Problema de Mínimos Quadrados

Sejam A uma matriz de ordem $m \times n$ e o vetor $b \in \mathbb{R}^m$. Determinemos $x^* \in \mathbb{R}^n$ tal que

$$Ax^* = b$$

podendo o N° de equações exceder o N° de incógnitas, isto é,

$$m \geqslant n.$$

Exemplo: Para $m = 3$ e $n = 1$, seja o problema

$$\underbrace{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}}_{A} x = \underbrace{\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}}_{b} \quad \Rightarrow \quad \begin{aligned} x &= 3 \\ x &= 2 \\ 0 &= 1 \end{aligned}$$

Ou seja,

não existe $x^* \in \mathbb{R}$ tal que resolva este sistema.

- Se $m > n$, então o sistema é chamado de sobre-determinado.

- Em geral, este tipo de sistemas não tem solução no sentido clássico, isto é, $b \notin \mathrm{Im}(A)$.

- A origem do problema de mínimos quadrados é a necessidade de uma noção de "soluções generalizadas" para um sistema linear $Ax = b$.

- A ideia é determinar um vector x tal que Ax seja "o mais próximo possível" de b.

- Ou equivalentemente, considerando o vector resíduo

$$r = b - Ax \in \mathbb{R}^m$$

  precisamos tornar r "suficientemente pequeno" para uma escolha conveniente de x, mas em geral não será zero.

Então o que significa resolver

um sistema sem solução?

- No caso de um sistema de equações sobre-determinado existe uma resposta natural.

- Como o resíduo r não será zero, vamos fazê-lo o mais pequeno possível.

- Desde que medir vetores envolve escolher uma norma.

- Então, nosso problema pode ser reescrito como:

Determinemos $x^* \in \mathbb{R}^n$ tal que:

a norma $\|b - Ax^*\|_p$ seja a menor possível.

Em termos matemáticos:

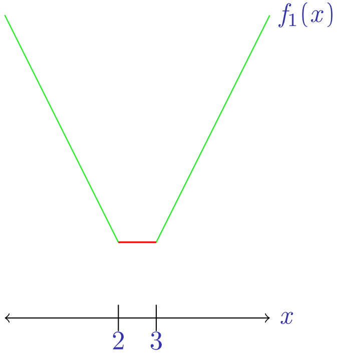
$$\min_{x \in \mathbb{R}^n} \|b - Ax\|_p$$

para alguma norma apropriada.

Exemplos

- Se $p = 1$, então

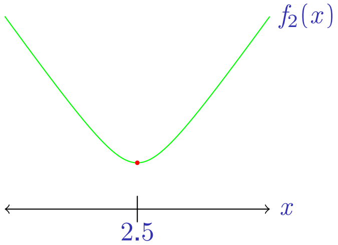
$$f_1(x) = \|b - Ax\|_1 = |x - 3| + |x - 2| + |0 - 1|$$



Assim, qualquer $x^* \in [2, 3]$ minimiza $\|b - Ax\|_1$, com 2 sendo o valor mínimo.

- Se $p = 2$, então

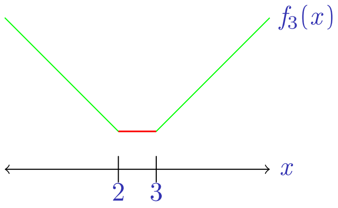
$$f_2(x) = \|b - Ax\|_2 = \sqrt{(x-3)^2 + (x-2)^2 + (0-1)^2}$$



Assim $x^* = 2.5$ minimiza $\|b - Ax\|_2$, com $\dfrac{\sqrt{6}}{2}$ sendo o valor mínimo.

- Se $p = \infty$, então

$$f_3(x) = \|b - Ax\|_\infty = \max\{|x-3|, |x-2|, 1\}$$



Assim, qualquer $x^* \in [2, 3]$ minimiza $\|b - Ax\|_\infty$, com 1 sendo o valor mínimo.

- Embora que, atualmente, existam boas técnicas para trabalhar com a minimização de $\| \cdot \|_1$ ou $\| \cdot \|_\infty$. Evitamos trabalhar com estas.

- Desde que

$$\|b - Ax\|_1 \quad \text{e} \quad \|b - Ax\|_\infty$$

  são funções não, necessariamente, diferenciáveis.

- A teoria apresentada a seguir não considera estas normas.

Em contraste, trabalhar com $\| \cdot \|_2$, é mais vantajoso por 2 razões:

- $\phi(x) = \frac{1}{2}\|b - Ax\|_2^2$ é uma função diferenciável em x. Assim,

  $$x^* \text{ minimiza } \phi(x) \quad \Leftrightarrow \quad \nabla\phi(x^*) = 0 \quad \equiv \quad A^T A x^* = A^T b$$

  Se A tem posto completo, então $A^T A$ é uma matriz positiva definida, e resolver este último problema é "fácil";

- $\| \cdot \|_2$ é preservada sob transformações ortogonais. Ou seja, podemos procurar uma matriz ortogonal Q tal que o problema equivalente

  $$\min_{x \in \mathbb{R}^n} \|Q^T A x - Q^T b\|_2$$

  é "fácil" de resolver.

Assim, o problema de mínimos quadrados é da forma

$$\boxed{(\mathrm{PMQ}) \qquad \min_{x\in\mathbb{R}^n} \|Ax-b\|_2^2.}$$

Lema sobre o (PMQ): Se $x^* \in \mathbb{R}^n$ é tal que

$$A^T(Ax^*-b) = \mathbf{0},$$

Então,

1. x^* é uma solução do (PMQ).
2. Mais ainda, se A tem posto completo, então esta solução é única.

Prova de 1: Desde que

$$A^T(Ax^* - b) = \mathbf{0} \quad \Rightarrow \quad (A(:,j))^T(Ax^* - b) = 0, \quad \forall j = 1, 2, \ldots n.$$

Então $Ax^* - b$ é ortogonal ao espaço formado pelas colunas de A. Agora, considerando qualquer $y \in \mathbb{R}^n$ temos que

$$A(y - x^*) = (y_1 - x_1^*)A(:,1) + (y_2 - x_2^*)A(:,2) + \cdots + (y_n - x_n^*)A(:,n)$$

Então, $A(y - x^*)$ pertence ao espaço formado pelas colunas de A. Logo,

$$(A(y - x^*))^T(Ax^* - b) = 0. \tag{1}$$

Assim,

$$\begin{aligned} \|Ay - b\|_2^2 &= \|Ay - Ax^* + Ax^* - b\|_2^2 = \|A(y - x^*) + Ax^* - b\|_2^2 \\ &= \|A(y - x^*)\|_2^2 + 2(A(y - x^*))^T(Ax^* - b) + \|Ax^* - b\|_2^2 \end{aligned} \tag{2}$$

Então, de (6) e (2) se segue que

$$\|Ay - b\|_2^2 = \|A(y - x^*)\|_2^2 + \|Ax^* - b\|_2^2 \geqslant \|Ax^* - b\|_2^2 \tag{3}$$

Portanto, x^* é uma solução de (PMQ).

Lema sobre o(PMQ): Se $x^* \in \mathbb{R}^n$ é tal que

$$A^T(Ax^* - b) = \mathbf{0},$$

Então,

1. x^* é uma solução do (PMQ).
2. Mais ainda, se A tem posto completo, então esta solução é única.

Prova de 2: Agora supondo que A tem posto completo, então as colunas de A são $\mathrm{L.I.}$, assim considerando $y \in \mathbb{R}^n$, com $y \neq x^*$, temos que

$$A(y - x^*) \neq \mathbf{0} \quad \Rightarrow \quad \|A(y - x^*)\|_2^2 > 0$$

Logo, em (3), temos que

$$\|Ay - b\|_2^2 = \|A(y - x^*)\|_2^2 + \|Ax^* - b\|_2^2 > \|Ax^* - b\|_2^2 \tag{4}$$

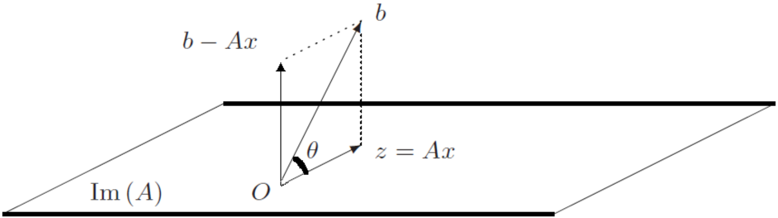
Portanto, x^* é a única solução de (PMQ).

No caso de $m = n$, se a matriz A é não singular, então existe um único minimizador $x^* = A^{-1}b$, com valor mínimo é igual a zero.

Assim, neste caso, o (PMQ) é equivalente a resolver um sistema linear.

Mais ainda, se uma solução do sistema linear $Ax = b$ existe, então também é uma solução do (PMQ). O inverso não necessariamente é verdade.

Interpretação geométrica



Recapitulando: O Problema

Sejam A uma matriz de ordem $m \times n$ e o vetor $b \in \mathbb{R}^m$. Determinemos $x^* \in \mathbb{R}^n$ tal que

$$Ax^* = b$$

podendo o N° de equações exceder o N° de incógnitas, isto é,

$$m \geqslant n.$$

Então, supondo que o posto de A é completo, isto é, $\mathrm{posto}(A) = n$, para resolver o problema estudaremos os métodos baseados em:

- Equações Normais;
- Decomposição QR;
- Decomposição por Valores Singulares.

Método de Equações Normais

Desde que, resolver

$$A^T A x = A^T b.$$

é equivalente a resolver o (PMQ), então analisemos as propriedades deste novo problema.

- $A^T A$ é uma matriz de ordem n;
- $A^T A$ é uma matriz simétrica;
- $\det(A^T A) \neq 0$;
- Existe $(A^T A)^{-1}$;
- $A^T A$ é uma matriz positiva definida;
- Todos os menores principais de $A^T A$ têm determinante positivo;
- $\det(A^T A) > 0$;
- $A^T A$ aceita uma decomposição de Cholesky, isto é, existe uma matriz triangular inferior G, de ordem n, tal que

$$A^T A = G G^T.$$

Algoritmo PMQ via EN: Dada a matriz A de ordem $m \times n$, com posto n, e $b \in \mathbb{R}^m$.

Passo 1: Construir a matriz $\tilde{A} = A^T A$ e o vetor $\tilde{b} = A^T b$;

Passo 2: Determinar a decomposição de Cholesky de $\tilde{A}$, isto é

$$\tilde{A} = GG^T;$$

Passo 3: Resolver em y o sistema triangular inferior:

$$Gy = \tilde{b};$$

Passo 4: Resolver em x o sistema triangular superior:

$$G^T x = y.$$

Lembrando a Decomposição de Cholesky

Algoritmo Cholesky: Faça:

$$g_{11} \leftarrow \sqrt{\tilde{a}_{11}}$$

Para $j = 2, \ldots, n$, faça:

$$g_{j1} \leftarrow \frac{\tilde{a}_{j1}}{g_{11}}$$

Para $i = 2, \ldots, n-1$, faça:

$$g_{ii} = \sqrt{\tilde{a}_{ii} - \sum_{k=1}^{i-1} g_{ik}^2}$$

Para $j = i+1, \ldots, n$, faça:

$$g_{ji} = \frac{\tilde{a}_{ji} - \sum_{k=1}^{i-1} g_{ik}\, g_{jk}}{g_{ii}}$$

Faça

$$g_{nn} = \sqrt{\tilde{a}_{nn} - \sum_{k=1}^{n-1} g_{nk}^2}$$

Exemplo: Usando o Algoritmo PMQ via EN determinemos a solução do seguinte (PMQ):

$$\begin{bmatrix} 4 & 5 \\ 3 & 5 \\ 0 & -\dfrac{1}{2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -3 \\ -2 \end{bmatrix}$$

Solução: Como

$$A = \begin{bmatrix} 4 & 5 \\ 3 & 5 \\ 0 & -\frac{1}{2} \end{bmatrix} \quad \Rightarrow \quad A^T = \begin{bmatrix} 4 & 3 & 0 \\ 5 & 5 & -\frac{1}{2} \end{bmatrix}$$

Passo 1: Construção de $\tilde{A}$ e $\tilde{b}$:

$$\tilde{A} = A^T A = \begin{bmatrix} 25 & 35 \\ 35 & \frac{201}{4} \end{bmatrix} \quad \text{e} \quad \tilde{b} = A^T b = \begin{bmatrix} -5 \\ -9 \end{bmatrix}$$

Passo 2: Obtendo a Decomposição de Cholesky de $\tilde{A}$, temos que

$$G = \begin{bmatrix} 5 & 0 \\ 7 & \frac{\sqrt{5}}{2} \end{bmatrix}, \quad G^T = \begin{bmatrix} 5 & 7 \\ 0 & \frac{\sqrt{5}}{2} \end{bmatrix} \quad \text{e} \quad \tilde{A} = GG^T.$$

Passo 3: resolvendo em y o sistema triangular inferior

$$\begin{bmatrix} 5 & 0 \\ 7 & \dfrac{\sqrt{5}}{2} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} -5 \\ -9 \end{bmatrix} \quad \Rightarrow \quad y^* = \begin{bmatrix} -1 \\ -\dfrac{4}{\sqrt{5}} \end{bmatrix}.$$

Passo 4: resolvendo em x o sistema triangular superior

$$\begin{bmatrix} 5 & 7 \\ 0 & \dfrac{\sqrt{5}}{2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 \\ -\dfrac{4}{\sqrt{5}} \end{bmatrix} \quad \Rightarrow \quad x^* = \begin{bmatrix} \dfrac{51}{25} \\ -\dfrac{8}{5} \end{bmatrix}.$$

Portanto, $x^* = \begin{bmatrix} \dfrac{51}{25} \\ -\dfrac{8}{5} \end{bmatrix}$ é o vetor de $\mathbb{R}^2$ que minimiza o (PMQ), com

valor mínimo $\|Ax^* - b\|_2^2 = \dfrac{49}{5}$.

Exemplo: Usando o Algoritmo PMQ via EN determinemos a solução do seguinte (PMQ):

$$\begin{bmatrix} 1 & -2 & 1 \\ 2 & -1 & 0 \\ 1 & -1 & 1 \\ -1 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -7 \\ -2 \\ 7 \\ -2 \end{bmatrix}$$

Solução: Como

$$A = \begin{bmatrix} 1 & -2 & 1 \\ 2 & -1 & 0 \\ 1 & -1 & 1 \\ -1 & 0 & 2 \end{bmatrix} \quad \Rightarrow \quad A^T = \begin{bmatrix} 1 & 2 & 1 & -1 \\ -2 & -1 & -1 & 0 \\ 1 & 0 & 1 & 2 \end{bmatrix}$$

Passo 1: Construção de $\tilde{A}$ e $\tilde{b}$:

$$\tilde{A} = A^T A = \begin{bmatrix} 7 & -5 & 0 \\ -5 & 6 & -3 \\ 0 & -3 & 6 \end{bmatrix} \quad \text{e} \quad \tilde{b} = A^T b = \begin{bmatrix} -2 \\ 9 \\ -4 \end{bmatrix}$$

Passo 2: Obtendo a Decomposição de Cholesky de $\tilde{A}$, temos que

$$G = \begin{bmatrix} \sqrt{7} & 0 & 0 \\ -\frac{5}{\sqrt{7}} & \sqrt{\frac{17}{7}} & 0 \\ 0 & -3\sqrt{\frac{7}{17}} & \sqrt{\frac{39}{17}} \end{bmatrix}, \quad G^T = \begin{bmatrix} \sqrt{7} & -\frac{5}{\sqrt{7}} & 0 \\ 0 & \sqrt{\frac{17}{7}} & -3\sqrt{\frac{7}{17}} \\ 0 & 0 & \sqrt{\frac{39}{17}} \end{bmatrix} \quad \text{e} \quad \tilde{A} = GG^T.$$

Passo 3: resolvendo em y o sistema triangular inferior

$$\begin{bmatrix} \sqrt{7} & 0 & 0 \\ -\frac{5}{\sqrt{7}} & \sqrt{\frac{17}{7}} & 0 \\ 0 & -3\sqrt{\frac{7}{17}} & \sqrt{\frac{39}{17}} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} -2 \\ 9 \\ -4 \end{bmatrix} \quad \Rightarrow \quad y^* = \begin{bmatrix} -\frac{2}{\sqrt{7}} \\ \frac{53}{\sqrt{119}} \\ 7\sqrt{\frac{13}{51}} \end{bmatrix}.$$

Passo 4: resolvendo em x o sistema triangular superior

$$\begin{bmatrix} \sqrt{7} & -\frac{5}{\sqrt{7}} & 0 \\ 0 & \sqrt{\frac{17}{7}} & -3\sqrt{\frac{7}{17}} \\ 0 & 0 & \sqrt{\frac{39}{17}} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -\frac{2}{\sqrt{7}} \\ \frac{53}{\sqrt{119}} \\ 7\sqrt{\frac{13}{51}} \end{bmatrix} \quad \Rightarrow \quad x^* = \begin{bmatrix} 4 \\ 6 \\ \frac{7}{3} \end{bmatrix}.$$

Portanto, $x^* = \begin{bmatrix} 4 \\ 6 \\ \frac{7}{3} \end{bmatrix}$ é o vetor de $\mathbb{R}^3$ que minimiza o (PMQ), com valor mínimo $\|Ax^* - b\|_2^2 = \dfrac{208}{3}$.

Eficiência deste método

- O custo computacional assintótico do Algoritmo PMQ via EN é da ordem
$$mn^2 + \frac{n^3}{3}$$
operações de ponto flutuante por segundo;
- Ao usar a `decomposição de Cholesky` para a obtenção do sistema normal
$$A^T Ax = A^T b$$
torna-se mais rápido comparado com os algoritmos análogos, tipo LU e QR;
- Em muitas aplicações $m >> n$. Neste caso, a matriz $A^T A$ é relativamente pequena e o custo computacional maior será a construção de $A^T A$ $(O(mn^2))$ e não na solução do sistema linear $A^T Ax = A^T b$ $(O(n^3))$.

Possível Instabilidade das Equações Normais

Exemplo: Seja a matriz

$$A = \begin{bmatrix} 1 & 1 & 1 \\ \varepsilon & 0 & 0 \\ 0 & \varepsilon & 0 \\ 0 & 0 & \varepsilon \end{bmatrix} \quad \Rightarrow \quad A^T A = \begin{bmatrix} 1+\varepsilon^2 & 1 & 1 \\ 1 & 1+\varepsilon^2 & 1 \\ 1 & 1 & 1+\varepsilon^2 \end{bmatrix}$$

- Assim, os postos tanto de A como de $A^T A$ dependem de $\varepsilon \neq 0$;
- No entanto, a não-singularidade de $A^T A$ é um assunto mais delicado para valores pequenos de ε;
- Num computador, $\mathrm{posto}(A) = 3$ e $\mathrm{posto}(A^T A) = 1$.

- Desde que A tem posto completo, então o sistema normal

$$A^T Ax = A^T b$$

é não singular, isto é, $\det(A^T A) \neq 0$;

- Embora que o uso das equações normais para resolver (PMQ) seja mais atraente, pois conceitualmente é simples;

- Erros de arredondamento no cálculo de $A^T A$ podem ocasionar perda de dígitos significativos, em consequência a perda da não-singularidade da matriz $A^T A$.

Bibliografia

1. Álgebra Linear e suas aplicações; Strang, G., Cengage Learning, 4° Edição, 2010.
2. Álgebra Linear; Poole, D., Thomson Learning, 2004.
3. Matrix Computations; Golub, G., Van Loan, C.F., Johns Hopkins University Press, 4° Edição, 2012.
4. Applied Numerical Linear Algebra; Demmel, J., SIAM, 1997.
5. Introduction to Applied Linear Algebra; Boyd, S., Vandenberghe, L., Cambridge University Press, 1° Edição, 2018.
6. Numerical Linear Algebra with Julia; Darve, E., Wootters, M., SIAM, 1° Edição, 2021.
7. Linear Algebra and Optimization for Machine Learning; Aggarwal, C. C., Springer, 1° Edição, 2020.
8. Algoritmos Numéricos; Campos, F.F. filho, LTC, 3 Edição, 2018.