Otimização Não Linear

CM106/CMM204/CMI043

Tópico 02 - Funções Quadráticas

Abel Soares Siqueira - UFPR

2020/s1

Notação

- $x \in \mathbb{R}^n$ quer dizer x é um vetor, mas também pode ser visto como uma matriz $n \times 1$. $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$
- O produto interno entre x e y é denotado por $x^T y = \sum_{i=1}^{n} x_i y_i$.
- A norma (norma 2) de x é dada por $\|x\| = \|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$.

Notação

- Se $f : \mathbb{R}^n \to \mathbb{R}$, então o gradiente e a Hessiana de f são

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} \quad \text{e} \quad \nabla^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 x_n} \\ \frac{\partial^2 f}{\partial x_2 x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n x_1} & \frac{\partial^2 f}{\partial x_n x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

- Se $F : \mathbb{R}^n \to \mathbb{R}^m$ é uma função vetorial, então existem m funções $F_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, m$, tais que F e sua Jacobiana J são dadas por

$$F(x) = \begin{bmatrix} F_1(x) \\ F_2(x) \\ \vdots \\ F_m(x) \end{bmatrix} \quad \text{e} \quad J = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \frac{\partial F_1}{\partial x_2} & \cdots & \frac{\partial F_1}{\partial x_n} \\ \frac{\partial F_2}{\partial x_1} & \frac{\partial F_2}{\partial x_2} & \cdots & \frac{\partial F_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial x_1} & \frac{\partial F_m}{\partial x_2} & \cdots & \frac{\partial F_m}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \nabla F_1^T \\ \nabla F_2^T \\ \vdots \\ \nabla F_m^T \end{bmatrix}$$

Função Quadrática

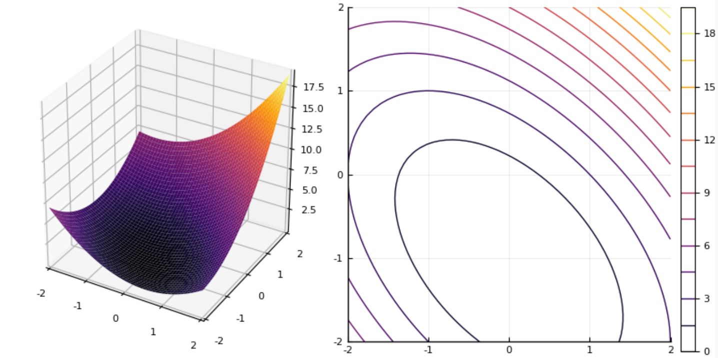
- Em 1d $f(x) = ax^2 + bx + c$

  A curvatura de f é dada pelo sinal de a.
- Em 2 variáveis, com rotações e translações pode ser escrita como $f(x) = \alpha_1 x_1^2 + \alpha_2 x_2^2$

  α_1 e α_2 definem a quadrática
- Em n variáveis, $f(x) = \frac{1}{2}x^T Ax - b^T x + c$

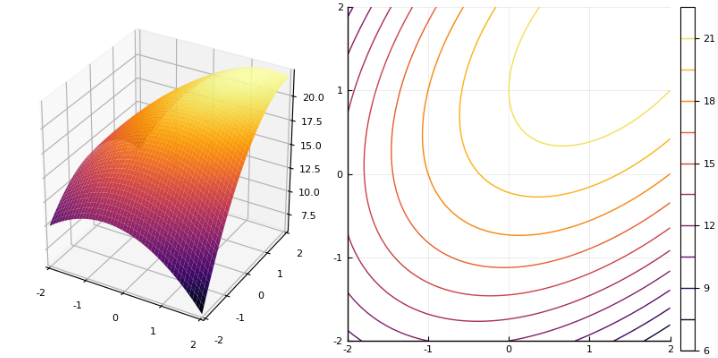
  $c \in \mathbb{R}$, $b \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times n}$, A simétrica.
- A forma da f depende de A.
- A ideia de vértice têm que ser estendida.

Função Quadrática - $f(x) = \alpha_1 x_1^2 + \alpha_2 x_2^2$



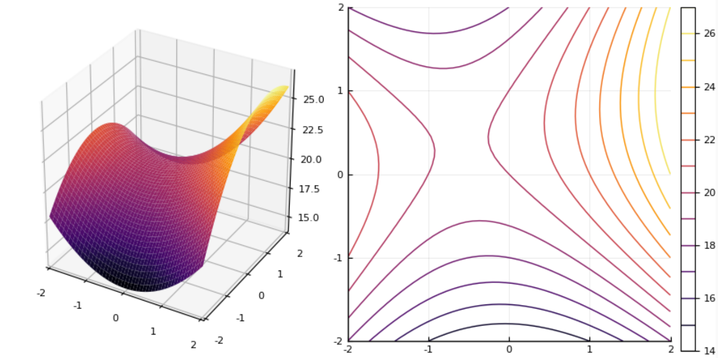
$$\alpha_1 > 0, \alpha_2 > 0.$$

Função Quadrática - $f(x) = \alpha_1 x_1^2 + \alpha_2 x_2^2$



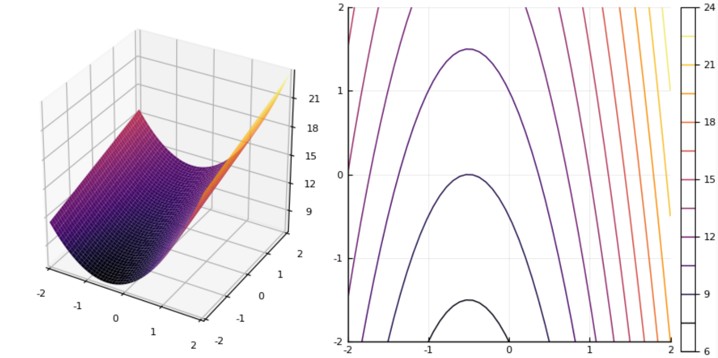
$\alpha_1 < 0, \alpha_2 < 0.$

Função Quadrática - $f(x) = \alpha_1 x_1^2 + \alpha_2 x_2^2$



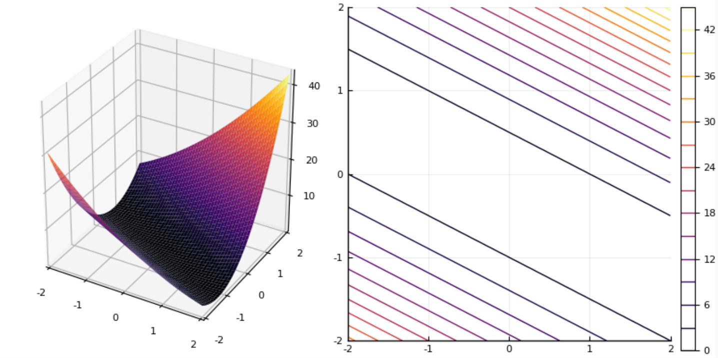
$$\alpha_1 > 0, \alpha_2 < 0.$$

Função Quadrática - $f(x) = \alpha_1 x_1^2 + \alpha_2 x_2^2$



$$\alpha_1 > 0, \alpha_2 = 0.$$

Função Quadrática - $f(x) = \alpha_1 x_1^2 + \alpha_2 x_2^2$



$$\alpha_1 > 0, \alpha_2 = 0.$$

Função Quadrática

- De uma maneira geral, podemos rotacionar e transladar f para obter $f(x) = \alpha_1 x_1^2 + \cdots + \alpha_n x_n^2$.
- Esses valores são os autovalores de A. As rotações vêm dos autovetores.
- Se todos os autovalores forem positivos, existe um único minimizador.
- Se todos forem positivos ou nulos, pode ser que existam infinitos ou nenhum.
- Se algum for negativo, existe um direção que decresce infinitamente.

Def.: Uma matriz simétrica A é dita definida positiva se $x^T Ax > 0$ para todo $x \neq 0$.

Def.: Uma matriz simétrica A é dita semi-definida positiva se $x^T Ax \geq 0$ para todo x.

Def.: Uma matriz simétrica A é dita indefinida se existem x e y tais que $x^T Ax > 0$ e $y^T Ay < 0$.

Teo.: Seja A simétrica. Os autovalores de A dizem a definição de A.

Dem.: Como A é simétrica vale o Teo. espectral: existe base $\{v_1, \ldots, v_n\}$ ortonormal de autovetores. Sejam λ_i os autovalores associados, i.e., $Av_i = \lambda_i v_i$. Tome $x \in \mathbb{R}^n$, existem $\alpha_1, \ldots, \alpha_n$ tais que $x = \alpha_1 v_1 + \cdots + \alpha_n v_n$. Logo,

$$x^T A x = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j v_i^T A v_j = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \lambda_j v_i^T v_j = \sum_{i=1}^{n} \alpha_i^2 \lambda_i v_i^T v_i = \sum_{i=1}^{n} \alpha_i^2 \lambda_i.$$

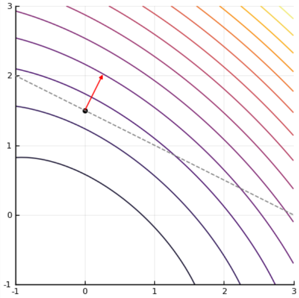
As relações entre os sinais de λ_i e o sinal de $x^T A x$ fica evidente.

Minimizador, vértice e pontos críticos

- No caso de A definida positiva, o minimizador de $f(x) = \frac{1}{2}x^TAx - b^Tx + c$ é a única solução de $Ax = b$.
- Se A é definida negativa, a solução de $Ax = b$ é um maximizador.
- Se A é indefinida, a solução é chamada de ponto de sela.
- Se A é singular, pode ser que $Ax = b$ tenha zero ou infinitas soluções.

Método de Cauchy

- Hipótese A definida positiva.
- Se x_k é o minimizador, temos $Ax_k = b$, isto é, $\nabla f(x_k) = 0$.
- **Direções de descida**: d tal que $d^T \nabla f(x_k) < 0$

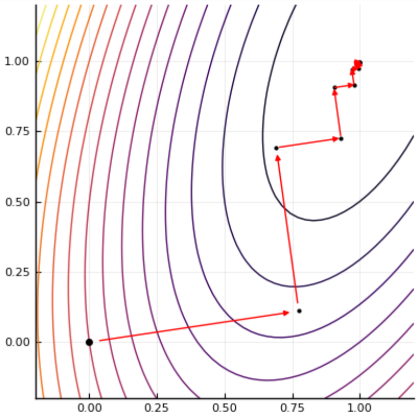


Método de Cauchy

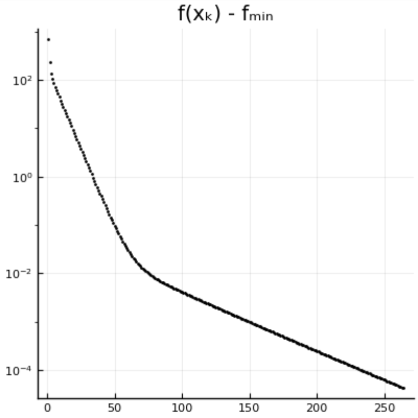
- $\min_d d^T \nabla f(x_k) : \ \|d\| = 1?\ d = -\alpha \nabla f(x_k)$.
- Método do Gradiente Descendente:

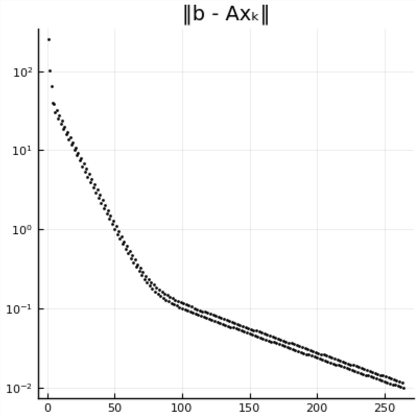
  $x_{k+1} = x_k - \alpha_k \nabla f(x_k) = x_k - \alpha_k (Ax_k - b)$.
- α_k constante funciona se o valor for pequeno o suficiente. Tradicional para Machine Learning.
- Cauchy: e se α_k for escolhido para que $f(x_{k+1})$ seja o menor possível?
- $\alpha_k = \dfrac{(Ax_k - b)^T (Ax_k - b)}{(Ax_k - b)^T A (Ax_k - b)}$

Método de Cauchy



Método de Cauchy





Método dos Gradientes Conjugados

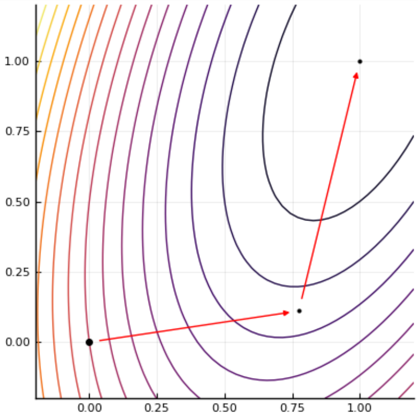
- Considere duas direções linearmente independentes v_1 e v_2, e um ponto inicial x_0.
- Vamos definir x_1 como o minimizador de $f(x_0 + \alpha_1 v_1)$ em função de α_1.
- Vamos definir x_2 como o minimizador de $f(x_0 + \beta_1 v_1 + \beta_2 v_2)$ em função de β_1 e β_2.
- Note que em geral, minimizar $x_1 + \alpha_2 v_2$ não dará x_2.
- x_1 minimiza numa reta, e x_2 num plano.
- Note que minimizar num plano é melhor que minimizar numa reta e depois noutra reta (em geral).
- Note que minimizar no plano é mais caro, mas e se v_2 puder ser escolhida para facilitar?

Método dos Gradientes Conjugados

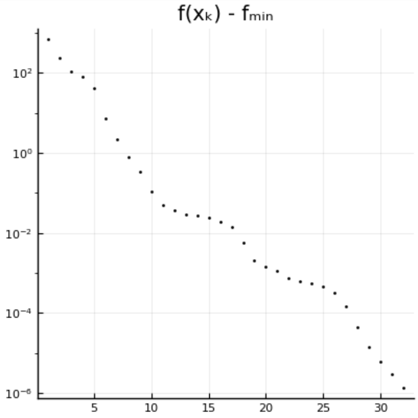
- Definimos dois vetores v e w como A-conjugados se $v^T A w = 0$ para A definida positiva.
- A partir de x_0, definimos $v_1 = b - Ax_0 = r_0$ e as sequências

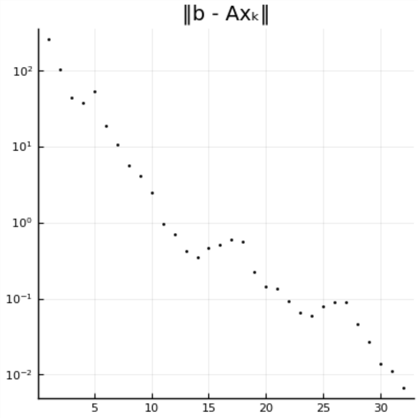
$$
\begin{aligned}
\alpha_k &= \arg\min_{\alpha} f(x_{k-1} + \alpha v_k) = \frac{v_k^T r_{k-1}}{v_k^T A v_k} \\
x_k &= x_{k-1} + \alpha_k v_k \\
r_k &= b - Ax_k \\
v_{k+1} &= r_k - \sum_{i=1}^{k} \frac{v_i^T A r_k}{v_i^T A v_i} v_i
\end{aligned}
$$

Método dos Gradientes Conjugados



Método dos Gradientes Conjugados





Método dos Gradientes Conjugados

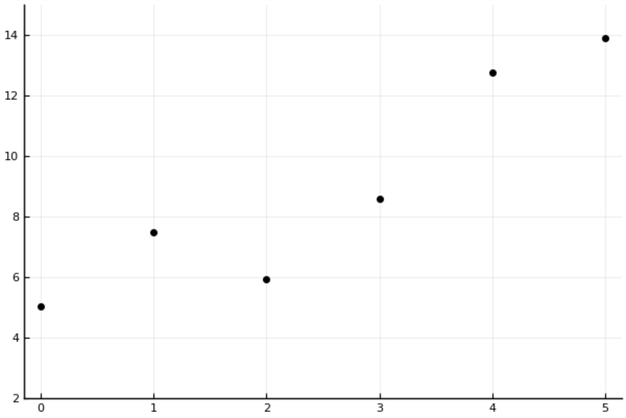
- Várias melhorias e resultados para este método.
- **Teo.:** $r_{k+1}^T v_i = 0$ para $i = 1, \dots, k$ e x_k é minimizador de $f(x)$ no conjunto $x_0 + \text{span}\{v_1, \dots, v_k\}$.
- $v_{k+1} = r_k - \sum_{i=1}^{k-1} \frac{v_i^T A r_k}{v_i^T A v_i} v_i = r_k + \beta_{k+1} v_k$ com $\beta_{k+1} = -\frac{v_k^T A r_{k+1}}{v_k^T A v_k}$.
- $r_k = b - A x_k = b - A(x_{k-1} + \alpha_k v_k) = r_{k-1} - \alpha_k A v_k$
- **Teo.:** $r_k^T r_i = 0$ e $v_{k+1}^T A v_{i+1} = 0$ para $i = 0, \dots, k-1$.
- **Teo.:** $\text{span}\{r_0, r_1, \dots, r_k\} = \text{span}\{r_0, A r_0, \dots, A^k r_0\} = \text{span}\{v_1, \dots, v_{k+1}\}$.
- $\alpha_k = \frac{r_{k-1}^T r_{k-1}}{v_k^T A v_k}$ e $\beta_{k+1} = \frac{r_k^T r_k}{r_{k-1}^T r_{k-1}}$.

Método dos Gradientes Conjugados

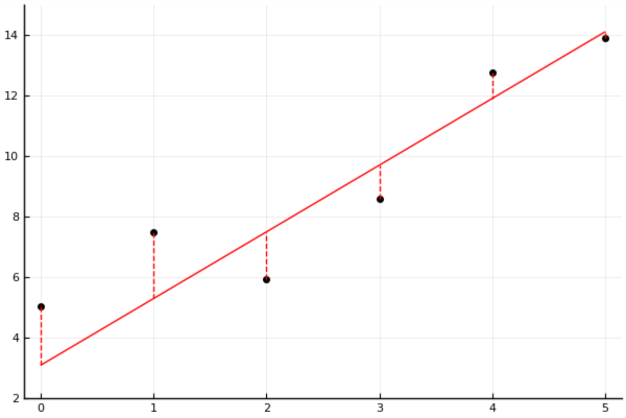
1. Dados A e b.
2. Defina $r_0 = b - Ax_0$, $v_1 = r_0$, $k = 1$.
3. $\alpha_k = \dfrac{r_{k-1}^T r_{k-1}}{v_k^T A v_k}$
4. $x_k = x_{k-1} + \alpha_k v_k$
5. $r_k = r_{k-1} - \alpha_k A v_k$
6. Se $r_k = 0$, FIM.
7. $\beta_{k+1} = \dfrac{r_k^T r_k}{r_{k-1}^T r_{k-1}}$
8. $v_{k+1} = r_k + \beta_{k+1} v_k$
9. Incremente k e volte ao passo 3.

Quadrados Mínimos

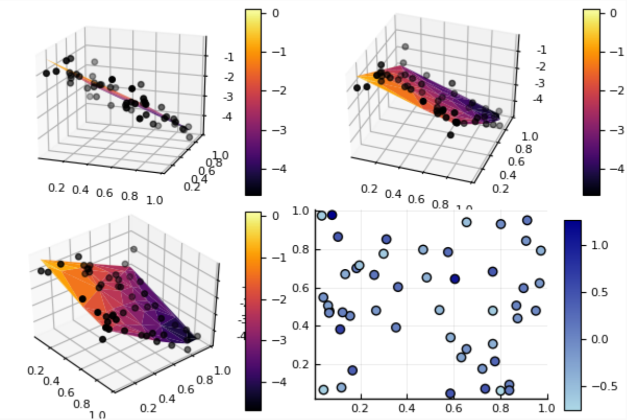
Ajuste de retas por quadrados mínimos



Ajuste de retas por quadrados mínimos



Ajuste de retas por quadrados mínimos



Ajuste de retas por quadrados mínimos

- Dados: $\Omega = \{(x_i, y_i),\ i = 1, \dots, n\} \subset \mathbb{R}^n \times \mathbb{R}$.
- Ajuste $y \approx X\beta + \epsilon$, onde $\epsilon \in \mathbb{R}^m$ e $\beta \in \mathbb{R}^n$.
- $\min\limits_{\beta} \ \frac{1}{2}\|X\beta - y\|^2$.

Quadrados Mínimos

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

- Quadrados mínimos é uma função quadrática.
- $\nabla f(x) = A^T(Ax - b)$, e sempre existe x tal que $\nabla f(x) = 0$.
- $\nabla^2 f(x) = A^T A$, é semi-definida ou definida positiva.

Sumário

- Uma função quadrática pode ser definida pelo seu "vértice", onde $\nabla f(x) = 0$, e sua concavidade dada por $\nabla^2 f(x)$.
- A concavidade é definida pelo sinal dos autovalores.
- Se a concavidade é estritamente positiva, i.e. a quadrática é estritamente convexa, sempre temos solução única.
- O método de Cauchy é uma maneira computacional de encontrar esse método.
- O método dos Gradientes Conjugados é um método melhor.
- Quadrados mínimos é um caso de programação quadrática.

FIM