}$. Então existe $p \in \mathcal{P}_k$ tal que, para todo $t \in [0, L]$,

$$T_{2k+1}\left(\frac{\sqrt{t}}{\sqrt{L}}\right) = (-1)^k(2k+1)\frac{\sqrt{t}}{\sqrt{L}}p(t).$$

Polinômios de Chebyshev

Propriedades

- $\|T_k\| = \sup\{|T_k(t)| \mid t \in [-1,1]\}$
- T_k é uma função par (ímpar) quando k é par (ímpar).

Lema

Sejam $L > 0$ e $k \in \mathbb{N}$. Então existe $p \in \mathcal{P}_k$ tal que, para todo $t \in [0, L]$,

$$T_{2k+1}\left(\frac{\sqrt{t}}{\sqrt{L}}\right) = (-1)^k (2k+1)\frac{\sqrt{t}}{\sqrt{L}} p(t).$$

Propriedades

- $\|T_k\| = \sup\{|T_k(t)| \mid t \in [-1,1]\}$
- T_k é uma função par (ímpar) quando k é par (ímpar).

Lema

Sejam $L > 0$ e $k \in \mathbb{N}$. Então existe $p \in \mathcal{P}_k$ tal que, para todo $t \in [0, L]$,

$$T_{2k+1}\left(\frac{\sqrt{t}}{\sqrt{L}}\right) = (-1)^k(2k+1)\frac{\sqrt{t}}{\sqrt{L}}p(t).$$

Considere a sequência (x^k) gerada pelo algoritmo de gradientes conjugados para minimizar uma quadrática com Hessiana A. Então,

$$f(x^k) - f(x^*) \leq \frac{L\|x^0 - x^*\|^2}{2(2k+1)^2},$$

onde x^* é o minimizador de f e L o maior autovalor de A.

- Considere $d^0, d^1, \ldots, d^{k-1}$ as direções geradas pelo algoritmo.
- x^k é o minimizador de f na variedade afim $x^0 + [d^0, d^1, \ldots, d^{k-1}] = V_k$.
- $f(x^k) - f(x^*) \leq \dfrac{1}{2}\|x^0 - x^*\|^2 \left\| A\big(p(A)\big)^2 \right\|$, para todo $p \in \mathcal{P}_k$.
- $\left\| A\big(p(A)\big)^2 \right\| \leq \max_{t\in[0,L]} \left\{ t\big(p(t)\big)^2 \right\} = \dfrac{L}{(2k+1)^2} \max_{t\in[0,L]} \left\{ T^2_{2k+1}\left(\dfrac{\sqrt{t}}{\sqrt{L}}\right) \right\}$
- $\max_{t\in[0,L]} \left\{ T^2_{2k+1}\left(\dfrac{\sqrt{t}}{\sqrt{L}}\right) \right\} \leq 1$

Complexidade algorítmica

Considere a sequência (x^k) gerada pelo algoritmo de gradientes conjugados para minimizar uma quadrática com Hessiana A. Então,

$$f(x^k) - f(x^*) \leq \frac{L\|x^0 - x^*\|^2}{2(2k+1)^2},$$

onde x^* é o minimizador de f e L o maior autovalor de A.

- Considere $d^0, d^1, \ldots, d^{k-1}$ as direções geradas pelo algoritmo.
- x^k é o minimizador de f na variedade afim $x^0 + [d^0, d^1, \ldots, d^{k-1}] = V_k$.
- $f(x^k) - f(x^*) \leq \frac{1}{2}\|x^0 - x^*\|^2 \left\| A\left(p(A)\right)^2 \right\|$, para todo $p \in \mathcal{P}_k$.
- $\left\| A\left(p(A)\right)^2 \right\| \leq \max_{t \in [0,L]} \left\{ t\left(p(t)\right)^2 \right\} = \frac{L}{(2k+1)^2} \max_{t \in [0,L]} \left\{ T_{2k+1}^2\left(\frac{\sqrt{t}}{\sqrt{L}}\right) \right\}$
- $\max_{t \in [0,L]} \left\{ T_{2k+1}^2\left(\frac{\sqrt{t}}{\sqrt{L}}\right) \right\} \leq 1$

CENGAGE Learning

Complexidade algorítmica

Considere a sequência (x^k) gerada pelo algoritmo de gradientes conjugados para minimizar uma quadrática com Hessiana A. Então,

$$f(x^k) - f(x^*) \leq \frac{L\|x^0 - x^*\|^2}{2(2k+1)^2},$$

onde x^* é o minimizador de f e L o maior autovalor de A.

- Considere $d^0, d^1, \ldots, d^{k-1}$ as direções geradas pelo algoritmo.
- x^k é o minimizador de f na variedade afim $x^0 + [d^0, d^1, \ldots, d^{k-1}] = V_k$.
- $f(x^k) - f(x^*) \leq \frac{1}{2}\|x^0 - x^*\|^2 \left\| A\big(p(A)\big)^2 \right\|$, para todo $p \in \mathcal{P}_k$.
- $\left\| A\big(p(A)\big)^2 \right\| \leq \max_{t \in [0,L]} \left\{ t\big(p(t)\big)^2 \right\} = \frac{L}{(2k+1)^2} \max_{t \in [0,L]} \left\{ T_{2k+1}^2\left(\frac{\sqrt{t}}{\sqrt{L}}\right) \right\}$
- $\max_{t \in [0,L]} \left\{ T_{2k+1}^2\left(\frac{\sqrt{t}}{\sqrt{L}}\right) \right\} \leq 1$

Complexidade algorítmica

Considere a sequência (x^k) gerada pelo algoritmo de gradientes conjugados para minimizar uma quadrática com Hessiana A. Então,

$$f(x^k) - f(x^*) \leq \frac{L\|x^0 - x^*\|^2}{2(2k+1)^2},$$

onde x^* é o minimizador de f e L o maior autovalor de A.

- Considere $d^0, d^1, \ldots, d^{k-1}$ as direções geradas pelo algoritmo.
- x^k é o minimizador de f na variedade afim $x^0 + [d^0, d^1, \ldots, d^{k-1}] = V_k$.
- $f(x^k) - f(x^*) \leq \dfrac{1}{2}\|x^0 - x^*\|^2 \left\| A\left(p(A)\right)^2 \right\|$, para todo $p \in \mathcal{P}_k$.
- $\left\| A\left(p(A)\right)^2 \right\| \leq \max_{t\in[0,L]} \left\{ t\left(p(t)\right)^2 \right\} = \dfrac{L}{(2k+1)^2} \max_{t\in[0,L]} \left\{ T_{2k+1}^2\left(\dfrac{\sqrt{t}}{\sqrt{L}}\right) \right\}$
- $\max_{t\in[0,L]} \left\{ T_{2k+1}^2\left(\dfrac{\sqrt{t}}{\sqrt{L}}\right) \right\} \leq 1$

Considere a sequência (x^k) gerada pelo algoritmo de gradientes conjugados para minimizar uma quadrática com Hessiana A. Então,

$$f(x^k) - f(x^*) \leq \frac{L\|x^0 - x^*\|^2}{2(2k+1)^2},$$

onde x^* é o minimizador de f e L o maior autovalor de A.

- Considere $d^0, d^1, \ldots, d^{k-1}$ as direções geradas pelo algoritmo.
- x^k é o minimizador de f na variedade afim $x^0 + [d^0, d^1, \ldots, d^{k-1}] = V_k$.
- $f(x^k) - f(x^*) \leq \dfrac{1}{2}\|x^0 - x^*\|^2 \left\| A\left(p(A)\right)^2 \right\|$, para todo $p \in \mathcal{P}_k$.
- $\left\| A\left(p(A)\right)^2 \right\| \leq \max_{t\in[0,L]} \left\{ t\left(p(t)\right)^2 \right\} = \dfrac{L}{(2k+1)^2} \max_{t\in[0,L]} \left\{ T_{2k+1}^2\left(\dfrac{\sqrt{t}}{\sqrt{L}}\right) \right\}$
- $\max_{t\in[0,L]} \left\{ T_{2k+1}^2\left(\dfrac{\sqrt{t}}{\sqrt{L}}\right) \right\} \leq 1$

Considere a sequência (x^k) gerada pelo algoritmo de gradientes conjugados para minimizar uma quadrática com Hessiana A. Então,

$$f(x^k) - f(x^*) \leq \frac{L\|x^0 - x^*\|^2}{2(2k+1)^2},$$

onde x^* é o minimizador de f e L o maior autovalor de A.

- Considere $d^0, d^1, \ldots, d^{k-1}$ as direções geradas pelo algoritmo.
- x^k é o minimizador de f na variedade afim $x^0 + [d^0, d^1, \ldots, d^{k-1}] = V_k$.
- $f(x^k) - f(x^*) \leq \dfrac{1}{2}\|x^0 - x^*\|^2 \left\| A\left(p(A)\right)^2 \right\|$, para todo $p \in \mathcal{P}_k$.
- $\left\| A\left(p(A)\right)^2 \right\| \leq \max_{t\in[0,L]} \left\{ t\left(p(t)\right)^2 \right\} = \dfrac{L}{(2k+1)^2} \max_{t\in[0,L]} \left\{ T_{2k+1}^2\left(\dfrac{\sqrt{t}}{\sqrt{L}}\right) \right\}$
- $\max_{t\in[0,L]} \left\{ T_{2k+1}^2\left(\dfrac{\sqrt{t}}{\sqrt{L}}\right) \right\} \leq 1$

Algoritmo Quase-Newton

Dados $x^0 \in \mathbb{R}^n$, $H_0 \in \mathbb{R}^{n \times n}$ definida positiva
$k = 0$
REPITA enquanto $\nabla f(x^k) \neq 0$
    Defina $d^k = -H_k \nabla f(x^k)$
    Obtenha $t_k > 0$ que minimiza $f(x^k + td^k)$ em $[0, \infty)$
    Faça $x^{k+1} = x^k + t_k d^k$
    Determine H_{k+1} definida positiva
    $k = k + 1$

Observações:

- Se $H_k = I$, a direção de busca é a do método do gradiente.
- Se $H_k = \left(\nabla^2 f(x^k)\right)^{-1}$, a direção é a do método de Newton.

CENGAGE Learning

Algoritmo Quase-Newton

Dados $x^0 \in \mathbb{R}^n$, $H_0 \in \mathbb{R}^{n \times n}$ definida positiva
$k = 0$
REPITA enquanto $\nabla f(x^k) \neq 0$
Defina $d^k = -H_k \nabla f(x^k)$
Obtenha $t_k > 0$ que minimiza $f(x^k + td^k)$ em $[0, \infty)$
Faça $x^{k+1} = x^k + t_k d^k$
Determine H_{k+1} definida positiva
$k = k + 1$

Observações:

- Se $H_k = I$, a direção de busca é a do método do gradiente.
- Se $H_k = \left(\nabla^2 f(x^k)\right)^{-1}$, a direção é a do método de Newton.

Dados $x^0 \in \mathbb{R}^n$, $H_0 \in \mathbb{R}^{n\times n}$ definida positiva
$k = 0$
REPITA enquanto $\nabla f(x^k) \neq 0$
    Defina $d^k = -H_k \nabla f(x^k)$
    Obtenha $t_k > 0$ que minimiza $f(x^k + td^k)$ em $[0, \infty)$
    Faça $x^{k+1} = x^k + t_k d^k$
    Determine H_{k+1} definida positiva
    $k = k + 1$

Observações:

- Se $H_k = I$, a direção de busca é a do método do gradiente.
- Se $H_k = \left(\nabla^2 f(x^k)\right)^{-1}$, a direção é a do método de Newton.

- Método DFP - Davidon, Fletcher e Powell.

$$H_{k+1} = H_k + \frac{p^k (p^k)^T}{(p^k)^T q^k} - \frac{H_k q^k (q^k)^T H_k}{(q^k)^T H_k q^k}$$

- Método BFGS - Broyden, Fletcher, Goldfarb e Shanno

$$H_{k+1}^{BFGS} = H_k + \left(1 + \frac{(q^k)^T H_k q^k}{(p^k)^T q^k}\right) \frac{p^k (p^k)^T}{(p^k)^T q^k} - \frac{p^k (q^k)^T H_k + H_k q^k (p^k)^T}{(p^k)^T q^k}$$

- Método DFP - Davidon, Fletcher e Powell.

$$H_{k+1} = H_k + \frac{p^k(p^k)^T}{(p^k)^T q^k} - \frac{H_k q^k (q^k)^T H_k}{(q^k)^T H_k q^k}$$

- Método BFGS - Broyden, Fletcher, Goldfarb e Shanno

$$H_{k+1}^{BFGS} = H_k + \left(1 + \frac{(q^k)^T H_k q^k}{(p^k)^T q^k}\right) \frac{p^k(p^k)^T}{(p^k)^T q^k} - \frac{p^k(q^k)^T H_k + H_k q^k (p^k)^T}{(p^k)^T q^k}$$

Convergência finita para funções quadráticas

Teorema

O métodos DFP e BFGS quando aplicados para minimizar uma função quadrática estritamente convexa encontram o minimizador em no máximo n iterações.

O método de região de confiança

O subproblema quadrático

$$
\begin{array}{ll}
\text{minimizar} & m_k(d) = f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T B_k d \\
\text{sujeito a} & \|d\| \leq \Delta_k
\end{array}
$$

O método de região de confiança

Dados: $x^0 \in \mathbb{R}^n$, $\Delta_0 > 0$ e $\eta \in [0, \frac{1}{4})$

$k = 0$

REPITA enquanto $\nabla f(x^k) \neq 0$

Obtenha d^k, solução "aproximada" do subproblema

Calcule $\rho_k = \dfrac{f(x^k) - f(x^k + d^k)}{m_k(0) - m_k(d^k)}$

SE $\rho_k > \eta$

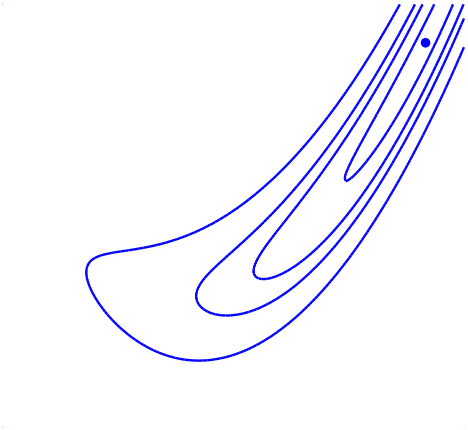
$x^{k+1} = x^k + d^k$, $\Delta_{k+1} = 2\Delta_k$

SENÃO

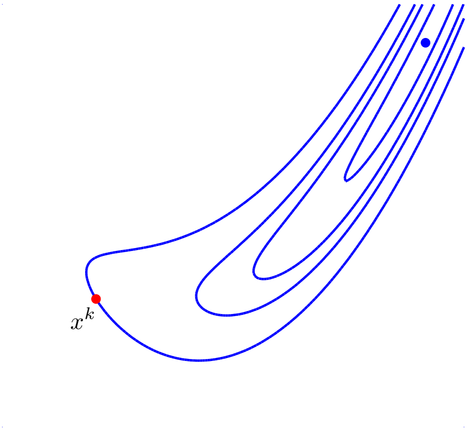
$x^{k+1} = x^k$, $\Delta_{k+1} = \dfrac{\Delta_k}{2}$

$k = k + 1$

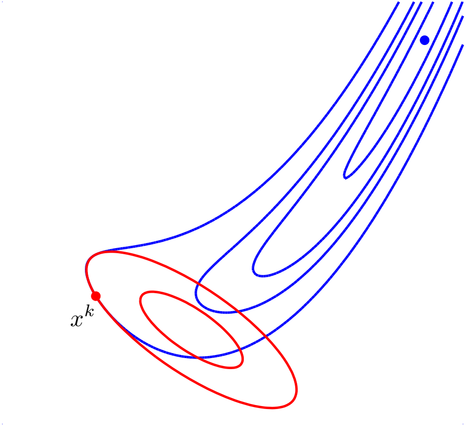
Função objetivo e seu minimizador



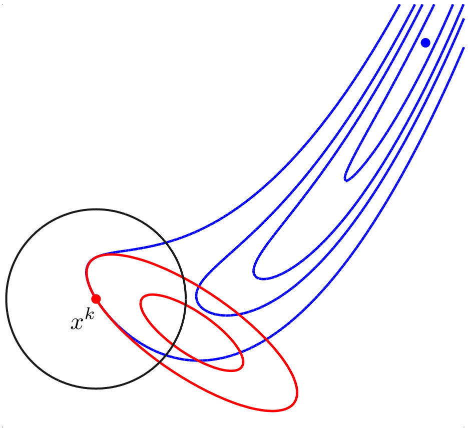
Ponto corrente



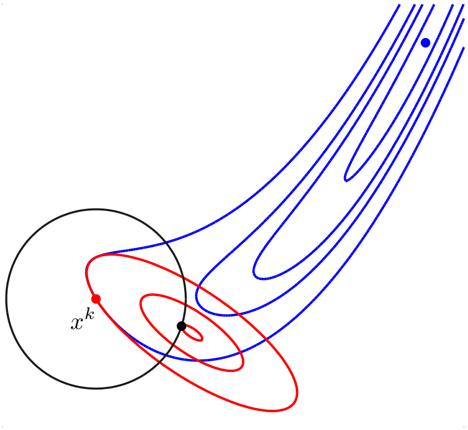
Modelo quadrático



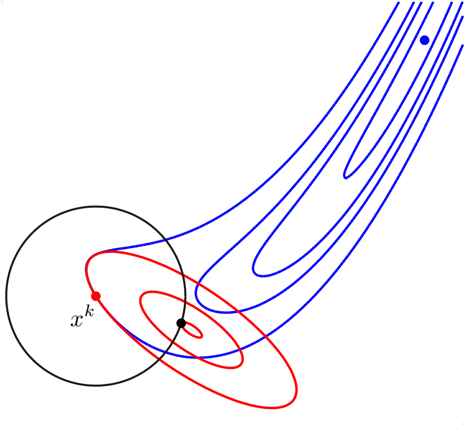
Região de confiança



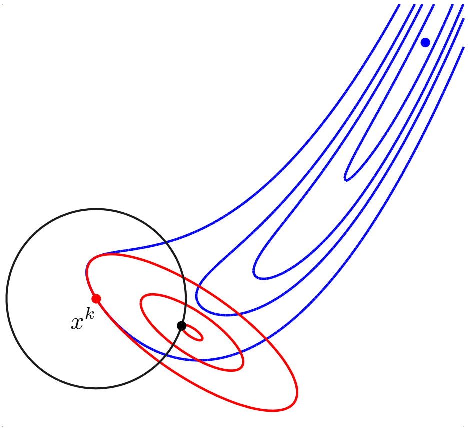
Minimizador do modelo na região de confiança



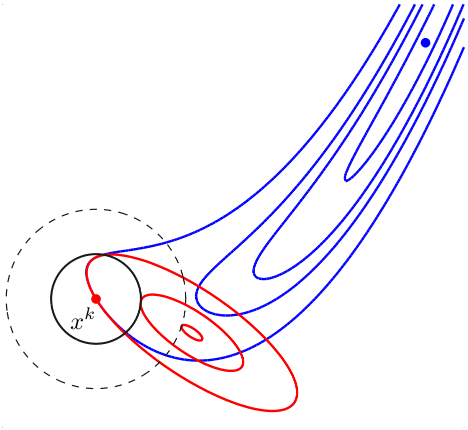
Redução insuficiente na função objetivo



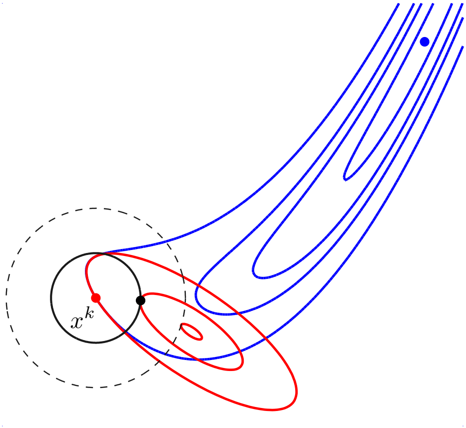
Digamos, f caiu de 10 para 6 e m_k caiu de 25 para 5



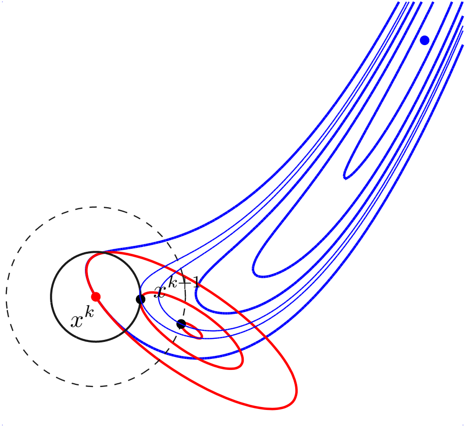
Redução do raio da região de confiança



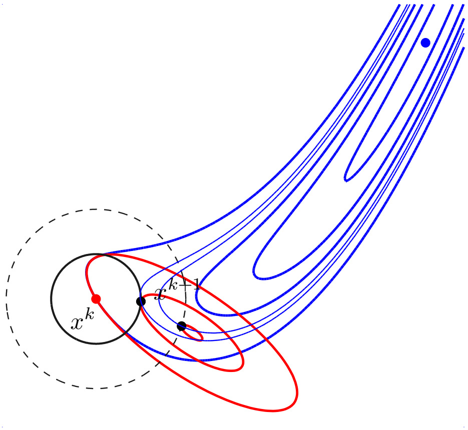
Minimizador do modelo



Redução suficiente



Digamos, f caiu de 10 para 7 e m_k caiu de 25 para 15

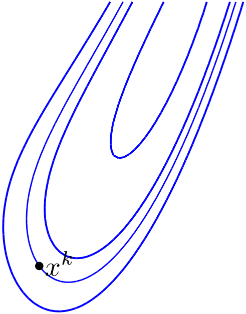


Ponto de Cauchy

$$x_c^k = x^k + d^k{}_c$$

onde $d^k{}_c = -t_k g_k$ com $t_k > 0$ solução do problema

$$\begin{array}{ll} \text{minimizar} & m_k(-tg_k) = f(x^k) - t\|g_k\|^2 + \frac{1}{2}t^2 {g_k}^T B_k g_k \\ \text{sujeito a} & \|tg_k\| \le \Delta_k. \end{array}$$



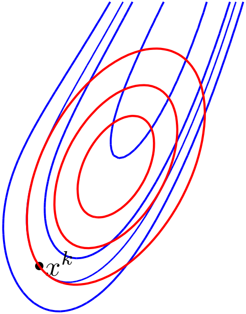
Solução aproximada do subproblema

Ponto de Cauchy

$$x_c^k = x^k + d^k{}_c$$

onde $d^k{}_c = -t_k g_k$ com $t_k > 0$ solução do problema

$$\begin{array}{ll} \text{minimizar} & m_k(-tg_k) = f(x^k) - t\|g_k\|^2 + \frac{1}{2}t^2 g_k{}^T B_k g_k \\ \text{sujeito a} & \|tg_k\| \leq \Delta_k. \end{array}$$

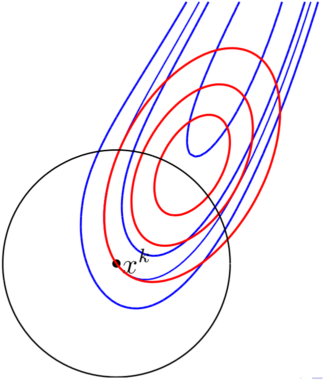


Ponto de Cauchy

$$x_c^k = x^k + d^k{}_c$$

onde $d^k{}_c = -t_k g_k$ com $t_k > 0$ solução do problema

$$\begin{array}{ll} \text{minimizar} & m_k(-tg_k) = f(x^k) - t\|g_k\|^2 + \frac{1}{2}t^2 g_k{}^T B_k g_k \\ \text{sujeito a} & \|tg_k\| \le \Delta_k. \end{array}$$

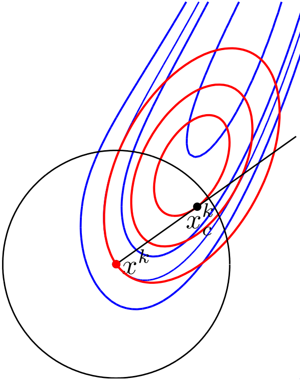


Ponto de Cauchy

$$x_c^k = x^k + d^k{}_c$$

onde $d^k{}_c = -t_k g_k$ com $t_k > 0$ solução do problema

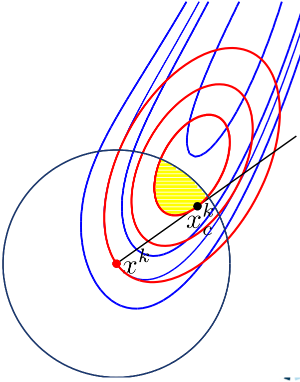
$$\begin{array}{ll} \text{minimizar} & m_k(-tg_k) = f(x^k) - t\|g_k\|^2 + \frac{1}{2}t^2 {g_k}^T B_k g_k \\ \text{sujeito a} & \|tg_k\| \leq \Delta_k. \end{array}$$



Solução aproximada do subproblema: qualquer solução que forneça uma redução de pelo menos uma fração da redução obtida pelo passo $d^k{}_c$ de Cauchy.

Ponto de Cauchy

$$x_c^k = x^k + d^k{}_c$$



Convergência global

Hipóteses

- f é de classe C^1, com ∇f Lipschitz.
- A solução d^k do subproblema satisfaz
  - $pred = m_k(0) - m_k(d^k) \geq c_1(m_k(0) - m_k(d^k{}_c))$, com $c_1 \in (0,1)$.
  - $\|d^k\| \leq \gamma\Delta_k$, com $\gamma \geq 1$.
- As Hessianas B_k são uniformemente limitadas.
- f é limitada inferiormente no conjunto de nível $\mathcal{N} = \left\{x \mid f(x) \leq f(x^0)\right\}$.

Convergência global

Hipóteses

- f é de classe $\mathcal{C}^1$, com ∇f Lipschitz.
- A solução d^k do subproblema satisfaz
  - $pred = m_k(0) - m_k(d^k) \geq c_1 (m_k(0) - m_k(d^k{}_c))$, com $c_1 \in (0,1)$.
  - $\|d^k\| \leq \gamma \Delta_k$, com $\gamma \geq 1$.
- As Hessianas B_k são uniformemente limitadas.
- f é limitada inferiormente no conjunto de nível $N = \left\{ x \mid f(x) \leq f(x^0) \right\}$.

Convergência global

Hipóteses

- f é de classe C^1, com ∇f Lipschitz.
- A solução d^k do subproblema satisfaz
  - $pred = m_k(0) - m_k(d^k) \geq c_1(m_k(0) - m_k(d^k{}_c))$, com $c_1 \in (0,1)$.
  - $\|d^k\| \leq \gamma \Delta_k$, com $\gamma \geq 1$.
- As Hessianas B_k são uniformemente limitadas.
- f é limitada inferiormente no conjunto de nível $N = \left\{ x \mid f(x) \leq f(x^0) \right\}$.

Convergência global

Hipóteses

- f é de classe C^1, com ∇f Lipschitz.
- A solução d^k do subproblema satisfaz
  - $pred = m_k(0) - m_k(d^k) \geq c_1(m_k(0) - m_k(d^k{}_c))$, com $c_1 \in (0,1)$.
  - $\|d^k\| \leq \gamma\Delta_k$, com $\gamma \geq 1$.
- As Hessianas B_k são uniformemente limitadas.
- f é limitada inferiormente no conjunto de nível $N = \left\{x \mid f(x) \leq f(x^0)\right\}$.

Convergência global

Hipóteses

- f é de classe C^1, com ∇f Lipschitz.
- A solução d^k do subproblema satisfaz
  - $pred = m_k(0) - m_k(d^k) \geq c_1(m_k(0) - m_k(d^k{}_c))$, com $c_1 \in (0,1)$.
  - $\|d^k\| \leq \gamma\Delta_k$, com $\gamma \geq 1$.
- As Hessianas B_k são uniformemente limitadas.
- f é limitada inferiormente no conjunto de nível $N = \left\{x \mid f(x) \leq f(x^0)\right\}$.

Convergência global

Teoremas

- $\liminf_{k\to\infty} \|\nabla f(x^k)\| = 0.$
- Suponha que $\eta > 0$. Então $\nabla f(x^k) \to 0.$

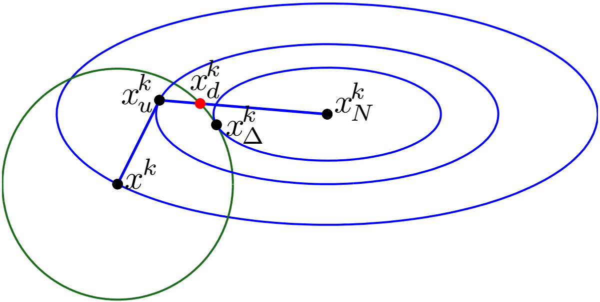
Convergência global

Teoremas

- $\liminf_{k\to\infty} \|\nabla f(x^k)\| = 0.$
- Suponha que $\eta > 0$. Então $\nabla f(x^k) \to 0$.

Método dogleg consiste em minimizar o modelo, sujeito à região de confiança, na poligonal que liga os pontos x^k, $x^k{}_u$ e $x^k{}_N$, onde

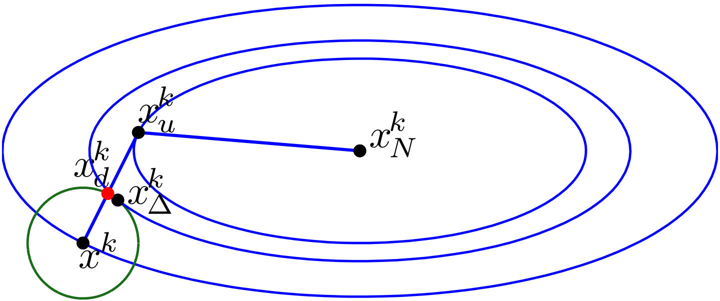
- $x^k{}_u$: minimizador do modelo na direção oposta ao gradiente.
- $x^k{}_N$: minimizador irrestrito do modelo, isto é, o ponto de Newton.



Este método se aplica quando a Hessiana do modelo é definida positiva.

Método dogleg consiste em minimizar o modelo, sujeito à região de confiança, na poligonal que liga os pontos x^k, ${x^k}_u$ e ${x^k}_N$, onde

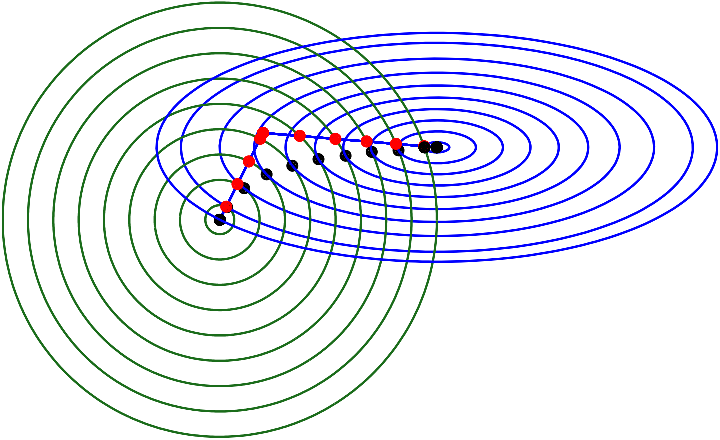
- ${x^k}_u$: minimizador do modelo na direção oposta ao gradiente.
- ${x^k}_N$: minimizador irrestrito do modelo, isto é, o ponto de Newton.



Este método se aplica quando a Hessiana do modelo é definida positiva.

Método dogleg

Trajetória em função do raio da região de confiança, do ponto dogleg, ${x^k}_d$, bem como dos pontos ${x^k}_\Delta = x^k + d^k$, onde d^k é a solução exata do subproblema.



Algoritmo dogleg

Dados: $x^k \in \mathbb{R}^n$, $\Delta_k > 0$

Calcule ${d^k}_u = -\dfrac{{g_k}^T g_k}{{g_k}^T B_k g_k} g_k$

SE $\|{d^k}_u\| > \Delta_k$

$\quad {d^k} = -\dfrac{\Delta_k}{\|g_k\|} g_k$

SENÃO

$\quad$ Determine ${d^k}_N$ tal que $B_k {d^k}_N = -g_k$

$\quad$ SE $\|{d^k}_N\| \leq \Delta_k$

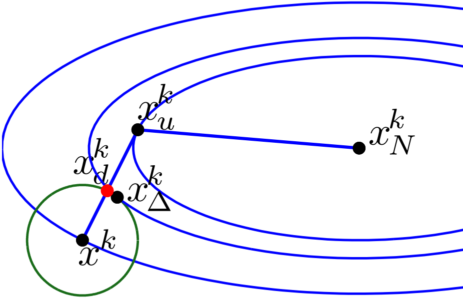
$\quad\quad d^k = {d^k}_N$

$\quad$ SENÃO

$\quad\quad$ Determine $\alpha_k \in [0,1]$ tal que

$$\|{d^k}_u + \alpha_k({d^k}_N - {d^k}_u)\| = \Delta_k$$

$$d^k = {d^k}_u + \alpha_k({d^k}_N - {d^k}_u)$$



$${x^k}_u = x^k + {d^k}_u$$

$${x^k}_N = x^k + {d^k}_N$$

$${x^k}_d = x^k + d^k$$

CENGAGE Learning

Dados: $x^k \in \mathbb{R}^n$, $\Delta_k > 0$

Calcule $d^k{}_u = -\dfrac{{g_k}^T g_k}{{g_k}^T B_k g_k} g_k$

SE $\|d^k{}_u\| > \Delta_k$

$$d^k = -\frac{\Delta_k}{\|g_k\|} g_k$$

SENÃO

Determine $d^k{}_N$ tal que $B_k d^k{}_N = -g_k$

SE $\|d^k{}_N\| \leq \Delta_k$

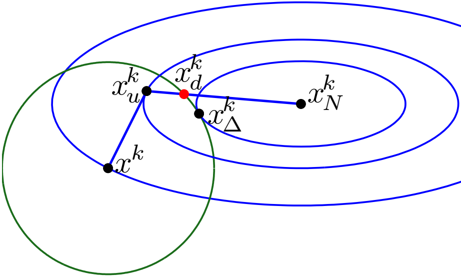
$$d^k = d^k{}_N$$

SENÃO

Determine $\alpha_k \in [0,1]$ tal que

$$\|d^k{}_u + \alpha_k (d^k{}_N - d^k{}_u)\| = \Delta_k$$

$$d^k = d^k{}_u + \alpha_k (d^k{}_N - d^k{}_u)$$



$$x^k{}_u = x^k + d^k{}_u$$

$$x^k{}_N = x^k + d^k{}_N$$

$$x^k{}_d = x^k + d^k$$

Método GC-Steihaug

Dados: $d_s^0 = 0$, $r^0 = g$, $p^0 = -r^0$. Faça $j = 0$

REPITA enquanto o passo d^k não for obtido

SE $(p^j)^T B p^j \leq 0$

Calcule $t \in \mathbb{R}$ tal que $d = d_s^j + t p^j$ minimiza m e $\|d\| = \Delta$. Faça $d^k = d$.

SENÃO

Calcule $t_j = \dfrac{(r^j)^T r^j}{(p^j)^T B p^j}$ e defina $d_s^{j+1} = d_s^j + t_j p^j$

SE $\|d_s^{j+1}\| \geq \Delta$

Calcule $t \in \mathbb{R}$ tal que $d = d_s^j + t p^j$ satisfaz $\|d\| = \Delta$. Faça $d^k = d$.

SENÃO

$r^{j+1} = r^j + t_j B p^j$

SE $r^{j+1} = 0$, faça $d^k = d_s^{j+1}$.

SENÃO

$$\beta_j = \frac{(r^{j+1})^T r^{j+1}}{(r^j)^T r^j} \quad \text{e} \quad p^{j+1} = -r^{j+1} + \beta_j p^j$$

$j = j + 1$

Principais referências

C. C. Gonzaga.
Um curso de programação não linear.
Notas de aula - UFSC, 2004.

A. Izmailov and M. Solodov.
Otimização: Métodos Computacionais, volume 2.
IMPA, Rio de Janeiro, 2007.

D. G. Luenberger.
Linear and Nonlinear Programming.
Addison - Wesley Publishing Company, New York, 1986.

J. Nocedal and S. J. Wright.
Numerical Optimization.
Springer Series in Operations Research. Springer-Verlag, 1999.