MAURA FERREIRA MARINELLI

Método de Quadrados Mínimos

Trabalho de Conclusão de Curso apresentado ao
Curso de Matemática - Habilitação Licenciatura
Departamento de Matemática
Centro de Ciências Físicas e Matemáticas
Universidade Federal de Santa Catarina

Orientadora: Joana B. O. Quandt

Florianópolis -SC
Abril - 2002

Método de Quadrados Mínimos

Esta monografia foi julgada adequada como TRABALHO DE CONCLUSÃO DE CURSO no curso de Matemática - Habilitação Licenciatura e aprovada em sua forma final pela Banca Examinadora designada pela Portaria n^o. 016 / SGC / 2002.

Prof^o.Nereu Estanislau Burin
Professor responsável pela disciplina

Banca Examinadora:

Joana B.O. Quandt
Prof^a. Joana B. de Oliveira Quandt
Orientadora

Zatelli
Prof^a. Albertina Zatelli

WQuandt
Prof^o. Waldir Quandt

Florianópolis, 29 de Abril de 2002.

Sumário

Introdução

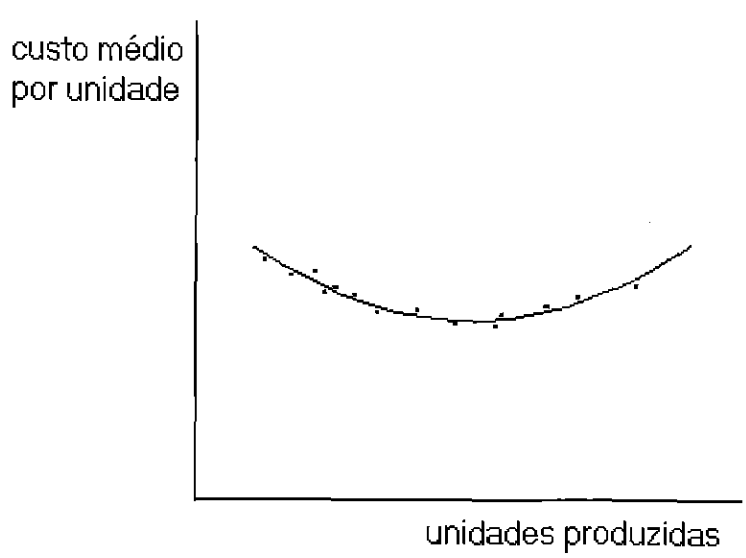
O primeiro problema de quadrados mínimos registrado na história foi em 1740, quando Jacques Cassini construiu uma lista de dados coletados em medições astronômicas feitas por vários astrônomos desde 140 A.C.. Esta lista continha diferentes resultados para a mesma quantidade. O problema então, tornou-se achar a "melhor equação" que satisfizesse a esses dados.

A primeira solução coerente apareceu em 1805, no livro de Legendre, onde o mesmo propôs a minimização dos erros das medições.

Em 1809, Gauss publicou um livro de órbitas planetárias, onde introduziu o Método de Quadrados Mínimos, citando seu uso desde 1795.

Laplace e outros matemáticos logo adotaram o método de quadrados mínimos, sendo que este se tornou indispensável para analisar dados astronômicos. Porém, demorou aproximadamente um século para que o método fosse realmente adotado na Biologia e nas Ciências Sociais.

O método de quadrados mínimos é utilizado na Administração, como por exemplo quando em uma empresa relacionamos o custo médio por unidade produzida com a quantidade de unidades produzidas em um dia. Uma curva típica de custo médio tem a aparência de uma parábola com a concavidade voltada para cima, como mostra o gráfico a seguir.



Neste trabalho, no capítulo 1, veremos algumas notações, definições e propriedades básicas de Álgebra Linear que serão necessárias para as deduções que serão feitas no capítulo seguinte. No capítulo 2 será visto o método de quadrados mínimos para determinar uma solução aproximada de um sistema linear incompatível $A\mathbf{x} = \mathbf{b}$. Para isto serão vistas dois tipos de decomposição de matrizes, a decomposição QR e a decomposição em valores singulares (D.V.S), e também a definição de pseudoinversa de uma matriz.

Ao longo do trabalho, serão vistos alguns exemplos e ilustrações para a melhor compreensão dos tópicos.

Capítulo 1

Notações, Definições e Propriedades Básicas

Neste capítulo veremos algumas notações, definições e resultados básicos importantes para o estudo do método de quadrados mínimos.

A maior parte das demonstrações serão omitidas, especialmente aquelas vistas nos cursos de Álgebra Linear, do curso de graduação em Matemática - habilitação Licenciatura. Para facilitar as demonstrações que serão vistas no capítulo 2, veremos aqui algumas definições usuais, como multiplicação de matrizes, porém com outro enfoque.

1. $\mathbb{R}^n$ é o conjunto dos vetores

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

2. $\mathbf{x}^T = (x_1, x_2, ..., x_n)$; vetor transposto.

3. $\langle \mathbf{x}, \mathbf{y} \rangle$: produto interno. Em geral é usado

   $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T\mathbf{y} = x_1y_1 + x_2y_2 + ... + x_ny_n$, produto interno usual em $\mathbb{R}^n$.

4. $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{y} \rangle} = \sqrt{\mathbf{x}^T\mathbf{x}} = \sqrt{x_1^2 + x_2^2 + ... + x_n^2}$,
   norma euclidiana do vetor $\mathbf{x}$, com o produto interno usual de $\mathbb{R}^n$.

5. $\mathbb{R}^{m\times n}$ é o conjunto das matrizes reais de ordem $m \times n$.

6. $dimV$: dimensão de um espaço vetorial V finitamente gerado.

7. $\mathbf{O}$: elemento neutro de um espaço vetorial V.

Definição 1.1. MULTIPLICAÇÃO DE UMA MATRIZ POR UM VETOR

Sejam $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}_{m\times n}$ e $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_{n\times 1}$.

Então, $A\mathbf{x} = \begin{bmatrix} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n \\ a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n \\ \vdots & & \vdots & & \ddots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \dots & + & a_{mn}x_n \end{bmatrix}_{m\times 1}$,

ou seja,

$$A\mathbf{x} = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \dots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}.$$

Portanto, $A\mathbf{x}$ é uma combinação linear das colunas de A, e os coeficientes que aparecem na combinação linear são as coordenadas de $\mathbf{x}$.

Sejam $\mathbf{a}_1$, $\mathbf{a}_2$, ..., $\mathbf{a}_n$ as colunas de A.

Então,

$$A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \dots + x_n\mathbf{a}_n. \tag{1.1}$$

□

Teorema 1.2. MULTIPLICAÇÃO DE MATRIZES

Sejam A uma matriz $m \times n$, B uma matriz $n \times p$ e C uma matriz $m \times p$, tal que $C = AB$. Então a k-ésima coluna de C é uma combinação linear das colunas de A, cujos coeficientes são os elementos da k-ésima coluna de B, para todo k tal que $1 \leq k \leq p$.

Demonstração.

Sejam

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \quad , \quad B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1p} \\ b_{21} & b_{22} & \dots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{np} \end{bmatrix}$$

e

$$C = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1p} \\ c_{21} & c_{22} & \dots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \dots & c_{mp} \end{bmatrix}.$$

Multiplicando as matrizes A e B temos:

$$C = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + \dots + a_{1n}b_{n1} & \dots & a_{11}b_{1p} + a_{12}b_{2p} + \dots + a_{1n}b_{np} \\ a_{21}b_{11} + a_{22}b_{21} + \dots + a_{2n}b_{n1} & \dots & a_{21}b_{1p} + a_{22}b_{2p} + \dots + a_{2n}b_{np} \\ \vdots & \ddots & \vdots \\ a_{m1}b_{11} + a_{m2}b_{21} + \dots + a_{mn}b_{n1} & \dots & a_{m1}b_{1p} + a_{m2}b_{2p} + \dots + a_{mn}b_{np} \end{bmatrix}$$

Sejam $\mathbf{a}_i$, $\mathbf{b}_k$, $\mathbf{c}_k$, $i = 1, \dots, n$; $k = 1, \dots, p$, as colunas de A, B e C, respectivamente.

$$C = \left[\ \mathbf{a}_1 b_{11} + \mathbf{a}_2 b_{21} + \dots + \mathbf{a}_n b_{n1} \quad \dots. \quad \mathbf{a}_1 b_{1p} + \mathbf{a}_2 b_{2p} + \dots + \mathbf{a}_n b_{np} \ \right]$$

Portanto,

$$\begin{aligned} \mathbf{c}_1 &= b_{11}\mathbf{a}_1 + b_{21}\mathbf{a}_2 + \dots + b_{n1}\mathbf{a}_n \\ \mathbf{c}_2 &= b_{12}\mathbf{a}_1 + b_{22}\mathbf{a}_2 + \dots + b_{n2}\mathbf{a}_n \\ &\vdots \\ \mathbf{c}_p &= b_{1p}\mathbf{a}_1 + b_{2p}\mathbf{a}_2 + \dots + b_{np}\mathbf{a}_n \end{aligned} \qquad (1.2)$$

Assim, cada coluna $\mathbf{c}_k$ de C é uma combinação linear das n colunas de A, com coeficientes b_{ik}, $1 \leq i \leq n$, $1 \leq k \leq p$.

$\square$

Utilizando a demonstração anterior podemos obter uma outra propriedade.

Propriedade 1.3. Sejam A uma matriz $m \times n$, B uma matriz $n \times p$ e C uma matriz $m \times p$, tal que $C = AB$.

Então, cada coluna k de AB pode ser obtida multiplicando-se a matriz A pela coluna k de B.

Demontração.

Utilizando a *definição 1.1* podemos reescrever (1.2) da seguinte forma:

$$\begin{aligned}
\mathbf{c}_1 &= A\mathbf{b}_1 \\
\mathbf{c}_2 &= A\mathbf{b}_2 \\
&\vdots \\
\mathbf{c}_p &= A\mathbf{b}_p
\end{aligned}$$

ou seja,

$$AB = C = [\mathbf{c}_1,\ \mathbf{c}_2, ...,\ \mathbf{c}_p] = [A\mathbf{b}_1,\ A\mathbf{b}_2, ...,\ A\mathbf{b}_p]\,. \tag{1.3}$$

$\square$

Definição 1.4. SISTEMAS DE EQUAÇÕES LINEARES

Sejam a_{ij} e b_i números reais, com $1 \leq i \leq m$ e $1 \leq j \leq n$. Então um sistema linear de m equações e n incógnitas é um sistema da seguinte forma:

$$\begin{aligned}
&a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n = b_1 \\
&a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n = b_2 \\
&\vdots \\
&a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n = b_m,
\end{aligned} \tag{1.4}$$

sendo $x_1, x_2, ..., x_n$ as incógnitas, a_{ij} os coeficientes e b_i os termos independentes do sistema linear.

Definição 1.5. NOTAÇÃO MATRICIAL DE UM SISTEMA LINEAR E MATRIZ AUMENTADA

Seja o sistema linear de m equações e n incógnitas dado acima em (1.4).

Se definirmos:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}_{m \times n}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_{n \times 1}$$

$$\text{e} \qquad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}_{m \times 1},$$

então

$$A\mathbf{x} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} =$$

$$= \begin{bmatrix} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n \\ a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n \\ \vdots & & \vdots & & \ddots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \dots & + & a_{mn}x_n \end{bmatrix},$$

ou seja, $A\mathbf{x}$ representa o "lado" esquerdo do sistema linear (1.4) dado acima.

Logo, o sistema linear (1.4) pode ser reescrito da seguinte forma:

$$A\mathbf{x} = \mathbf{b}.$$

A matriz

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right] = \left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right],$$

é chamada de **matriz aumentada** do sistema linear.

Teorema 1.6. Seja um sistema linear $A\mathbf{x} = \mathbf{b}$ com m equações e n incógnitas. Este sistema poderá ter:

- nenhuma solução; neste caso, dizemos que o sistema é Incompatível.
- exatamente um solução; neste caso, dizemos que o sistema é Compatível Determinado.
- uma infinidade de soluções; dizemos que o sistema é Compatível Indeterminado.

Teorema 1.7. POSTO DE UMA MATRIZ

Seja A uma matriz de ordem $m \times n$.
Seja B a matriz obtida a partir do escalonamento de A, B na forma escada.
Então as afirmações abaixo são equivalentes:
· r é o número máximo de colunas linearmente independentes de A.
· r é o número máximo de linhas linearmente independentes de A.
· r é o número de linhas não nulas de B.
Este número é chamado de *posto*(A).

$$\text{Notação: } Posto(A) = r.$$

Observação: $r \leqslant min\{m,n\}$, ou seja, $posto(A) \leqslant min\{m,n\}$.

Definição 1.8. Seja $[A \mid \mathbf{b}]$ a matriz aumentada do sistema $A\mathbf{x} = \mathbf{b}$, de m equações e n incógnitas.

Se $Posto[A \mid \mathbf{b}] \neq Posto(A)$, então o sistema é *Imcompatível.*
Se $Posto[A \mid \mathbf{b}] = Posto(A)$, então o sistema é *Compatível.*
E ainda, se:
• $posto(A) = n$, então o sistema é *Compatível Determinado.*
• se $r = posto(A) < n$, então existem $n - r$ variáveis livres e o sistema é *Compatível Indeterminado.*

Definição 1.9. ESPAÇO COLUNA DE UMA MATRIZ A

Seja $A = [\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n]$ uma matriz $m \times n$, sendo $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n$ as colunas de A.. O espaço coluna de A é o subespaço de $\mathbb{R}^m$ gerado pelas colunas de A, isto é, é o conjunto de todas as combinações lineares das colunas de A. Ele é denotado por $Col(A)$. Então,

$$Col(A) = \{A\mathbf{x} \;/\; \mathbf{x} \in \mathbb{R}^n\} = \{x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + ... + x_n\mathbf{a}_n;\; x_1, x_2, ..., x_n \in \mathbb{R}\}$$

O espaço coluna de A também é conhecido como Imagem de A.

$$Col(A) = Im(A)$$

Definição 1.10. ESPAÇO LINHA DE UMA MATRIZ A

Seja $A \in \mathbb{R}^{m\times n}$; o espaço linha de A é o subespaço de $\mathbb{R}^n$ gerado pelas linhas de A ; como as linhas de A são as colunas de A^T, então o espaço linha de A é dado por

$$Col(A^T) = \left\{A^T\mathbf{x} \;/\mathbf{x} \in \mathbb{R}^m\right\} = Im(A^T).$$

Definição 1.11. ESPAÇO NULO DE UMA MATRIZ A

Seja $A \in \mathbb{R}^{m\times n}$; o espaço nulo de A é o conjunto solução do sistema homogêneo $A\mathbf{x} = \mathbf{O}$; ele é subespaço de $\mathbb{R}^n$, e é denotado por $N(A)$. Então,

$$N(A) = \{\mathbf{x} \in \mathbb{R}^n \ / \ A\mathbf{x} = \mathbf{O}\ \}.$$

Definição 1.12. ESPAÇO NULO À ESQUERDA DE UMA MATRIZ A

Seja $A \in \mathbb{R}^{m\times n}$; segue-se que $A^T \in \mathbb{R}^{n\times m}$; o espaço nulo à esquerda de A é o espaço nulo de A^T, e é um subespaço de $\mathbb{R}^m$. Então,

$$N(A^T) = \left\{\mathbf{x} \in \mathbb{R}^m \ / \ A^T\mathbf{x} = \mathbf{O}\ \right\}.$$

Definição 1.13. OS SUBESPAÇOS FUNDAMENTAIS DE UMA MATRIZ A

Seja $A \in \mathbb{R}^{m\times n}$; são quatro os subespaços fundamentais de A:

- o espaço coluna de A, ou seja, $Col(A)$;
- o espaço linha de A, ou seja, $Col(A^T)$;
- o espaço nulo de A, $N(A)$;
- o espaço nulo à esquerda de A, $N(A^T)$.

Teorema 1.14. Um sistema linear de equações $A\mathbf{x} = \mathbf{b}$, é compatível se, e somente se $\mathbf{b}$ está no espaço coluna de A.

Demonstração.

$$\text{Sejam } A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \quad \text{e} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Por (1.1), temos que

$$A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + ... + x_n\mathbf{a}_n,$$

sendo $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n$ as colunas de A.

Portanto, o sistema linear $A\mathbf{x} = \mathbf{b}$ pode ser escrito como

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + ... + x_n\mathbf{a}_n = \mathbf{b}.$$

Suponhamos que $A\mathbf{x} = \mathbf{b}$ seja um sistema Compatível; então, existem valores $x_1, x_2, ..., x_n$ que satisfazem o sistema, ou seja, $\mathbf{b}$ é uma combinação linear das colunas de A, que geram o subespaço $Col(A)$. Logo, $\mathbf{b}$ pertence ao espaço coluna de A.

Suponhamos agora que $\mathbf{b}$ pertença ao espaço coluna de A, ou seja, existem $x_1, x_2, ..., x_n \in \mathbb{R}$ tais que

$$\mathbf{b} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + ... + x_n\mathbf{a}_n.$$

Mas $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + ... + x_n\mathbf{a}_n = A\mathbf{x}$.

Portanto, $\mathbf{x}$ é solução para o sistema $A\mathbf{x} = \mathbf{b}$.

□

Teorema 1.15. Seja $A\mathbf{x} = \mathbf{b}$ um sistema linear de m equações e n incógnitas.
Seja $posto(A) = r$.
Então:

- $dim(Col(A)) = r$
- $dim(N(A)) = n - r$
- $dim(Col(A^T)) = r$
- $dim(N(A^T)) = m - r$

Definição 1.16. VETORES ORTOGONAIS

Seja V um espaço vetorial com produto interno. Sejam $\mathbf{u}$ e $\mathbf{v}$ vetores em V. $\mathbf{u}$ e $\mathbf{v}$ são vetores ortogonais se $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.

Notação: $\mathbf{u} \perp \mathbf{v}$

Definição 1.17. CONJUNTO ORTOGONAL

Seja V um espaço vetorial com produto interno. Sejam $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ vetores em V. Se $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$, quando $i \neq j$, então $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ é um conjunto ortogonal de vetores.

Teorema 1.18. Seja V um espaço vetorial com produto interno. Seja $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ um conjunto ortogonal de vetores não nulos de V.

Então $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ são vetores linearmente independentes.

Definição 1.19. CONJUNTO ORTONORMAL

Seja V um espaço vetorial com produto interno. Seja $B = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ um subconjunto de V. Se B é um conjunto ortogonal e $||\mathbf{v}_i|| = 1$, $i = 1, 2, ..., n$, então B é um conjunto ortonormal de vetores.

Definição1.20. MATRIZ ORTOGONAL

Se Q é uma matriz quadrada $n \times n$ e $QQ^T = Q^TQ = I$, I : matriz identidade $n \times n$, então Q é chamada de **matriz ortogonal**.

Segue-se que

$$Q^{-1} = Q^T.$$

Observação: $|det(Q)| = 1$, sendo $det(Q)$ o determinante de Q.

Teorema 1.21. Seja Q uma matriz $n \times n$. As seguintes afirmações são equivalentes:

a) Q é ortogonal

b) as colunas de Q formam um conjunto ortonormal em $\mathbb{R}^n$.

c) as linhas de Q formam um conjunto ortonormal em $\mathbb{R}^{1\times n}$.

Propriedade 1.22. Seja A uma matriz $m \times n$ tal que as colunas de A formam um conjunto ortonormal em $\mathbb{R}^m$ e as linhas de A formam um conjunto ortonormal em $\mathbb{R}^{1\times n}$.

Sejam $\mathbf{u}$ e $\mathbf{v}$ vetores de $\mathbb{R}^n$. Então:

a) A preserva o produto interno entre os vetores, isto é,

$$\langle A\mathbf{u}, A\mathbf{v}\rangle = \langle \mathbf{u}, \mathbf{v}\rangle.$$

b) A preserva a norma de $\mathbf{v}$, isto é,

$$\|A\mathbf{v}\| = \|\mathbf{v}\|.$$

Definição 1.23. SUBESPAÇOS ORTOGONAIS

Sejam S e W dois subespaços de $\mathbb{R}^n$. S e W são ortogonais se $\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{x}^T\mathbf{y} = 0$, para todo $\mathbf{x} \in S$ e todo $\mathbf{y} \in W$.

Notação: $S \perp W$.

Definição 1.24. COMPLEMENTO ORTOGONAL DE UM SUBESPAÇO

Seja V um espaço vetorial com produto interno. Seja W um subespaço de V.

O complemento ortogonal de W é o conjunto de todos os vetores em V que são ortogonais a todos os vetores contidos em W, e é denotado por $W^\perp$. Então,

$$W^\perp = \{\mathbf{x} \in V \; / \; \langle \mathbf{x}, \mathbf{y}\rangle = 0, \text{ para qualquer } \mathbf{y} \in W\}.$$

Teorema 1.25. Seja V um espaço vetorial com produto interno e W um subespaço de V. Então:

a) $W^{\perp}$ também é um subespaço de V.
b) $W \cap W^{\perp} = \{\mathbf{O}\}$.
c) O complemento ortogonal de $W^{\perp}$ é W, ou seja, $(W^{\perp})^{\perp} = W$.
d) $\mathbf{v} \in W^{\perp}$ se e somente se ele é ortogonal a um conjunto de geradores de W.

Teorema 1.26. Seja A uma matriz $m \times n$. O espaço linha de A e o espaço nulo de A são complementos ortogonais um do outro e estão contidos em $\mathbb{R}^n$. E o espaço coluna de A é complemento ortogonal ao espaço nulo de A^T, estando ambos contidos em $\mathbb{R}^m$.

Isto é,

$$N(A) = \left(Col(A^T)\right)^{\perp} \text{ e } N(A^T) = (Col(A))^{\perp}$$

Demonstração.

Vamos mostrar apenas que $N(A) = \left(Col(A^T)\right)^{\perp}$.
Sejam $\mathbf{w} \in N(A)$ e $\mathbf{v} \in Col(A^T)$.
Então $A\mathbf{w} = \mathbf{O}$ e $\mathbf{v} = A^T\mathbf{x}$, para algum vetor $\mathbf{x} \in \mathbb{R}^m$.
Portanto,

$$\langle \mathbf{w}, \mathbf{v} \rangle = \mathbf{w}^T\mathbf{v} = \mathbf{w}^T\left(A^T\mathbf{x}\right) = \left(\mathbf{w}^T A^T\right)\mathbf{x} = (A\mathbf{w})^T\mathbf{x} = \mathbf{O}^T\mathbf{x} = 0,$$

ou seja,

$$N(A) \perp Col(A^T).$$

Isso implica que $N(A) \subset \left(Col(A^T)\right)^{\perp}$.

Seja $\mathbf{v} \in \left(Col(A^T)\right)^{\perp}$; então $\mathbf{v}$ é ortogonal a cada elemento do espaço linha de A. Em particular, $\mathbf{v}$ é ortogonal a cada coluna de A^T.

Porém as colunas de A^T são as linhas de A; logo $\mathbf{v}$ é ortogonal a cada linha de A.

Portanto $A\mathbf{v} = \mathbf{O}$, isto é, $\mathbf{v} \in N(A)$.
Assim

$$\left(Col(A^T)\right)^{\perp} \subset N(A).$$

Logo

$$N(A) = \left(Col(A^T)\right)^{\perp}.$$

Isso prova a primeira afirmação. A segunda segue da primeira substituindo-se A^T pela matriz B.

□

Teorema 1.27. Seja V um espaço vetorial de dimensão finita n com produto interno. Seja W um subespaço de V. Então,

$$V = W \oplus W^{\perp}$$

e

$$dim(W) + dim(W^{\perp}) = dim(V) = n$$

Além disso, se $\{\mathbf{x}_1,\ \mathbf{x}_2, ...,\ \mathbf{x}_r\}$ é base de W e $\{\mathbf{x}_{r+1},\ \mathbf{x}_{r+2}, ...,\ \mathbf{x}_n\}$ é base de $W^{\perp}$, então $\{\mathbf{x}_1,\ \mathbf{x}_2, ...,\ \mathbf{x}_r,\ \mathbf{x}_{r+1}, ...,\ \mathbf{x}_n\}$ é base de $\mathbb{R}^n$.

Observação: V é soma direta de W e $W^{\perp}$, isto é, todo vetor em V pode ser expresso de maneira única como uma soma $\mathbf{u} + \mathbf{v}$, sendo $\mathbf{u} \in W$e $\mathbf{v} \in W^{\perp}$.

Teorema 1.28. TEOREMA DA DIMENSÃO

Seja A uma matriz $m \times n$. Então

$$dim(Col(A)) + dim(N(A)) = n$$

Exemplo1.1. Seja $A = \begin{bmatrix} 2 & 3 & 1 \\ 4 & 5 & 6 \end{bmatrix} \in \mathbb{R}^{2\times3}$. Então $A^T = \begin{bmatrix} 2 & 4 \\ 3 & 5 \\ 1 & 6 \end{bmatrix} \in \mathbb{R}^{3\times2}$.

Sabe-se que $N(A) \subseteq \mathbb{R}^3$, $Col(A^T) \subseteq \mathbb{R}^3$, $Col(A) \subseteq \mathbb{R}^2$ e $N(A^T) \subseteq \mathbb{R}^2$.

$N(A) = \{\mathbf{v} \in \mathbb{R}^3/A\mathbf{v} = \mathbf{O}\} \neq \emptyset$, pois todo sistema homogêneo é compatível, já que $\mathbf{v} = \mathbf{O}$ é solução.

A matriz aumentada do sistema $A\mathbf{v} = \mathbf{O}$ é dada por:

$$\left[\begin{array}{ccc|c} 2 & 3 & 1 & 0 \\ 4 & 5 & 6 & 0 \end{array}\right].$$

Escalonando a matriz aumentada acima, através de operações elementares, temos:

$$\left[\begin{array}{ccc|c} 2 & 3 & 1 & 0 \\ 4 & 5 & 6 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 2 & 3 & 1 & 0 \\ 0 & -1 & 4 & 0 \end{array}\right] \longrightarrow$$

$$\left[\begin{array}{ccc|c} 2 & 3 & 1 & 0 \\ 0 & 1 & -4 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 2 & 0 & 13 & 0 \\ 0 & 1 & -4 & 0 \end{array}\right].$$

O sistema tem uma infinidade de soluções:

$$x = -\frac{13}{2}z \quad \text{e} \quad y = 4z \ , \ \ z \in \mathbb{R}$$

Portanto

$$N(A) = \left\{(-\frac{13}{2}z, 4z, z); \ z \in \mathbb{R}\right\} = \{z(-13, 8, 2); \ z \in \mathbb{R}\}$$

$$N(A) = [(-13, 8, 2)] .$$

Seja $\mathbf{v}_1 = (-13, 8, 2)^T$

Então $dim(N(A)) = 1$.

$N(A)$ é a reta que passa pela origem e cujo vetor direção é $\mathbf{v}_1$.

Sabemos que $Col(A^T) = \left\{\mathbf{b} \in \mathbb{R}^3/ \ \exists \ \mathbf{v} \in \mathbb{R}^2 \text{ satisfazendo } A^T\mathbf{v} = \mathbf{b}\right\}$.

Escalonando a matriz aumentada do sistema $A^T\mathbf{v} = \mathbf{b}$, obtemos:

$$
\left[\begin{array}{cc|c} 2 & 4 & b_1 \\ 3 & 5 & b_2 \\ 1 & 6 & b_3 \end{array}\right] \longrightarrow \left[\begin{array}{cc|c} 1 & 6 & b_3 \\ 3 & 5 & b_2 \\ 2 & 4 & b_1 \end{array}\right] \longrightarrow \left[\begin{array}{cc|c} 1 & 6 & b_3 \\ 0 & -13 & b_2 - 3b_3 \\ 2 & 4 & b_1 \end{array}\right] \longrightarrow
$$

$$
\longrightarrow \left[\begin{array}{cc|c} 1 & 6 & b_3 \\ 0 & -13 & b_2 - 3b_3 \\ 0 & -8 & b_1 - 2b_3 \end{array}\right] \longrightarrow \left[\begin{array}{cc|c} 1 & 6 & b_3 \\ 0 & 1 & \frac{1}{13}(-b_2 + 3b_3) \\ 0 & -8 & b_1 - 2b_3 \end{array}\right] \longrightarrow
$$

$$
\longrightarrow \left[\begin{array}{cc|c} 1 & 6 & b_3 \\ 0 & 1 & \frac{1}{13}(-b_2 + 3b_3) \\ 0 & 0 & \frac{13b_1 - 8b_2 - 2b_3}{13} \end{array}\right] \tag{1.5}
$$

Neste caso existe uma restrição, pois para que o sistema seja compatível devemos ter

$$
Posto\,[A \mid \mathbf{b}] = Posto(A).
$$

Portanto, $\frac{1}{13}(13b_1 - 8b_2 - 2b_3) = 0$.

Deste modo

$$
Col(A^T) = \{(b_1, b_2, b_3) \in \mathbb{R}^3 \ / \ 13b_1 - 8b_2 - 2b_3 = 0\}
$$

Como $b_3 = \frac{13}{2}b_1 - 4b_2$, temos

$$
\begin{aligned}
Col(A^T) &= \left\{(b_1, b_2, \frac{13}{2}b_1 - 4b_2)/b_1, b_2 \in \mathbb{R}\right\} = \\
&= \left\{b_1(1, 0, \frac{13}{2}) + b_2(0, 1, -4); b_1, b_2 \in \mathbb{R}\right\}.
\end{aligned}
$$

$$
Col(A^T) = \{b_1(2, 0, 13) + b_2(0, 1, -4); b_1, b_2 \in \mathbb{R}\}.
$$

$Col(A^T) = [(2,0,13),(0,1,-4)]$.

Sejam $\mathbf{v}_2 = (2,0,13)^T$ e $\mathbf{v}_3 = (0,1,-4)^T$.

Já que $\mathbf{v}_2$ e $\mathbf{v}_3$ geram $Col(A^T)$ e são linearmente independentes, pois um não é múltiplo escalar do outro, temos que $\{\mathbf{v}_2, \mathbf{v}_3\}$ é base do espaço coluna de A^T.

Então $dim(Col(A^T)) = 2$.

$Col(A^T)$ é o plano gerado por $\mathbf{v}_2$ e $\mathbf{v}_3$, e que passa pela origem

Observamos que $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = 0$ e $\langle \mathbf{v}_1, \mathbf{v}_3 \rangle = 0$, ou seja, o gerador do espaço nulo de A é ortogonal aos geradores do espaço coluna de A^T, ou seja

$$Col\left(A^T\right) = \left(N\left(A\right)\right)^{\perp}$$

e

$$dim(N(A)) + dim(Col(A^T)) = dim(R^3) = 3.$$

Observação: $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ é base de $\mathbb{R}^3$.

Sabemos que $N(A^T) = \left\{\mathbf{v} \in \mathbb{R}^2 \;/\; A^T\mathbf{v} = \mathbf{O}\right\}$ e que $A^T\mathbf{v} = \mathbf{O}$ é um sistema compatível.

Com a matriz

$$\left[\begin{array}{cc|c} 2 & 4 & 0 \\ 3 & 5 & 0 \\ 1 & 6 & 0 \end{array}\right] \quad ,$$

utilizando o escalonamento feito anteriormente para obter (1.5), temos

$$\begin{aligned} x + 6y &= 0 \\ y &= 0, \end{aligned}$$

$$\text{ou seja, } x = y = 0.$$

O conjunto solução é a solução trivial.

Portanto $N\left(A^T\right) = \{(0,0)\}$ e $dim(N(A^T)) = 0$.

Sabemos que $Col(A) = \{A\mathbf{v} = \mathbf{b} \ / \ \mathbf{v} \in R^2\}$.

Escalonando a matriz aumentada do sistema $A\mathbf{v} = \mathbf{b}$, obtemos:

$$\left[\begin{array}{ccc|c} 2 & 3 & 1 & b_1 \\ 4 & 5 & 6 & b_2 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 2 & 3 & 1 & b_1 \\ 0 & -1 & 4 & b_2 - 2b_1 \end{array}\right] \longrightarrow$$

$$\longrightarrow \left[\begin{array}{ccc|c} 2 & 3 & 1 & b_1 \\ 0 & 1 & -4 & 2b_1 - b_2 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 2 & 0 & 13 & 3b_2 - 5b_1 \\ 0 & 1 & -4 & 2b_1 - b_2 \end{array}\right]$$

Neste caso, $posto(A) = 2$ e , portanto, não há restrição para $\mathbf{b}$.

Logo, $Col(A) = \mathbb{R}^2$, ou seja, o espaço coluna de A é o próprio $\mathbb{R}^2$.

Assim $dim(Col(A)) = 2$, e temos

$$dim\left(N\left(A^T\right)\right) + dim(Col\,(A)) = dim\,(\mathbb{R}^2) = 2.$$

$\square$

Teorema 1.29. Sejam uma matriz $A \in \mathbb{R}^{n \times n}$, $\mathbf{x}$, $\mathbf{b} \in \mathbb{R}^n$ e $A\mathbf{x} = \mathbf{b}$ um sistema linear. Então as seguintes afirmações são equivalentes:

- A é inversível ($det(A) \neq 0$).
- $Col(A) = \mathbb{R}^n$.
- as colunas de A são linearmente independentes e geram $\mathbb{R}^n$, e portanto formam uma base de $\mathbb{R}^n$.
- $A\mathbf{x} = \mathbf{O}$ tem apenas a solução trivial.
- $A\mathbf{x} = \mathbf{b}$ é compatível para qualquer vetor $\mathbf{b} \in \mathbb{R}^n$.
- $A\mathbf{x} = \mathbf{b}$ tem exatamente uma solução para qualquer vetor $\mathbf{b} \in \mathbb{R}^n$.

Teorema 1.30. TEOREMA DE PITÁGORAS GENERALIZADO

Se **u** e **v** são vetores ortogonais num espaço com produto interno, então

$$\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$$

Teorema 1.31. Seja V um espaço vetorial com produto interno e $\mathbf{v}$ um vetor em V. Seja $B = \{\mathbf{v}_1,\ \mathbf{v}_2, ...,\ \mathbf{v}_n\}$ uma **base ortogonal** de V. Então todo vetor $\mathbf{v} \in V$ pode ser escrito como combinação linear da base B da seguinte forma:

$$\mathbf{v} = \frac{\langle \mathbf{v}, \mathbf{v}_1\rangle}{\langle \mathbf{v}_1, \mathbf{v}_1\rangle}\mathbf{v}_1 + \frac{\langle \mathbf{v}, \mathbf{v}_2\rangle}{\langle \mathbf{v}_2, \mathbf{v}_2\rangle}\mathbf{v}_2 + ... + \frac{\langle \mathbf{v}, \mathbf{v}_n\rangle}{\langle \mathbf{v}_n, \mathbf{v}_n\rangle}\mathbf{v}_n.$$

Os coeficientes $\mathbf{a}_j = \dfrac{\langle \mathbf{v}, \mathbf{v}_j\rangle}{\langle \mathbf{v}_j, \mathbf{v}_j\rangle}, j = 1, 2, ..., n$ são chamados de **coeficientes de Fourier** de $\mathbf{v}$.

Teorema 1.32. Sejam V um espaço vetorial com produto interno e W um subespaço de V. Todo vetor $\mathbf{v}$ em V pode ser expresso de maneira única como:

$$\mathbf{v} = \mathbf{w}_1 + \mathbf{w}_2,$$

pois $V = W \oplus W^{\perp}$, sendo $\mathbf{w}_1 \in W$ e $\mathbf{w}_2 \in W^{\perp}$.
O vetor $\mathbf{w}_1$ é chamado de projeção ortogonal de $\mathbf{v}$ em W .

$$\text{Notação: } \mathbf{w}_1 = proj_W\mathbf{v}$$

O vetor $\mathbf{w}_2$ é chamado de componente de $\mathbf{v}$ ortogonal a W.

$$\text{Notação: } \mathbf{w}_2 = proj_{W^{\perp}}\mathbf{v}$$

Portanto

$$\mathbf{v} = proj_W\mathbf{v} + proj_{W^{\perp}}\mathbf{v}$$

Já que $\mathbf{w}_2 = \mathbf{v} - \mathbf{w}_1$, temos $proj_{W^{\perp}}\mathbf{v} = \mathbf{v} - proj_W\mathbf{v}$, ou seja,

$$\mathbf{v} = proj_W\mathbf{v} + (\mathbf{v} - proj_W\mathbf{v}).$$

Teorema 1.33. PROJEÇÃO ORTOGONAL DE UM VETOR EM UM SUBESPAÇO

Seja V um espaço vetorial com produto interno. Seja $B = \{\mathbf{v}_1,\ \mathbf{v}_2, ...,\ \mathbf{v}_r\}$ uma **base ortogonal** de um subespaço W de V. Seja $\mathbf{v} \in V$. A projeção de $\mathbf{v}$ em W é denotada por $proj_W \mathbf{v}$ e é dada pela fórmula;

$$p = proj_W \mathbf{v} = \frac{\langle \mathbf{v}, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 + \frac{\langle \mathbf{v}, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2 + ... + \frac{\langle \mathbf{v}, \mathbf{v}_r \rangle}{\langle \mathbf{v}_r, \mathbf{v}_r \rangle}\mathbf{v}_r.$$

Teorema 1.34. TEOREMA DA MELHOR APROXIMAÇÃO

Seja V um espaço vetorial com produto interno. Seja W um subespaço de V. Seja $\mathbf{u}$ um vetor em V e suponha que $\mathbf{u}$ não pertença a W. Então $proj_W \mathbf{u}$ é a melhor aproximação de $\mathbf{u}$ em W, isto é,

$$\|\mathbf{u} - proj_W \mathbf{u}\| < \|\mathbf{u} - \mathbf{w}\|,$$

para qualquer vetor $\mathbf{w}$ em W, diferente de $proj_W \mathbf{u}$.

Teorema 1.35. PROCESSO DE ORTOGONALIZAÇÃO DE GRAM-SCHMIDT

Seja V um espaço vetorial com produto interno. Seja $B = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ uma base qualquer de V.

Sejam $\mathbf{v}_1 = \mathbf{u}_1$ e

$$\mathbf{v}_n = \mathbf{u}_n - \frac{\langle \mathbf{u}_n, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 - \frac{\langle \mathbf{u}_n, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2 - ... - \frac{\langle \mathbf{u}_n, \mathbf{v}_{n-1} \rangle}{\langle \mathbf{v}_{n-1}, \mathbf{v}_{n-1} \rangle}\mathbf{v}_{n-1},\ n \geqslant 2.$$

Deste modo obtemos uma base ortogonal $\overline{B} = \{\mathbf{v}_1,\ \mathbf{v}_2, ...,\ \mathbf{v}_n\}$ de V.

Sejam $\mathbf{q}_1 = \dfrac{\mathbf{v}_1}{\|\mathbf{v}_1\|}$, $\mathbf{q}_2 = \dfrac{\mathbf{v}_2}{\|\mathbf{v}_2\|}$, ... , $\mathbf{q}_n = \dfrac{\mathbf{v}_n}{\|\mathbf{v}_n\|}$.

Então $\overline{\overline{B}} = \{\mathbf{q}_1,\ \mathbf{q}_2, ...,\ \mathbf{q}_n\}$ é uma base ortonormal de V.

Observação:

1) Como $proj_{\mathbf{v}_j}\mathbf{u}_i = \dfrac{\langle \mathbf{u}_i, \mathbf{v}_j \rangle}{\langle \mathbf{v}_j, \mathbf{v}_j \rangle}\mathbf{v}_j$, segue-se que $\mathbf{v}_n$ pode ser reescrito como:

$$\mathbf{v}_n = \mathbf{u}_n - \left(proj_{\mathbf{v}_1}\mathbf{u}_n + proj_{\mathbf{v}_2}\mathbf{u}_n + ... + proj_{\mathbf{v}_{n-1}}\mathbf{u}_n\right)$$

2) Para qualquer $1 \leq k \leq n$, $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}$ é uma base ortogonal que gera o mesmo subespaço gerado pela base $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$, ou seja,

$$[\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k] = [\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k],\ 1 \leq k \leq n$$

Corolário 1.36. Todo espaço vetorial V com dimensão finita tem uma base ortonormal.

Teorema 1.37. PROCESSO DE GRAM-SCHMIDT MODIFICADO

Seja V um espaço vetorial com produto interno. Seja $B = \{\mathbf{u}_1,\ \mathbf{u}_2, ...,\ \mathbf{u}_n\}$ uma base **qualquer** de V. Por definição:

$$\mathbf{q}_1 = \frac{1}{\|\mathbf{u}_1\|}\mathbf{u}_1.$$

O próximo passo é subtrair de cada vetor $\mathbf{u}_k$, $k = 2, 3, ..., n$, a projeção ortogonal de $\mathbf{u}_k$ sobre $\mathbf{q}_1$.

Como $proj_{\mathbf{q}_1}\mathbf{u}_k = \dfrac{\langle \mathbf{u}_k, \mathbf{q}_1 \rangle}{\langle \mathbf{q}_1, \mathbf{q}_1 \rangle}\mathbf{q}_1$, temos que:

$$\mathbf{u}_k^1 = \mathbf{u}_k - \frac{\langle \mathbf{u}_k, \mathbf{q}_1 \rangle}{\langle \mathbf{q}_1, \mathbf{q}_1 \rangle}\mathbf{q}_1\ ,\quad k = 2, 3, ..., n$$

Deste modo obtemos novos vetores $\mathbf{u}_k^1$, $k = 2, 3, ..., n$, ortogonais a $\mathbf{q}_1$.

Agora definimos

$$\mathbf{q}_2 = \frac{1}{\|\mathbf{u}_2^1\|}\mathbf{u}_2^1 \ ,$$

sendo que $\mathbf{q}_2$ já é ortogonal a $\mathbf{q}_1$.

Analogamente, subtraímos de cada novo vetor $\mathbf{u}_k^1$, $k = 3, 4, ..., n$, a projeção ortogonal de $\mathbf{u}_k^1$ sobre $\mathbf{q}_1$:

$$\mathbf{u}_k^2 = \mathbf{u}_k^1 - \frac{\langle \mathbf{u}_k^1, \mathbf{q}_2 \rangle}{\langle \mathbf{q}_2, \mathbf{q}_2 \rangle}\mathbf{q}_2 \ , \quad k = 3, 4, ..., n.$$

Assim obtemos novos vetores $\mathbf{u}_k^2$, $k = 3, 4, ..., n$, ortogonais a $\mathbf{q}_2$.

Deste modo, sucessivamente, definimos $\mathbf{q}_3, \mathbf{q}_4, ..., \mathbf{q}_n$, sendo

$$\mathbf{q}_n = \frac{1}{\|\mathbf{u}_n^{n-1}\|}\mathbf{u}_n^{n-1}$$

ortogonal a $\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_{n-1}$.

Logo, $\overline{B} = \{\mathbf{q}_1, \ \mathbf{q}_2, ..., \ \mathbf{q}_n\}$ é uma base ortonormal de vetores de V.

Definição 1.38. AUTOVETORES E AUTOVALORES

Seja A uma matriz $n \times n$. Um escalar λ é um autovalor ou valor característico de A se existe um vetor não-nulo $\mathbf{x}$ tal que $A\mathbf{x} = \lambda\mathbf{x}$. O vetor $\mathbf{x}$ é um autovetor ou vetor característico associado a λ.

Teorema 1.39. TEOREMA ESPECTRAL PARA MATRIZES SIMÉTRICAS

Seja $A \in \mathbb{R}^{n \times n}$ simétrica. Então:

a) A possui n autovalores $\lambda_1, \lambda_2, ..., \lambda_n$ e são todos reais;

b) os autovetores de A associados a autovalores distintos são ortogonais entre si;

c) é sempre possível obter uma **base ortonormal de autovetores** $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$;

d) vale a relação $Q^T A Q = D$;

as colunas de Q são $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ e, portanto, Q é uma matriz ortogonal;

D é uma matriz diagonal, sendo que a diagonal principal desta matriz é composta por $\lambda_1, \lambda_2, ..., \lambda_n$.

Observação:

do item *d)* acima segue-se que A pode ser decomposta na forma, $A = QDQ^T$.

Teorema 1.40. Se $\mathbf{v}$ é um autovetor de $A^T A$ associados a um autovalor não nulo λ, então $A\mathbf{v}$ é um autovetor de AA^T associado ao mesmo autovalor.

Demonstração.

Seja $A^T A\mathbf{v} = \lambda\mathbf{v}$. Então

$$AA^T (A\mathbf{v}) = A\left(A^T A\mathbf{v}\right) = A(\lambda\mathbf{v}) = \lambda A\mathbf{v}.$$

$\square$

Teorema 1.41. Seja A uma matriz $m \times n$. Então $A^T A$ é uma matriz simétrica $n \times n$, tal que

a) $N(A^T A) = N(A)$
b) $Col(A^T A) = Col(A^T)$.

Demonstração.

Seja A uma matriz de ordem $m \times n$; logo, A^T é uma matriz de ordem $n \times m$.

Então, $A^T A$ é de ordem $n \times n$ e

$$\left(A^T A\right)^T = A^T \left(A^T\right)^T = A^T A,$$

ou seja, $A^T A$ é uma matriz simétrica.

a) Será provado agora que $N(A^T A) = N(A)$.
Suponha que $\mathbf{x}$ pertença ao espaço nulo de $A^T A$.
Então,

$$A^T A\mathbf{x} = \mathbf{O}, \quad \text{ou seja,} \quad A^T (A\mathbf{x}) = \mathbf{O}\ .$$

Logo,

$$A\mathbf{x} \ \in N\left(A^T\right).$$

Mas, por (1.1),

$$A\mathbf{x} \in Col(A).$$

Então, $A\mathbf{x} \in Col(A)$ e $A\mathbf{x} \in N\left(A^T\right)$.
Como, pelo *teorema 1.26*, $N\left(A^T\right)$ é o complemento ortogonal de $Col(A)$, segue, pelo *teorema 1.25*, que

$$A\mathbf{x} = \mathbf{O}.$$

Logo,

$$\mathbf{x} \ \in N(A).$$

Portanto,

$$N(A^T A) \subset N(A). \tag{1.6}$$

Seja $\mathbf{x}$ pertencente a $N(A)$; então,

$$A\mathbf{x} = \mathbf{O}.$$

Assim, $A^T (A\mathbf{x}) = A^T\mathbf{O} = \mathbf{O}$, ou seja,

$$\mathbf{x} \in N(A^T A).$$

Portanto,

$$N(A^T A) \subset N(A). \tag{1.7}$$

Logo, de (1.6) e (1.7), segue-se que

$$N(A) = N(A^T A).$$

b) Temos, do *teorema 1.26*, que

$$Col(A^T A) = \left(N(A^T A)\right)^{\perp}.$$

Mas sabemos, do resultado anterior, que $N(A^T A) = N(A)$.
Deste modo,

$$Col(A^T A) = (N(A))^{\perp}.$$

Do *teorema 1.26*,

$$(N(A))^{\perp} = Col(A^T).$$

Logo,

$$Col(A^T A) = Col(A^T).$$

□

Observação: Do teorema acima, segue-se que

$$Posto(A^T A) = Posto(A^T) = Posto(A) = Posto(AA^T).$$

Teorema 1.42. Seja A uma matriz $m \times n$ e $posto(A) = n$. Então $A^T A$ é uma matriz $n \times n$ inversível.

Capítulo 2

Método de Quadrados Mínimos

2.1. INTRODUÇÃO

Neste capítulo veremos o método de quadrados mínimos para a obtenção da solução aproximada de sistemas lineares incompatíveis e também algumas aplicações.

Para isso deverão ser vistas algumas decomposições de matrizes, que são as "ferramentas" necessárias para a obtenção de uma solução aproximada de um sistema incompatível pelo método citado acima.

2.2 O PROBLEMA DE QUADRADOS MÍNIMOS

Seja $A\mathbf{x} = \mathbf{b}$ um sistema linear de m equações e n incógnitas incompatível.

Já que não existe solução para $A\mathbf{x} = \mathbf{b}$, vamos determinar um vetor $\mathbf{x}$ tal que $\|A\mathbf{x} - \mathbf{b}\|$ seja minimizada, ou seja, tal que a distância entre $\mathbf{b}$ e $A\mathbf{x}$ seja a menor possível. Quanto menor for esta distância, melhor será a solução aproximada.

Seja $\widehat{\mathbf{x}}$ essa solução aproximada. Podemos dizer que $A\widehat{\mathbf{x}} - \mathbf{b}$ é o "vetor erro" que resulta do fato de $\widehat{\mathbf{x}}$ ser a melhor solução aproximada do sistema $A\mathbf{x} = \mathbf{b}$.

Se o sistema fosse compatível e $\widehat{\mathbf{x}}$ fosse a solução exata, então o erro seria nulo:

$$\mathbf{e} = A\widehat{\mathbf{x}} - \mathbf{b} = \mathbf{O}.$$

O problema formulado como foi feito acima, isto é, dado um sistema $A\mathbf{x} = \mathbf{b}$ de m equações e n incógnitas, determinar um vetor $\widehat{\mathbf{x}}$ que minimiza $\|A\mathbf{x} - \mathbf{b}\|$, é chamado **problema de quadrados mínimos**.

Um tal vetor é chamado uma **solução de quadrados mínimos** de $A\mathbf{x} = \mathbf{b}$.

Observação: Para entender a origem do termo *quadrados mínimos*, seja $\widehat{\mathbf{x}}$ uma solução deste tipo e $\mathbf{e} = A\widehat{\mathbf{x}} - \mathbf{b}$, o vetor erro.

Se $\mathbf{e} = (e_1,\ e_2, ...,\ e_m)$, então a solução de quadrados mínimos minimiza

$$\|\mathbf{e}\| = (e_1^2 + e_2^2 + ... + e_m^2)^{\frac{1}{2}}.$$

Portanto,

$$\|\mathbf{e}\|^2 = e_1^2 + e_2^2 + ... + e_m^2 \qquad \text{também é minimizada,}$$

e para que isto ocorra, cada parcela deve ser o menor possível.
Por isso é dado o nome de Quadrados Mínimos .

2.2.1. SOLUÇÃO DO PROBLEMA DE QUADRADOS MÍNIMOS.

Seja $W = Col(A)$.

Se $A\mathbf{x} = \mathbf{b}$ é incompatível, então, pelo *teorema 1.14*, $\mathbf{b} \notin W$, ou seja, $\mathbf{b}$ não é combinação linear das colunas de A.

A solução de quadrados mínimos é um vetor $\widehat{\mathbf{x}}$ em $\mathbb{R}^n$ tal que $A\widehat{\mathbf{x}}$ seja o vetor "mais próximo" do vetor $\mathbf{b}$.

Pelo *teorema 1.34* temos que $proj_W\mathbf{b}$ é o vetor em W mais próximo de $\mathbf{b}$.

Portanto, uma solução de quadrados mínimos deve satisfazer o sistema

$$A\mathbf{x} = proj_W\mathbf{b} = \mathbf{p}. \tag{2.1}$$

Esse sistema pode ser resolvido diretamente. Primeiramente temos que calcular o vetor $\mathbf{p}$. Para isto precisamos antes obter uma base ortogonal do espaço coluna de A que será utilizada no cálculo da projeção ortogonal de $\mathbf{b}$ em W. Só então resolveríamos o sistema $A\mathbf{x} = \mathbf{p}$.

Por causa das questões computacionais citadas acima, será dada uma nova abordagem para a obtenção da solução do problema.

Pelo *teorema 1.34* temos que

$$\mathbf{b} - proj_W\mathbf{b} \in W^{\perp}$$

e pelo *teorema 1.26* que $W^{\perp} = N(A^T)$.
Segue-se que

$$\mathbf{b} - A\mathbf{x} \in N(A^T),$$

ou seja,

$$A^T(\mathbf{b} - A\mathbf{x}) = \mathbf{O}.$$

Portanto,

$$A^T\mathbf{b} - A^TA\mathbf{x} = \mathbf{O}.$$

Logo, determinar uma solução de quadrados mínimos de $A\mathbf{x} = \mathbf{b}$ é a mesma coisa que achar a solução exata do sistema

$$A^TA\mathbf{x} = A^T\mathbf{b}.$$

O sistema $A^TA\mathbf{x} = A^T\mathbf{b}$ é chamado **sistema normal** associado a $A\mathbf{x} = \mathbf{b}$, e as equações que o compõem são chamadas *equações normais;* esse sistema envolve **n** equações e **n** incógnitas, pois A^TA é $n \times n$.

□

Teorema 2.1. Seja $A\mathbf{x} = \mathbf{b}$ um sistema linear de m equações e n incógnitas.
a) O sistema normal $A^TA\mathbf{x} = A^T\mathbf{b}$ é sempre compatível.
b) Qualquer solução $\mathbf{x}$ é uma solução do sistema normal se e somente se $\mathbf{x}$ é solução de quadrados mínimos de $A\mathbf{x} = \mathbf{b}$.

Demonstração.

a) Consideramos a matriz aumentada $\left[A^TA \;\; \vdots \;\; A^T\mathbf{b}\right]$.
As colunas de A^TA geram $Col(A^TA)$, e sabemos, pela *definição 1.1*, que $A^T\mathbf{b}$ é uma combinação linear das colunas de A^T.
Sejam $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ as colunas de A^T. Então

$$A^T\mathbf{b} = b_1\mathbf{a}_1 + b_2\mathbf{a}_2 + \ldots + b_m\mathbf{a}_m,$$

ou seja, $A^T b$ é uma combinação linear das colunas de A^T.
Logo,

$$A^T\mathbf{b} \in Col(A^T).$$

Como, pelo *teorema 1.41*,

$$Col(A^T) = Col(A^T A),$$

temos que

$$A^T\mathbf{b} \in Col(A^T A),$$

e pelo *teorema 1.14*, temos que

$$A^T A\mathbf{x} = A^T\mathbf{b} \text{ é sempre compatível.}$$

b) Considere $\mathbf{x}_1$ e $\mathbf{x}_2$ soluções do sistema normal. Então,

$$A^T A\mathbf{x}_1 = A^T\mathbf{b} \tag{2.2}$$
$$A^T A\mathbf{x}_2 = A^T\mathbf{b} \tag{2.3}$$

Subtraindo (2.3) de (2.2) temos

$$A^T A(\mathbf{x}_1 - \mathbf{x}_2) = A^T\mathbf{b} - A^T\mathbf{b} = \mathbf{O},$$

ou seja,

$$A^T A(\mathbf{x}_1 - \mathbf{x}_2) = \mathbf{O}.$$

Portanto, $(\mathbf{x}_1 - \mathbf{x}_2) \in N(A^T A)$.
Como , pelo *teorema 1.41*, $N(A^T A) = N(A)$, concluímos que $(\mathbf{x}_1 - \mathbf{x}_2) \in N(A)$, ou seja,

$$A(\mathbf{x}_1 - \mathbf{x}_2) = \mathbf{O}.$$

Seja $\mathbf{x}_1 - \mathbf{x}_2 = \mathbf{y}$. Deste modo $\mathbf{x}_1 = \mathbf{y} + \mathbf{x}_2$.

Então $A\mathbf{x}_1 - \mathbf{b} = A(\mathbf{y} + \mathbf{x}_2) - \mathbf{b} = A\mathbf{y} + A\mathbf{x}_2 - \mathbf{b} = A\mathbf{x}_2 - \mathbf{b}$.

Logo, $A\mathbf{x}_1 - \mathbf{b} = A\mathbf{x}_2 - \mathbf{b}$.

Isto significa que $\|\mathbf{v}\|^2 = \|A\mathbf{x} - \mathbf{b}\|^2$ terá o mesmo valor tanto para $\mathbf{x} = \mathbf{x}_1$ quanto para $\mathbf{x} = \mathbf{x}_2$.

Portanto, qualquer solução de $A^T A\mathbf{x} = A^T\mathbf{b}$ é uma solução de quadrados mínimos de $A\mathbf{x} = \mathbf{b}$.

Já vimos anteriormente, nas deduções feitas para estabelecer o sistema normal, que qualquer solução de quadrados mínimos de um sistema $A\mathbf{x} = \mathbf{b}$ é solução do sistema normal.

$\square$

Em alguns casos podemos garantir que um sistema linear tenha uma única solução de quadrados mínimos.

O próximo teorema nos dá um resultado útil para tais casos de unicidade.

Teorema 2.2. Um sistema linear $A\mathbf{x} = \mathbf{b}$ tem uma única solução de quadrados mínimos, se e somente se $Posto(A) = n$.

E ainda, esta solução é dada por:

$$\mathbf{x} = \left(A^T A\right)^{-1} A^T\mathbf{b}.$$

Demonstração.

Suponhamos que $posto(A) = n$; então, pelo *teorema 1.42*, temos que $A^T A$ é inversível.

Assim, do sistema normal $A^T A\mathbf{x} = A^T\mathbf{b}$ segue-se que

$$\mathbf{x} = \left(A^T A\right)^{-1} A^T\mathbf{b},$$

e da unicidade da inversa de uma matriz, segue-se que $\mathbf{x}$ a única solução de quadrados mínimos de $A\mathbf{x} = \mathbf{b}$.

Suponhamos agora que exista uma única solução de quadrados mínimos de $A\mathbf{x} = \mathbf{b}$.

Se $\widehat{\mathbf{x}}$ é essa solução, então

$$A\widehat{\mathbf{x}} = proj_W \mathbf{b} = \mathbf{p}, \qquad \text{sendo } W = Col(A).$$

Suponhamos que $posto(A) < n$.
Como, pelo *teorema 1.28*, $n = dim(N(A)) + posto(A)$, segue-se que

$$dim(N(A)) > 0.$$

Logo, existe $\overline{\mathbf{x}} \neq \mathbf{O}$ pertencente ao espaço nulo de A, isto é, existe $\overline{\mathbf{x}} \neq \mathbf{O}$ tal que $A\overline{\mathbf{x}} = \mathbf{O}$.
Mas, $A\overline{\mathbf{x}} + A\widehat{\mathbf{x}} = \mathbf{O} + \mathbf{p} = \mathbf{p}$.
Segue-se que

$$A\left(\overline{\mathbf{x}} + \widehat{\mathbf{x}}\right) = \mathbf{p},$$

isto é, $\overline{\mathbf{x}} + \widehat{\mathbf{x}}$ é também solução de quadrados mínimos de $A\mathbf{x} = \mathbf{b}$,

e como $\overline{\mathbf{x}} \neq \mathbf{O}$, então $\overline{\mathbf{x}} + \widehat{\mathbf{x}} \neq \widehat{\mathbf{x}}$, o que é absurdo.

Logo,

$$posto(A) = n.$$

□

Observação:
- Para obter a solução $\widehat{\mathbf{x}}$ resolvemos o sistema $A^T A\mathbf{x} = A^T\mathbf{b}$.
- Se $posto(A) < n$, então o sistema normal será indeterminado.

Exemplo 2.1. SOLUÇÃO DE QUADRADOS MÍNIMOS DE UM SISTEMA LINEAR

$$\text{Sejam} \quad A = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 1 & -2 \\ 1 & 1 & 0 \\ 1 & 1 & -1 \end{bmatrix}_{4\times 3} \qquad \text{e} \qquad \mathbf{b} = \begin{bmatrix} 6 \\ 0 \\ 9 \\ 3 \end{bmatrix}_{4\times 1}$$

A tem as colunas linearmente independentes. Logo, $posto(A) = 3$, de forma que existe uma única solução de quadrados mínimos.

$$A^TA = \begin{bmatrix} 1 & 2 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ -1 & -2 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 2 & 1 & -2 \\ 1 & 1 & 0 \\ 1 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 7 & 4 & -6 \\ 4 & 3 & -3 \\ -6 & -3 & 6 \end{bmatrix}$$

$$A^T\mathbf{b} = \begin{bmatrix} 1 & 2 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ -1 & -2 & 0 & -1 \end{bmatrix} \begin{bmatrix} 6 \\ 0 \\ 9 \\ 3 \end{bmatrix} = \begin{bmatrix} 18 \\ 12 \\ -9 \end{bmatrix}.$$

Portanto o sistema normal $A^TA\mathbf{x} = A^T\mathbf{b}$ é:

$$\begin{bmatrix} 7 & 4 & -6 \\ 4 & 3 & -3 \\ -6 & -3 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 18 \\ 12 \\ -9 \end{bmatrix}.$$

Assim,

$$x_1 = 12 \ , \ x_2 = -3 \ , \ x_3 = 9.$$

Então

$\widehat{\mathbf{x}} = \begin{bmatrix} 12 \\ -3 \\ 9 \end{bmatrix}$ é a única solução de quadrados mínimos do sistema $A\mathbf{x} = \mathbf{b}$.

Seja $W = Col(A)$. Como $proj_W\mathbf{b} = A\mathbf{x}$, então:

$$proj_W\mathbf{b} = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 1 & -2 \\ 1 & 1 & 0 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 12 \\ -3 \\ 9 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 9 \\ 0 \end{bmatrix}.$$

$\square$

2.2.2. AJUSTE DE CURVAS

A aplicação mais comum das soluções de quadrados mínimos é o **ajuste de curvas**, como por exemplo, ajustar funções a dados obtidos através de experimentos.

Sejam $(x_1, y_1), (x_2, y_2), ..., (x_m, y_m)$ os m pontos resultantes de um certo número de medições, ou seja, quando $x = x_i$ foi observado que $y = y_i$, sendo $i = 1, 2, ..., m$.

O nosso objetivo é encontrar uma relação funcional $y = f(x)$ que represente os m pontos da melhor maneira possível. Esta função é chamada de **curva de quadrados mínimos**.

Por exemplo, se um corpo se move com velocidade constante v, a relação entre o tempo t e a posição y atingida pelo corpo é dada pela equação linear

$$y = \alpha + vt. \tag{2.4}$$

Suponha que medimos y em vários instantes t e obtivemos os seguintes resultados

t	0	3	5	8	10
y	2	5	6	9	11

Estes resultados estão marcados no *gráfico 2.1*. Percebemos que os pontos quase formam uma linha reta. Podem exister dois motivos pelos quais os pontos não estão totalmente alinhados, já que a relação entre t e α é linear:

1. a velocidade não ser constante, fazendo com que (2.4) seja falsa.
2. erros de medição.

Vamos levar em consideração o segundo item. Se t e y tivessem sido medidos corretamente, então a relação entre eles seria dada por (2.4). Como as medidas estão sujeitas a erros experimentais, a equação (2.4) não é satisfeita, isto é, não conseguimos achar valores exatos para α e v.

Portanto, temos que deduzir os "melhores valores possíveis" para α e v, ou seja, achar a reta que melhor se adapta aos dados obtidos no experimento.

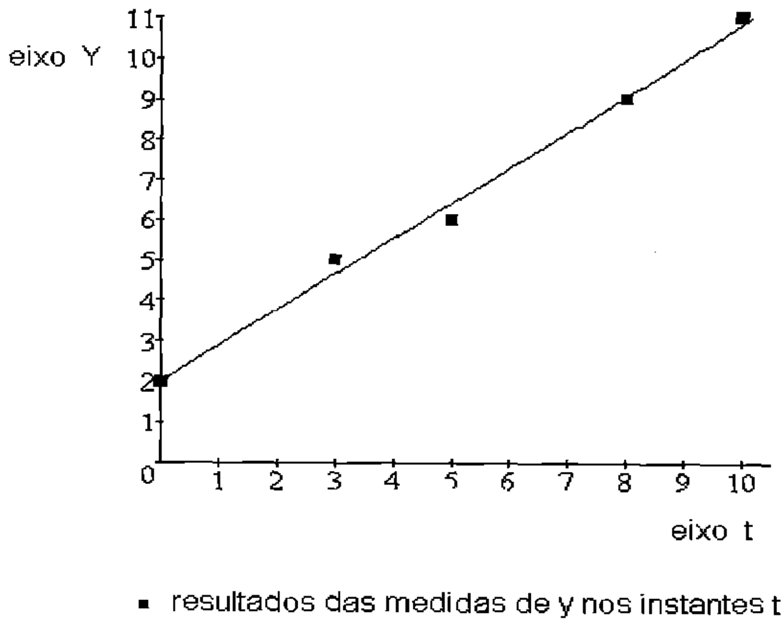


Gráfico 2.1

De forma geral, suponha que tenhamos colocado m pontos no plano cartesiano, e que a relação mais aproximada entre x e y seja dada por uma equação linear, ou seja, o gráfico é aproximadamente uma reta:

$$y = cx + d.$$

Na prática, os m pontos obtidos da experiência, raramente parecem estar alinhados com precisão sobre uma reta, mesmo quando as medições são feitas cuidadosamente. É inevitável o aparecimento de erros.

Se não houvesse erro a reta passaria pelos m pontos.
Então valeria o sistema linear

$$\begin{gathered} cx_1 + d = y_1 \\ cx_2 + d = y_2 \\ \vdots \\ cx_m + d = y_m. \end{gathered}$$

Sejam

$$A = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{bmatrix}_{m\times 2} \quad , \quad \mathbf{x} = \begin{bmatrix} c \\ d \end{bmatrix}_{2\times 1} \quad \text{e} \quad \mathbf{b} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}_{m\times 1} .$$

Assim, o sistema de m equações pode ser escrito na forma matricial

$$A\mathbf{x} = \mathbf{b}. \tag{2.5}$$

Cada ponto (x_i, y_i), obtido experimentalmente, corresponde a um ponto $(x_i, cx_i + d)$ pertencente à reta de quadrados mínimos. Portanto, y_i é um valor observado no experimento, e $cx_i + d$ é o valor obtido na *reta de quadrados mínimos*.

A diferença entre esses dois valores é

$$y_i - (cx_i + d) = e_i \qquad , i = 1, ..., m.$$

Assim,

$$y_i = cx_i + d + e_i \qquad , \ i = 1, ..., m,$$

sendo e_i o desvio vertical do ponto (x_i, y_i) até a reta procurada.

Temos que

$$\mathrm{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{bmatrix}_{m\times 1} \qquad \text{é o vetor erro.}$$

Exemplo 2.2. AJUSTE DE DADOS POR UMA RETA

Um determinado experimento produziu os dados:

$$P_1 = (0,0) \; ; \; P_2 = (1,2) \text{ e } P_3 = (2,7) \; .$$

Se pusermos estes pontos no gráfico cartesiano, podemos suspeitar que, a função que procuramos, para melhor se ajustar nos dados, seja uma reta

$$y = cx + d.$$

Suponha que exista uma reta que passe pelos três pontos. Deste modo temos:

$$\begin{array}{l} 0 = d \\ 2 = c1 + d \\ 7 = c2 + d \end{array} \qquad \text{, ou seja,}$$

$$\begin{bmatrix} 0 \\ 2 \\ 7 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix}.$$

Sejam

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} \quad , \quad \mathbf{x} = \begin{bmatrix} c \\ d \end{bmatrix} \quad \text{e} \quad \mathbf{b} = \begin{bmatrix} 0 \\ 2 \\ 7 \end{bmatrix}$$

Já que o sistema é Incompatível, ou seja, $\mathbf{b}$ não pertence ao espaço coluna de A, determinaremos uma solução aproximada pelo método de quadrados mínimos.

Escalonando a matriz A, obtemos a matriz reduzida

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{, ou seja,} \quad Posto(A) = n = 2.$$

Do *teorema 2.2* ,sabemos que existe uma única solução de quadrados mínimos. Temos que

$$A^T = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix}.$$

Portanto,

$$A^TA = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 3 \\ 3 & 3 \end{bmatrix}$$

e

$$A^T\mathbf{b} = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \\ 7 \end{bmatrix} = \begin{bmatrix} 16 \\ 9 \end{bmatrix}.$$

Resolvendo o sistema $A^TA\mathbf{x} = A^T\mathbf{b}$, obtemos a solução

$$c = \frac{7}{2} \quad \text{e} \quad d = -\frac{1}{2}.$$

Portanto, a única solução de quadrados mínimos é

$$\mathbf{x} = \begin{bmatrix} \frac{7}{2} \\ \\ -\frac{1}{2} \end{bmatrix}.$$

Logo, a função linear que melhor se ajusta aos dados é

$$y = \frac{7}{2}x - \frac{1}{2} \ .$$

Como a solução é aproximada, existe um vetor erro **e**.

Consideremos as abcissas dos pontos dados no problema, e calculemos as ordenadas correspondentes desses pontos na reta obtida.

$$\begin{array}{l} para \;\; x = 0, \;\; y = -\frac{1}{2} \\ para \;\; x = 1, \;\; y = 3 \\ para \;\; x = 2, \;\; y = \frac{13}{2} \end{array} \qquad \text{, ou seja,}$$

os pontos $Q_1 = \left(0, -\frac{1}{2}\right)$, $Q_2 = (1, 3)$, $Q_3 = \left(2, \frac{13}{2}\right)$ pertencem à reta.

O vetor

$$\mathbf{p} = \begin{bmatrix} -\frac{1}{2} \\ 3 \\ \frac{13}{2} \end{bmatrix} \quad \text{pertence à } Col(A),$$

e sabemos que $\mathbf{p} = proj_W \mathbf{b} = A\mathbf{x}$, $\mathbf{x}$ obtido acima.

Assim como, a reta de quadrados mínimos é a reta que melhor se ajusta aos dados do experimento, $\mathbf{p}$ é o vetor pertencente à $Col(A)$ mais próximo do vetor $\mathbf{b}$.

Subtraindo $\mathbf{p}$ de $\mathbf{b}$, obtemos o vetor erro:

$$\begin{bmatrix} 0 \\ 2 \\ 7 \end{bmatrix} - \begin{bmatrix} -\frac{1}{2} \\ 3 \\ \frac{13}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ -1 \\ \frac{1}{2} \end{bmatrix} = \mathbf{e}$$

$$\|\mathbf{e}\|^2 = \left(\frac{1}{2}\right)^2 + (-1)^2 + \left(\frac{1}{2}\right)^2 = \frac{3}{2} \quad \text{é mínima.}$$

Sabemos que $\mathbf{e} = \mathbf{b} - A\mathbf{x}$ é ortogonal ao subespaço $Col(A)$, ou seja, $\mathbf{e}$ pertence à $N(A^T)$.

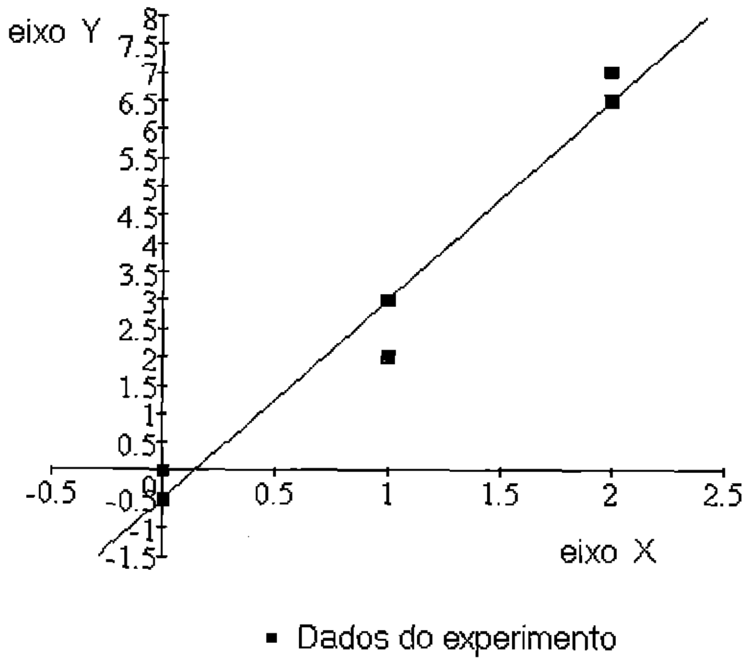


Gráfico 2.2

□

Algumas vezes não é conveniente ajustar os dados a uma reta, pois em certos experimentos já esperamos que a relação entre x e y seja dada por alguma outra função que não é a linear.

Deste modo, devemos encontrar um polinômio de grau estipulado que "melhor se ajuste" aos dados.

Suponha então, que temos m pontos representando os dados:

$$(x_1, y_1), (x_2, y_2), ..., (x_m, y_m),$$

e queremos obter um polinômio de grau p, que "melhor se ajuste" a eles.

Seja

$$y_i = c_p x_i^p + c_{p-1} x_i^{p-1} + \ldots + c_1 x_i + c_0 + e_i \quad , \qquad i = 1, 2, \ldots, m,$$

tal que $p \leqslant m - 1$.

Sejam

$$\mathbf{b} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}_{m \times 1} \quad , \quad A = \begin{bmatrix} x_1^p & \cdots & x_1 & 1 \\ x_2^p & \cdots & x_2 & 1 \\ \vdots & \ddots & \vdots & \vdots \\ x_m^p & \cdots & x_m & 1 \end{bmatrix}_{m \times (p+1)} \quad , \quad \mathbf{x} = \begin{bmatrix} c_p \\ \vdots \\ c_1 \\ c_0 \end{bmatrix}_{(p+1) \times 1}$$

$$\text{e} \qquad \mathrm{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{bmatrix}_{m \times 1} .$$

Portanto, podemos escrever as m equações na forma matricial como

$$\mathbf{b} = A\mathbf{x} + \mathrm{e} \ ,$$

ou seja,

$$\mathrm{e} = \mathbf{b} - A\mathbf{x}.$$

Exemplo 2.3. CURVA DE QUADRADOS MÍNIMOS

Em uma floresta tropical, foram obtidos os seguintes dados: $\mathbf{x}$ é o número de presas e $\mathbf{y}$ é o número de predadores (por quilometro quadrado) durante um certo tempo:

$$x = 2, y = 2 \quad ; \quad x = 4, y = 2 \quad ; \quad x = 3, y = 2 \quad ; \quad x = 5, y = 1.$$

Sejam $P_1 = (2,2)$, $P_2 = (4,2)$, $P_3 = (3,2)$, $P_4 = (5,1)$

Com os quatro pontos obtidos através das medições, poderíamos no máximo ajustar nossos dados a uma curva de terceiro grau. Porém nossa expectativa para este experimento, é que x e y tenham uma relação quadrática.

Portanto, nosso objetivo é achar uma curva

$$y = c_1x^2 + c_2x + d,$$

que melhor se ajuste aos dados do experimento.
Se nossos dados não tivessem erros, teríamos o seguinte sistema,

$$\begin{aligned} 2 &= 4c_1 + 2c_2 + d \\ 2 &= 16c_1 + 4c_2 + d \\ 2 &= 9c_1 + 3c_2 + d \\ 1 &= 25c_1 + 5c_2 + d \end{aligned},$$

ou seja,

$$\begin{bmatrix} 2 \\ 2 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 & 2 & 1 \\ 16 & 4 & 1 \\ 9 & 3 & 1 \\ 25 & 5 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ d \end{bmatrix}.$$

Sejam $\quad \mathbf{b} = \begin{bmatrix} 2 \\ 2 \\ 2 \\ 1 \end{bmatrix}, \quad A = \begin{bmatrix} 4 & 2 & 1 \\ 16 & 4 & 1 \\ 9 & 3 & 1 \\ 25 & 5 & 1 \end{bmatrix} \quad$ e $\quad \mathbf{x} = \begin{bmatrix} c_1 \\ c_2 \\ d \end{bmatrix}.$

Como o sistema é Incompatível, usaremos o método de quadrados mínimos.
As colunas de A são linearmente independentes, ou seja, $posto(A) = n = 3$.
Assim, a solução de quadrados mínimos é única.
Deste modo

$$A^TA = \begin{bmatrix} 4 & 16 & 9 & 25 \\ 2 & 4 & 3 & 5 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 4 & 2 & 1 \\ 16 & 4 & 1 \\ 9 & 3 & 1 \\ 25 & 5 & 1 \end{bmatrix} = \begin{bmatrix} 978 & 224 & 54 \\ 224 & 54 & 14 \\ 54 & 14 & 4 \end{bmatrix}.$$

e

$$A^T\mathbf{b} = \begin{bmatrix} 4 & 16 & 9 & 25 \\ 2 & 4 & 3 & 5 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 83 \\ 23 \\ 7 \end{bmatrix}.$$

Resolvendo o sistema $A^T A\mathbf{x} = A^T\mathbf{b}$, obtemos a solução

$$c_1 = -\frac{1}{4} \quad , \quad c_2 = \frac{29}{20} \quad \text{e} \qquad d = \frac{1}{20}.$$

Portanto, existe uma única solução de quadrados mínimos que é

$$\mathbf{x} = \begin{bmatrix} -\frac{1}{4} \\ \\ \frac{29}{20} \\ \\ \frac{1}{20} \end{bmatrix} .$$

Logo, a função de grau 2 que melhor se ajusta aos dados é

$$y = -\frac{1}{4}x^2 + \frac{29}{20}x + \frac{1}{20}.$$

Consideremos as abcissas dos pontos dados no problema, e calculemos as ordenadas correspondentes desses pontos na parábola obtida.

$$para \quad x = 2, \quad y = \frac{39}{20}$$

$$para \quad x = 4, \quad y = \frac{37}{20}$$

$$para \quad x = 3, \quad y = \frac{43}{20}$$

$$para \quad x = 5, \quad y = \frac{21}{20} \ .$$

Portanto, os pontos $Q_1 = \left(2, \frac{39}{20}\right)$, $Q_2 = \left(4, \frac{37}{20}\right)$, $Q_3 = \left(3, \frac{43}{20}\right)$ e $Q_4 = (5, \frac{21}{20})$ pertencem à curva de quadrados mínimos.

O vetor

$$\mathbf{p} = \begin{bmatrix} \frac{39}{20} \\ \frac{37}{20} \\ \frac{43}{20} \\ \frac{21}{20} \end{bmatrix} \qquad \text{pertence à } Col(A),$$

e sabemos que $\mathbf{p} = proj_W\mathbf{b} = A\mathbf{x}$, $\mathbf{x}$ obtido acima.

Assim como, a curva de quadrados mínimos é a curva que melhor se ajusta aos dados do experimento, $\mathbf{p}$ é o vetor pertencente à $Col(A)$ mais próximo do vetor $\mathbf{b}$.

Subtraindo $\mathbf{p}$ de $\mathbf{b}$ obtemos o vetor erro:

$$\begin{bmatrix} 2 \\ 2 \\ 2 \\ 1 \end{bmatrix} - \begin{bmatrix} \frac{39}{20} \\ \frac{37}{20} \\ \frac{43}{20} \\ \frac{21}{20} \end{bmatrix} = \begin{bmatrix} \frac{1}{20} \\ \frac{3}{20} \\ -\frac{3}{20} \\ -\frac{1}{20} \end{bmatrix} = \mathbf{e}$$

$$\|\mathbf{e}\|^2 = \left(\frac{1}{20}\right)^2 + \left(\frac{3}{20}\right)^2 + \left(-\frac{3}{20}\right)^2 + \left(-\frac{1}{20}\right)^2 = \frac{1}{20} \quad \text{é mínima.}$$

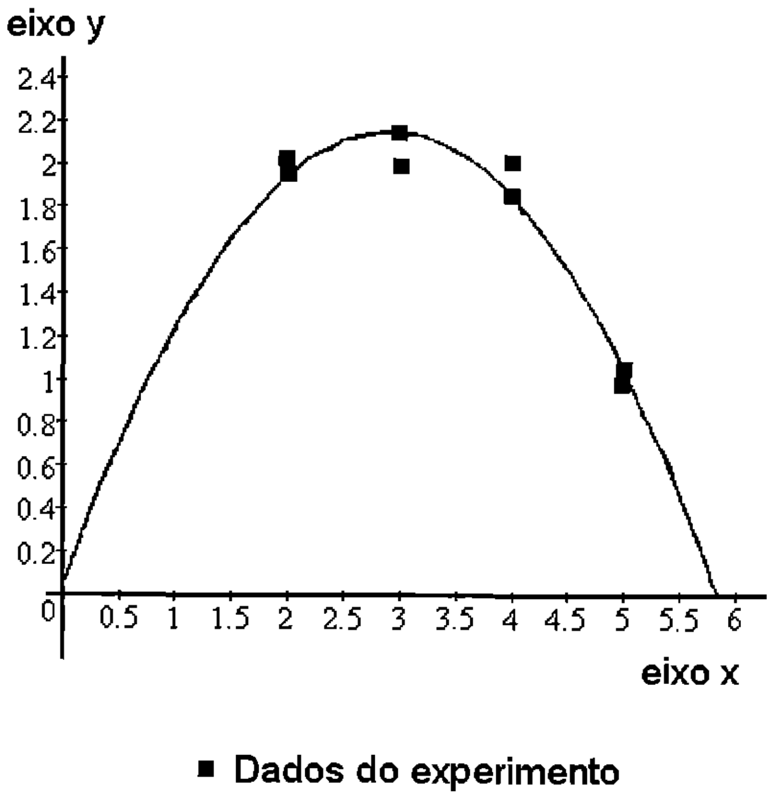

eixo y
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
eixo x
Dados do experimento

Gráfico 2.3

□

2.3. DECOMPOSIÇÃO QR

Discutiremos abaixo um tipo de decomposição de matrizes, que será utilizado na obtenção de soluções de problemas de quadrados mínimos.

Teorema 2.3. Seja A uma matriz $m \times n$ cujas colunas são vetores linearmente independentes. Então A pode ser fatorada como

$$A = QR,$$

sendo Q uma matriz $m \times n$ cujas colunas formam um conjunto ortonormal, e R é uma matriz triangular superior inversível $n \times n$.

Demonstração.

$$\text{Seja} \quad A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \in \mathbb{R}^{m \times n}$$

Denotaremos por $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n$ as colunas de A. Como elas são linearmente independentes, elas formam uma base para o espaço coluna de A.

Pode ser utilizado o processo de Gram-Schmidt para obtermos um conjunto $\{\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_n\}$ de vetores ortonormais a partir das colunas de A.

Seja Q a matriz cujas colunas são $\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_n$, isto é,

$$Q = [\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_n].$$

As colunas $\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_n$ de Q formam um base ortonormal para o espaço coluna de A.

De acordo com o *teorema 1.31*, toda coluna $\mathbf{a}_i$, $i = 1, 2, ..., n$, é combinação linear dos vetores da base $\{\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_n\}$:

$$\begin{aligned} \mathbf{a}_1 &= \langle \mathbf{a}_1, \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{a}_1, \mathbf{q}_2 \rangle \mathbf{q}_2 + ... + \langle \mathbf{a}_1, \mathbf{q}_n \rangle \mathbf{q}_n \\ \mathbf{a}_2 &= \langle \mathbf{a}_2, \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{a}_2, \mathbf{q}_2 \rangle \mathbf{q}_2 + ... + \langle \mathbf{a}_2, \mathbf{q}_n \rangle \mathbf{q}_n \\ &\vdots \\ \mathbf{a}_n &= \langle \mathbf{a}_n, \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{a}_n, \mathbf{q}_2 \rangle \mathbf{q}_2 + ... + \langle \mathbf{a}_n, \mathbf{q}_n \rangle \mathbf{q}_n. \end{aligned} \tag{2.6}$$

Para que possamos representar este sistema matricialmente, terá que existir uma terceira matriz $R \in \mathbb{R}^{n\times n}$ para conectar as matrizes A e Q. Assim sendo tomamos:

$$R = \begin{bmatrix} \langle \mathbf{a}_1, \mathbf{q}_1 \rangle & \langle \mathbf{a}_2, \mathbf{q}_1 \rangle & \cdots & \langle \mathbf{a}_n, \mathbf{q}_1 \rangle \\ \langle \mathbf{a}_1, \mathbf{q}_2 \rangle & \langle \mathbf{a}_2, \mathbf{q}_2 \rangle & \cdots & \langle \mathbf{a}_n, \mathbf{q}_2 \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{a}_1, \mathbf{q}_n \rangle & \langle \mathbf{a}_2, \mathbf{q}_n \rangle & \cdots & \langle \mathbf{a}_n, \mathbf{q}_n \rangle \end{bmatrix}$$

Pela multiplicação de matrizes vista no *teorema 1.2*,

$$QR = [Q\mathbf{r}_1 \;\; Q\mathbf{r}_2 \;\; \ldots \;\; Q\mathbf{r}_n],$$

sendo $\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_n$ as colunas da matriz R.

Portanto,

$$Q\mathbf{r}_i = \langle \mathbf{a}_i, \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{a}_i, \mathbf{q}_2 \rangle \mathbf{q}_2 + ... + \langle \mathbf{a}_i, \mathbf{q}_n \rangle \mathbf{q}_n, \quad i = 1, 2, ..., n.$$

Assim, de (2.6),

$$Q\mathbf{r}_i = \mathbf{a}_i.$$

Logo,

$$QR = A.$$

Porém, R é uma matriz triangular superior, pois do processo de Gram-Schmidt, temos que $\mathbf{q}_n$ é ortogonal aos vetores $\mathbf{a}_1$, $\mathbf{a}_2$, ..., $\mathbf{a}_{n-1}$, se $n \geq 2$. Logo:

$$R = \begin{bmatrix} \langle \mathbf{a}_1, \mathbf{q}_1 \rangle & \langle \mathbf{a}_2, \mathbf{q}_1 \rangle & \cdots & \langle \mathbf{a}_n, \mathbf{q}_1 \rangle \\ 0 & \langle \mathbf{a}_2, \mathbf{q}_2 \rangle & \cdots & \langle \mathbf{a}_n, \mathbf{q}_2 \rangle \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \langle \mathbf{a}_n, \mathbf{q}_n \rangle \end{bmatrix}$$

Portanto,

$$A = QR,$$

sendo Q uma matriz $m \times n$ com colunas ortonormais e R uma matriz $n \times n$ triangular superior.

Vamos mostrar agora que R é inversível.

Seja $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ uma solução do sistema $R\mathbf{x} = \mathbf{O}$.

Segue-se que $QR\mathbf{x} = Q\mathbf{O}$, ou seja, $A\mathbf{x} = \mathbf{O}$.

Mas, $A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + ... + x_n\mathbf{a}_n$.

Então,

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + ... + x_n\mathbf{a}_n = \mathbf{O}.$$

Como as colunas $\mathbf{a}_i$, $i = 1, ..., n$, são linearmente independentes, então

$$x_1 = x_2 = ... = x_n = 0.$$

Logo, $\mathbf{x} = \mathbf{O}$.

Como $\mathbf{x} = \mathbf{O}$ é a única solução do sistema homogêneo $R\mathbf{x} = \mathbf{O}$, então, pelo *teorema 1.29*, R é inversível.

□

Exemplo 2.4. Encontrar a fatoração QR da matriz A, sendo Q uma matriz 4×3 com colunas ortonormais, e R triangular superior de ordem 3×3 e inversível.

Seja $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}_{4\times 3}$

Sejam $\mathbf{a}_1 = (1,1,1,1)^T$, $\mathbf{a}_2 = (0,1,1,1)^T$ e $\mathbf{a}_3 = (0,0,1,1)^T$, os quais são linearmente independentes e não ortogonais entre si.

Como $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ geram o espaço coluna de A e são linearmente independentes, então $B = \{\mathbf{a}_1,\ \mathbf{a}_2,\ \mathbf{a}_3\}$ é uma base para $Col(A)$.

Através do processo de Gram-Schmidt calcularemos primeiramente os vetores ortogonais $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, a partir de $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$.

Sejam

$$\mathbf{v}_1 = \mathbf{a}_1 = (1,1,1,1)^T$$

$$\mathbf{v}_2 = \mathbf{a}_2 - \frac{\langle \mathbf{a}_2, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 = \left(-\frac{3}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)^T$$

$$\mathbf{v}_3 = \mathbf{a}_3 - \frac{\langle \mathbf{a}_3, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 - \frac{\langle \mathbf{a}_3, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2 = \left(0, -\frac{2}{3}, \frac{1}{3}, \frac{1}{3}\right)^T$$

$\overline{B} = \left\{(1,1,1,1)^T, \left(-\frac{3}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)^T, \left(0, -\frac{2}{3}, \frac{1}{3}, \frac{1}{3}\right)^T\right\}$ é uma base ortogonal do espaço coluna de A.

Sejam

$$\mathbf{q}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)^T$$

$$\mathbf{q}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \left(-\frac{3}{\sqrt{12}}, \frac{1}{\sqrt{12}}, \frac{1}{\sqrt{12}}, \frac{1}{\sqrt{12}}\right)^T$$

$$\mathbf{q}_3 = \frac{\mathbf{v}_3}{\|\mathbf{v}_3\|} = \left(0, -\frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}\right)^T$$

$\overline{\overline{B}} = \{\mathbf{q}_1,\ \mathbf{q}_2,\ \mathbf{q}_3\}$ é uma base ortonormal que gera o mesmo espaço gerado pela base $B = \{\mathbf{a}_1,\ \mathbf{a}_2,\ \mathbf{a}_3\}$.

Deste modo obtemos:

$$Q = [\mathbf{q}_1,\ \mathbf{q}_2,\ \mathbf{q}_3] = \begin{bmatrix} \frac{1}{2} & -\frac{3}{\sqrt{12}} & 0 \\ \frac{1}{2} & \frac{1}{\sqrt{12}} & \frac{2}{\sqrt{6}} \\ \frac{1}{2} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{6}} \\ \frac{1}{2} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{6}} \end{bmatrix}_{4\times 3},$$

$$R = \begin{bmatrix} \langle \mathbf{a}_1, \mathbf{q}_1\rangle & \langle \mathbf{a}_2, \mathbf{q}_1\rangle & \langle \mathbf{a}_3, \mathbf{q}_1\rangle \\ 0 & \langle \mathbf{a}_2, \mathbf{q}_2\rangle & \langle \mathbf{a}_3, \mathbf{q}_2\rangle \\ 0 & 0 & \langle \mathbf{a}_3, \mathbf{q}_3\rangle \end{bmatrix} = \begin{bmatrix} 2 & \frac{3}{2} & 1 \\ 0 & \frac{3}{\sqrt{12}} & \frac{3}{\sqrt{12}} \\ 0 & 0 & \frac{2}{\sqrt{6}} \end{bmatrix}_{3\times 3}$$

$$\text{e} \qquad QR = \begin{bmatrix} \frac{1}{2} & -\frac{3}{\sqrt{12}} & 0 \\ \frac{1}{2} & \frac{1}{\sqrt{12}} & \frac{2}{\sqrt{6}} \\ \frac{1}{2} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{6}} \\ \frac{1}{2} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{6}} \end{bmatrix} \times \begin{bmatrix} 2 & \frac{3}{2} & 1 \\ 0 & \frac{3}{\sqrt{12}} & \frac{3}{\sqrt{12}} \\ 0 & 0 & \frac{2}{\sqrt{6}} \end{bmatrix} = A$$

$\square$

Observação: A matriz Q pode ser calculada de outra forma, como por exemplo pelo método de Gram-Schmidt modificado, o qual nos dá melhor precisão numérica no caso de matrizes de grande porte.

2.4. QUADRADOS MÍNIMOS E DECOMPOSIÇÃO QR

Seja A uma matriz $m \times n$ e $A\mathbf{x} = \mathbf{b}$ um sistema linear.

Seja $Posto(A) = n$.

Então, do *teorema 2.2*, sabemos que existe uma única solução de quadrados mínimos:

$$\widehat{\mathbf{x}} = \left(A^T A\right)^{-1} A^T \mathbf{b}.$$

Como vimos, essa solução pode ser obtida resolvendo o sistema linear

$$A^T A\mathbf{x} = A^T \mathbf{b},$$

utilizando o método de Gauss (escalonando a matriz aumentada do sistema).

Outra forma de resolver o sistema normal acima é obtido utilizando a decomposição QR da matriz A.

Seja

$$A = QR, \qquad Q : m \times n \qquad \text{e} \qquad R : n \times n, \text{ inversível.}$$

Como

$$Q^T Q = I \ , \qquad I : n \times n$$

temos

$$A^T A = (QR)^T QR = R^T Q^T QR = R^T R.$$

Como $A^T A = R^T R$, segue-se que

$$\widehat{\mathbf{x}} = \left(A^T A\right)^{-1} A^T \mathbf{b} = \left(R^T R\right)^{-1} (QR)^T \mathbf{b} = R^{-1} \left(R^T\right)^{-1} R^T Q^T \mathbf{b}.$$

Logo,

$$\widehat{\mathbf{x}} = R^{-1} Q^T \mathbf{b}.$$

A solução $\widehat{\mathbf{x}}$ pode ser obtida resolvendo o *sistema triangular*

$$R\mathbf{x} = Q^T\mathbf{b}.$$

$\square$

Exemplo 2.5. SOLUÇÃO DE QUADRADOS MÍNIMOS E DECOMPOSIÇÃO QR

Sejam A e $\mathbf{b}$ as mesmas matrizes do *exemplo 2.1*:

$$A = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 1 & -2 \\ 1 & 1 & 0 \\ 1 & 1 & -1 \end{bmatrix}_{4\times 3}, \quad \mathbf{b} = \begin{bmatrix} 6 \\ 0 \\ 9 \\ 3 \end{bmatrix}_{4\times 1}.$$

Sabemos, pelo *exemplo 2.1*, que a solução é única.

Considere $\mathbf{a}_1 = (1,2,1,1)^T$; $\mathbf{a}_2 = (0,1,1,1)^T$ e $\mathbf{a}_3 = (-1,-2,0,-1)^T$, que são linearmente independentes e não ortogonais entre si.

Como $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ geram o espaço coluna de A e são linearmente independentes, então $B = \{\mathbf{a}_1,\ \mathbf{a}_2,\ \mathbf{a}_3\}$ é uma base para $Col(A)$.

Através do processo de Gram-Schmidt, calcularemos primeiramente os vetores ortogonais $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, a partir de $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$.

Sejam

$$\mathbf{v}_1 = \mathbf{a}_1 = (1,2,1,1)^T$$

$$\mathbf{v}_2 = \mathbf{a}_2 - \frac{\langle \mathbf{a}_2, \mathbf{v}_1\rangle}{\langle \mathbf{v}_1, \mathbf{v}_1\rangle}\mathbf{v}_1 = \left(-\frac{4}{7}, -\frac{1}{7}, \frac{3}{7}, \frac{3}{7}\right)^T$$

$$\mathbf{v}_3 = \mathbf{a}_3 - \frac{\langle \mathbf{a}_3, \mathbf{v}_1\rangle}{\langle \mathbf{v}_1, \mathbf{v}_1\rangle}\mathbf{v}_1 - \frac{\langle \mathbf{a}_3, \mathbf{v}_2\rangle}{\langle \mathbf{v}_2, \mathbf{v}_2\rangle}\mathbf{v}_2 = \left(\frac{1}{5}, -\frac{1}{5}, \frac{3}{5}, -\frac{2}{5}\right)^T$$

$\overline{B} = \{\mathbf{v}_1,\ \mathbf{v}_2,\ \mathbf{v}_3\}$ é uma base ortogonal do espaço coluna de A..

Sejam

$$\mathbf{q}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \frac{\sqrt{7}}{7}(1,2,1,1)^T$$

$$\mathbf{q}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \frac{\sqrt{35}}{5}\left(-\frac{4}{7},-\frac{1}{7},\frac{3}{7},\frac{3}{7}\right)^T$$

$$\mathbf{q}_3 = \frac{\mathbf{v}_3}{\|\mathbf{v}_3\|} = \frac{\sqrt{15}}{3}\left(\frac{1}{5},-\frac{1}{5},\frac{3}{5},-\frac{2}{5}\right)^T$$

$\overline{\overline{B}} = \{\mathbf{q}_1,\ \mathbf{q}_2,\ \mathbf{q}_3\}$ é uma base ortonormal que gera o mesmo espaço gerado pela base $B = \{\mathbf{a}_1,\ \mathbf{a}_2,\ \mathbf{a}_3\}$.

Deste modo obtemos:

$$Q = \begin{bmatrix} \frac{1}{7}\sqrt{7} & -\frac{4}{35}\sqrt{35} & \frac{1}{15}\sqrt{15} \\ \frac{2}{7}\sqrt{7} & -\frac{1}{35}\sqrt{35} & -\frac{1}{15}\sqrt{15} \\ \frac{1}{7}\sqrt{7} & \frac{3}{35}\sqrt{35} & \frac{1}{5}\sqrt{15} \\ \frac{1}{7}\sqrt{7} & \frac{3}{35}\sqrt{35} & -\frac{2}{15}\sqrt{15} \end{bmatrix}$$

e

$$Q^T = \begin{bmatrix} \frac{1}{7}\sqrt{7} & \frac{2}{7}\sqrt{7} & \frac{1}{7}\sqrt{7} & \frac{1}{7}\sqrt{7} \\ -\frac{4}{35}\sqrt{35} & -\frac{1}{35}\sqrt{35} & \frac{3}{35}\sqrt{35} & \frac{3}{35}\sqrt{35} \\ \frac{1}{15}\sqrt{15} & -\frac{1}{15}\sqrt{15} & \frac{1}{5}\sqrt{15} & -\frac{2}{15}\sqrt{15} \end{bmatrix}$$

E ainda,

$$R = \begin{bmatrix} \sqrt{7} & \frac{4}{7}\sqrt{7} & -\frac{6}{7}\sqrt{7} \\ 0 & \frac{1}{7}\sqrt{35} & \frac{3}{35}\sqrt{35} \\ 0 & 0 & \frac{1}{5}\sqrt{15} \end{bmatrix}.$$

Resolvendo o sistema $R\mathbf{x} = Q^T b$:

$$\begin{bmatrix} \sqrt{7} & \frac{4}{7}\sqrt{7} & -\frac{6}{7}\sqrt{7} \\ 0 & \frac{1}{7}\sqrt{35} & \frac{3}{35}\sqrt{35} \\ 0 & 0 & \frac{1}{5}\sqrt{15} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{18}{7}\sqrt{7} \\ \frac{12}{35}\sqrt{35} \\ \frac{9}{5}\sqrt{15} \end{bmatrix},$$

obtemos

$$x_1 = 12, \quad x_2 = -3, \quad x_3 = 9.$$

Logo,

$$\mathbf{x} = \begin{bmatrix} 12 \\ -3 \\ 9 \end{bmatrix} \quad \text{é a solução única de quadrados mínimos.}$$

$\square$

2.5. DECOMPOSIÇÃO EM VALORES SINGULARES

Sabemos, do *teorema 1.39*, que quando A é uma matriz $n \times n$ simétrica, é possível obtermos a seguinte decomposição de A:

$$A = QDQ^T,$$

sendo Q uma matriz ortogonal e D matriz diagonal.

Os n autovalores reais de A formam a diagonal de D e as colunas de Q são constituídas por autovetores que formam uma base ortonormal de $\mathbb{R}^n$.

Será apresentada agora uma decomposição geral para matrizes $m \times n$, chamada de **decomposição em valores singulares.**

Seja A uma matriz $m \times n$; então

$$A = U\Sigma V^T,$$

U e V são matrizes ortogonais, sendo U uma matriz $m \times m$ e V uma matriz $n \times n$ e Σ como dada abaixo.

A matriz Σ tem a seguinte forma:

$$\Sigma = \left[\begin{array}{cccc|c} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & 0 \\ & & & \sigma_r & \\ \hline & & 0 & & 0 \end{array}\right]_{m\times n} \quad , \text{com } \sigma_1 \geqslant \sigma_2 \geqslant \ldots \geqslant \sigma_r > 0,$$

sendo r o posto de A.

Teorema 2.4. DECOMPOSIÇÃO EM VALORES SINGULARES

Seja A uma matriz $m \times n$ tal que $posto(A) = r$. Então existem números reais $\sigma_1 \geqslant \sigma_2 \geqslant \ldots. \geqslant \sigma_r > 0$, $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ base ortonormal de $\mathbb{R}^n$ e $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ base ortonormal de $\mathbb{R}^m$, tais que,

$$\begin{array}{ll} A\mathbf{v}_i = \sigma_i \mathbf{u}_i, \quad i = 1, 2, \ldots, r & \qquad A\mathbf{v}_i = \mathbf{O}, \quad i = r+1, \ldots, n \\ A^T\mathbf{u}_i = \sigma_i \mathbf{v}_i, \quad i = 1, 2, \ldots, r & \qquad A^T\mathbf{u}_i = \mathbf{O}, \quad i = r+1, \ldots, m \end{array}$$

Os vetores $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ são autovetores de A^TA e $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_m$ são autovetores de AA^T.

E ainda, σ_1^2, $\sigma_2^2, ...,$ σ_r^2 são os autovalores não nulos de A^TA e AA^T.

Demonstração.

Seja A uma matriz $m \times n$.

Como A^TA é uma matriz simétrica $n \times n$, sabemos pelo *teorema 1.39*, que existe uma base ortonormal de $\mathbb{R}^n$, formada por autovetores de A^TA.

Seja $\{\mathbf{v}_1,\ \mathbf{v}_2, ...,\ \mathbf{v}_n\}$ essa base ortonormal de autovetores, e sejam $\lambda_1,\ \lambda_2, ...,$ λ_n os n autovalores reais associados.

Os autovalores de A^TA são não negativos, pois para todo $i = 1, ..., n$ temos:

$$\begin{aligned} \|A\mathbf{v}_i\|^2 &= \langle A\mathbf{v}_i, A\mathbf{v}_i\rangle = \left(\mathbf{v}_i^T A^T\right) A\mathbf{v}_i = \mathbf{v}_i^T \left(A^TA\right)\mathbf{v}_i = \mathbf{v}_i^T\lambda_i\mathbf{v}_i = \\ &= \lambda_i\mathbf{v}_i^T\mathbf{v}_i = \lambda_i\langle\mathbf{v}_i,\mathbf{v}_i\rangle = \lambda_i\|\mathbf{v}_i\|^2 = \lambda_i. \end{aligned}$$

Como $\|A\mathbf{v}_i\|^2 \geqslant 0$, então $\lambda_i \geqslant 0$, $\forall\ i = 1, ..., n$

Assuma $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ ordenados de modo que $\lambda_1 \geqslant \lambda_2 \geqslant ... \geqslant \lambda_n$.

Como, $posto(A) = r$ e $posto(A) = posto(A^TA)$, temos que

$$posto(A^TA) = r.$$

Como $dim\left(N(A^TA)\right) = n - r$, então existem r autovetores que não estão associados ao autovalor nulo.

Portanto,

$$\lambda_1 \geqslant \lambda_2 \geqslant ... \geqslant \lambda_r > 0 \ \text{ e } \ \lambda_{r+1} = ... = \lambda_n = 0\ .$$

Para $i = 1, ..., r$, definimos σ_i e $\mathbf{u}_i$ como:

$$\sigma_i = \sqrt{\lambda_i} \qquad \text{e} \qquad \mathbf{u}_i = \frac{1}{\sigma_i}A\mathbf{v}_i.$$

Logo,

$$A\mathbf{v}_i = \sigma_i\mathbf{u}_i, \quad i = 1, 2, ..., r \qquad \text{e} \qquad A\mathbf{v}_i = \mathbf{O}, \quad i = r + 1, ..., n.$$

Sejam $1 \leqslant i, j \leqslant r$; então,

$$\langle\mathbf{u}_i, \mathbf{u}_j\rangle = \mathbf{u}_i^T\mathbf{u}_j = \frac{\mathbf{v}_i^T A^T A\mathbf{v}_j}{\sigma_i\sigma_j} = \frac{\mathbf{v}_i^T\lambda_j\mathbf{v}_j}{\sigma_i\sigma_j} = \lambda_j\frac{\langle\mathbf{v}_i, \mathbf{v}_j\rangle}{\sigma_i\sigma_j} = 0\ ,\ \ i \neq j$$

e

$$\|\mathbf{u}_i\|^2 = \left\|\frac{A\mathbf{v}_i}{\sigma_i}\right\|^2 = \frac{\|A\mathbf{v}_i\|^2}{\|\sigma_i\|^2} = \frac{\lambda_i}{\sigma_i^2} = \frac{\lambda_i}{\lambda_i} = 1 \quad , \quad i = 1, ..., r.$$

Então, temos que $\mathbf{u}_1$, $\mathbf{u}_2, ...,$ $\mathbf{u}_r$ formam um conjunto ortonormal.
Multiplicando A^T por $\mathbf{u}_i$ obtemos:

$$A^T\mathbf{u}_i = \left(\frac{1}{\sigma_i}\right) A^T A\mathbf{v}_i = \left(\frac{\lambda_i}{\sigma_i}\right) \mathbf{v}_i = \frac{\sigma_i^2}{\sigma_i}\mathbf{v}_i = \sigma_i\mathbf{v}_i.$$

Logo,

$$A^T\mathbf{u}_i = \sigma_i\mathbf{v}_i, \quad i = 1, 2, ..., r.$$

Sabemos que $A^T A\mathbf{v}_i = \lambda_i\mathbf{v}_i$, portanto, pelo *teorema 1.40*, temos que

$$AA^T(A\mathbf{v}_i) = \lambda_i(A\mathbf{v}_i).$$

Como $A\mathbf{v}_i = \sigma_i\mathbf{u}_i$, obtemos

$$A^T A\sigma_i\mathbf{u}_i = \lambda_i\sigma_i\mathbf{u}_i, \text{ ou seja, } A^T A\mathbf{u}_i = \frac{1}{\sigma_i}\lambda_i\sigma_i\mathbf{u}_i.$$

Logo,

$$A^T A\mathbf{u}_i = \lambda_i\mathbf{u}_i, \; i = 1, ..., r,$$

isto é, os vetores $\mathbf{u}_1$, $\mathbf{u}_2, ...,$ $\mathbf{u}_r$ são autovetores de AA^T associados aos autovalores não nulos.
Vamos, agora estender $\{\mathbf{u}_1, \; \mathbf{u}_2, ..., \; \mathbf{u}_r\}$ a uma base ortonormal

$$\{\mathbf{u}_1, \; \mathbf{u}_2, ..., \; \mathbf{u}_m\} \text{ de } \mathbb{R}^m.$$

Já que $AA^T \in \mathbb{R}^{m\times m}$ e $posto(AA^T) = r$ temos que,

$$dim(N(AA^T)) = m - r.$$

Seja $\{\mathbf{u}_{r+1}, ..., \mathbf{u}_m\}$ uma base ortonormal de $N(AA^T)$.

Então $AA^T\mathbf{u}_i = \mathbf{O}$, $i = r+1, ..., m$, ou seja, $\mathbf{u}_{r+1}, ..., \mathbf{u}_m$ são autovetores de AA^T associados ao autovalor nulo.

Temos que $\mathbf{u}_{r+1}, ..., \mathbf{u}_m$ são ortogonais a $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r$.

Assim, $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_m$ é uma base ortonormal de $\mathbb{R}^m$ constituída de autovetores de AA^T.

Como $N(AA^T) = N(A^T)$,temos que

$$A^T\mathbf{u}_i = \mathbf{O}, \quad i = r+1, ..., m.$$

□

Definição 2.5. VALORES SINGULARES DE A

Os números $\sigma_1, \sigma_2, ..., \sigma_r$ são chamados de **valores singulares** de A.

Notação: $\sigma_i = \sqrt{\lambda_i}$, $i = 1, ..., r$.

Definição 2.6. VETORES SINGULARES DE A

Os vetores $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ são chamados vetores singulares à direita de A e $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_m$ são chamados vetores singulares à esquerda de A.

Observação:

1) A^T tem os mesmos valores singulares que A, e os vetores singulares à direita (à esquerda) de A^T são os vetores singulares à esquerda (à direita) de A.

2) Seja $posto(A) = r$. Então:

$\mathbf{v}_1, ..., \mathbf{v}_r$ formam uma base ortonormal para $Col(A^T)$.
$\mathbf{v}_{r+1}, ..., \mathbf{v}_n$ formam uma base ortonormal para $N(A)$.
$\mathbf{u}_1, ..., \mathbf{u}_r$ formam uma base ortonormal para $Col(A)$.
$\mathbf{u}_{r+1}, ..., \mathbf{u}_m$ formam uma base ortonormal para $N(A^T)$.

3) A decomposição em valores singulares de uma matriz A não é única, pois U e V não são matrizes únicas. Isto ocorre porque as colunas de V formam um conjunto ortonormal de autovetores de A^TA, sendo que esses autovetores não são únicos. Como as colunas de U dependem das colunas de V, temos que U também não é uma matriz única. Já a matriz Σ é única, pois existem valores singulares únicos para A^TA.

Exemplo.2.6. EXEMPLO DE DECOMPOSIÇÃO EM VALORES SINGULARES

Seja $$A = \begin{bmatrix} 2 & -2 \\ 1 & 1 \\ 2 & 2 \end{bmatrix}_{3\times 2}.$$

Como as colunas de A são linearmente independentes, temos que $posto(A) = 2$.

Cálculo dos valores singulares de A.

$$A^TA = \begin{bmatrix} 2 & 1 & 2 \\ -2 & 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & -2 \\ 1 & 1 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 9 & 1 \\ 1 & 9 \end{bmatrix} : \text{ matriz simétrica.}$$

Os autovalores são as raízes da equação $det(A^TA - \lambda I) = 0$, isto é, são as raízes de

$$det \begin{bmatrix} 9-\lambda & 1 \\ 1 & 9-\lambda \end{bmatrix} = 0$$

Resolvendo a equação, obtemos as raízes

$$\lambda_1 = 10 \quad \text{e} \quad \lambda_2 = 8.$$

Logo, os valores singulares de A são

$$\sigma_1 = \sqrt{\lambda_1} = \sqrt{10} \qquad \text{e} \qquad \sigma_2 = \sqrt{\lambda_2} = \sqrt{8}.$$

Como Σ é uma matriz 3×2, temos que

$$\Sigma = \begin{bmatrix} \sqrt{10} & 0 \\ 0 & \sqrt{8} \\ 0 & 0 \end{bmatrix}.$$

Cálculo dos autovetores associados a $\lambda_1 = 10$:

$$(A^T A - \lambda_1 I)\mathbf{z} = 0 \ , \quad \text{sendo } \mathbf{z} = \begin{bmatrix} x \\ y \end{bmatrix}, \ \ x \text{ e } y \in \mathbb{R}$$

Logo,

$$\begin{bmatrix} 9-\lambda & 1 \\ 1 & 9-\lambda \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Resolvendo o sistema temos que $x = y$.
Seja $x = 1$. Então:

$$\mathbf{z}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Cálculo dos autovetores associados a $\lambda_2 = 8$:

$$(A^T A - \lambda_2 I)\mathbf{z} = 0.$$

Logo,

$$\begin{bmatrix} 9-8 & 1 \\ 1 & 9-8 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Portanto $x = -y$.
Seja $x = 1$. Então

$$\mathbf{z}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Como $\lambda_1 \neq \lambda_2$, então $\langle \mathbf{z}_1, \mathbf{z}_2 \rangle = 0$.

Então $B = \left\{ \dfrac{\mathbf{z}_1}{\|\mathbf{z}_1\|}, \dfrac{\mathbf{z}_2}{\|\mathbf{z}_2\|} \right\}$: base ortonormal de $\mathbb{R}^2$.

Deste modo obtemos a matriz ortogonal

$$V = \begin{bmatrix} \dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{2}} \\ \dfrac{1}{\sqrt{2}} & -\dfrac{1}{\sqrt{2}} \end{bmatrix}.$$

Pelo *teorema 2.4*, temos que:

$$\mathbf{u}_1 = \frac{1}{\sigma_1} A\mathbf{v}_1 = \frac{1}{\sqrt{10}} \begin{bmatrix} 2 & -2 \\ 1 & 1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} \dfrac{1}{\sqrt{2}} \\ \dfrac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 0 \\ \dfrac{\sqrt{5}}{5,} \\ \dfrac{2\sqrt{5}}{5} \end{bmatrix} = \quad \text{e}$$

$$\mathbf{u}_2 = \frac{1}{\sigma_2} A\mathbf{v}_2 = \frac{1}{\sqrt{8}} \begin{bmatrix} 2 & -2 \\ 1 & 1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} \dfrac{1}{\sqrt{2}} \\ -\dfrac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

Agora, tomamos $\mathbf{u}_3 \in N(A^T)$, isto é, $\mathbf{u}_3$ deve satisfazer $A\mathbf{u}_3 = \mathbf{O}$ e $\|\mathbf{u}_3\| = 1$. Resolvendo

$$\begin{bmatrix} 2 & 1 & 2 \\ -2 & 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

obtemos

$$N(A^T) = \{(0, -2z, z)\,;\, z \in \mathbb{R}\}.$$

Seja $z = -1$, então

$$\mathbf{u}_3 = \begin{bmatrix} 0 \\ \dfrac{2}{\sqrt{5}} \\ -\dfrac{1}{\sqrt{5}} \end{bmatrix}.$$

Segue que

$$U = \begin{bmatrix} 0 & 1 & 0 \\ \dfrac{\sqrt{5}}{5} & 0 & \dfrac{2}{\sqrt{5}} \\ \dfrac{2\sqrt{5}}{5} & 0 & -\dfrac{1}{\sqrt{5}} \end{bmatrix}.$$

Logo,

$$A = U\Sigma V^T = \begin{bmatrix} 0 & 1 & 0 \\ \dfrac{\sqrt{5}}{5} & 0 & \dfrac{2}{\sqrt{5}} \\ \dfrac{2\sqrt{5}}{5} & 0 & -\dfrac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \sqrt{10} & 0 \\ 0 & \sqrt{8} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{2}} \\ \dfrac{1}{\sqrt{2}} & -\dfrac{1}{\sqrt{2}} \end{bmatrix}.$$

$\square$

2.6. PSEUDOINVERSA E QUADRADOS MÍNIMOS

Seja $A\mathbf{x} = \mathbf{b}$ um sistema linear com m equações e n incógnitas.

Vimos que se $posto(A) = n$, existe uma única solução para o problema de quadrados mínimos.

Se $posto(A) < n$, então $posto(A^T A) < n$ e o sistema normal tem uma infinidade de soluções.

Mesmo quando temos $m < n$, podemos ter um sistema indeterminado. Deste modo não atribuiremos relações de tamanho entre m e n.

Neste caso em que $posto(A) < n$, convencionou-se escolher dentre uma infinidade de soluções a solução de quadrados mínimos com comprimento mínimo, também conhecida como *solução de quadrados mínimos de norma mínima.*

Definição 2.7. Seja $A\mathbf{x} = \mathbf{b}$ um sistema linear $m \times n$. A **solução ótima** de quadrados mínimos de $A\mathbf{x} = \mathbf{b}$ é a solução deste tipo $\widehat{\mathbf{x}}$, tal que o comprimento $\|\widehat{\mathbf{x}}\|$ é o menor possível. Ela será denotada por $\mathbf{x}^+$.

Para obtermos a solução ótima $\mathbf{x}^+$, definiremos a **Pseudoinversa** de uma matriz A.

A pseudoinversa, também conhecida como a *inversa generalizada de Moore-Penrose*, é uma generalização interessante da inversa comum.

Assim como uma matriz quadrada A com determinante não nulo possui uma inversa A^{-1}, toda matriz $m \times n$ tem uma pseudoinversa, que será denotada por A^+.

Se $A\mathbf{x} = \mathbf{b}$ um sistema linear, com A de ordem $n \times n$, A inversível, então a solução do sistema é dada por

$$\mathbf{x} = A^{-1}\mathbf{b}.$$

Agora, suponha que $A\mathbf{x} = \mathbf{b}$ tenha m equações e n incógnitas. Será visto que a solução de quadrados mínimos de norma mínima $\mathbf{x}^+$ será dada por

$$\mathbf{x}^+ = A^+\mathbf{b}.$$

Definição 2.8. PSEUDOINVERSA DE UMA MATRIZ A

Seja A uma matriz $m \times n$ e $Posto(A) = r$. Seja a decomposição em valores singulares $A = U\Sigma V^T$, com $\sigma_1 \geqslant \sigma_2 \geqslant ... \geqslant \sigma_r \geqslant 0$. Então a pseudoinversa A^+ de A é a matriz $n \times m$

$$A^+ = V\Sigma^+ U^T,$$

sendo

$$\Sigma^+ = \left[\begin{array}{c|c} \Sigma_1^{-1} & 0 \\ \hline 0 & 0 \end{array}\right] = \left[\begin{array}{cccc|c} \sigma_1^{-1} & & & & \\ & \sigma_2^{-1} & & & \\ & & \ddots & & 0 \\ & & & \sigma_r^{-1} & \\ \hline & & 0 & & 0 \end{array}\right]_{n\times m}.$$

Então,

$$\Sigma^+ = \left[\begin{array}{cccc|c} \frac{1}{\sigma_1} & & & & \\ & \frac{1}{\sigma_2} & & & \\ & & \ddots & & 0 \\ & & & \frac{1}{\sigma_r} & \\ \hline & & 0 & & 0 \end{array}\right];$$

sendo que Σ^+ possui as últimas $n - r$ linhas nulas e as últimas $m - r$ colunas nulas.

Observação:

1) Como $$\Sigma = \left[\begin{array}{cccc|c} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & 0 \\ & & & \sigma_r & \\ \hline & & 0 & & 0 \end{array}\right] = \left[\begin{array}{c|c} \Sigma_1 & 0 \\ \hline 0 & 0 \end{array}\right]_{m\times n},$$

isto é, Σ tem as últimas $m-r$ linha nulas e as últimas $n-r$ colunas nulas, então $\Sigma\Sigma^{+} = I_m$.

2) A pseudoinversa de uma matriz A é única, apesar das matrizes U e V não serem únicas, como vimos anteriormente.

Teorema 2.9. Seja A uma matriz $m \times n$. Seja $A\mathbf{x} = \mathbf{b}$ um sistema linear com m equações e n incógnitas. Então $A\mathbf{x} = \mathbf{b}$ tem solução única $\mathbf{x}^{+}$ de quadrados mínimos se e somente se $posto(A) = n$, sendo $\mathbf{x}^{+} = A^{+}\mathbf{b}$.

E ainda, se $posto(A) < n$ então existe uma infinidade de soluções de quadrados mínimos para $A\mathbf{x} = \mathbf{b}$, e a de norma mínima é $\mathbf{x}^{+} = A^{+}\mathbf{b}$.

Demonstração.

Seja $\mathbf{x} \in \mathbb{R}^n$ e $\mathbf{b} \in \mathbb{R}^m$. Seja a decomposição em valores singulares de A

$$A = U\Sigma V^T, \text{ tal que } U \in \mathbb{R}^{m\times m}, \ V^T \in \mathbb{R}^{n\times n} \text{ e } \Sigma \in \mathbb{R}^{m\times n}.$$

Já sabemos que o sistema normal tem solução única se e somente se $posto(A) = n$.

Assim, basta mostrarmos que $\mathbf{x}^{+} = A^{+}\mathbf{b}$ é sempre uma solução do problema de quadrados mínimos.

É suficiente mostrarmos que $\mathbf{x}^{+} = A^{+}\mathbf{b}$ satisfaz $A^TA\mathbf{x} = A^T\mathbf{b}$.

Vejamos:

$$A^TA\mathbf{x}^{+} = A^TA(A^{+}\mathbf{b}) = (U\Sigma V^T)^T(U\Sigma V^T)(V\Sigma^{+}U^T)\mathbf{b} =$$

$$= V\Sigma^TU^TU\Sigma V^TV\Sigma^{+}U^T\mathbf{b} = V\Sigma^T\Sigma\Sigma^{+}U^T\mathbf{b} =$$

$$= V\Sigma^TU^T\mathbf{b} = A^T\mathbf{b}$$

Deste modo temos que $A^TA\mathbf{x}^+ = A^T\mathbf{b}$, ou seja, $\mathbf{x}^+$ é solução do sistema normal.

Logo $\mathbf{x}^+ = A^+\mathbf{b}$ é uma solução de quadrados mínimos, e é única se e somente se $posto(A) = n$.

Agora vamos considerar $posto(A) = r < n$.

Sabemos que uma solução de quadrados mínimos minimiza $\|\mathbf{b} - A\mathbf{x}\|^2$ e que $\|\mathbf{b} - A\mathbf{x}\|^2 = \|\mathbf{b} - U\Sigma V^T\mathbf{x}\|^2$.

Já que $\mathbf{b} - U\Sigma V^T\mathbf{x}$ é um vetor e, pela *propriedade 1.22*, U preserva a norma, e como $U^T = U^{-1}$, temos que

$$\|\mathbf{b} - U\Sigma V^T\mathbf{x}\|^2 = \|U^T(\mathbf{b} - U\Sigma V^T\mathbf{x})\|^2 = \|U^T\mathbf{b} - U^TU\Sigma V^T\mathbf{x}\|^2 = \|U^T\mathbf{b} - \Sigma V^T\mathbf{x}\|^2.$$

Seja
$\mathbf{c} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix} = U^T\mathbf{b} \in \mathbb{R}^m$; estamos denotando por $\mathbf{c}_1$ o vetor com as primeiras r coordenadas de $\mathbf{c}$, e por $\mathbf{c}_2$ o vetor com as últimas $m - r$ coordenadas.

Seja
$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = V^T\mathbf{x} \in \mathbb{R}^n$; estamos denotando por $\mathbf{y}_1$ o vetor com as primeiras r coordenadas de $\mathbf{y}$, e por $\mathbf{y}_2$ o vetor com as últimas $n - r$ coordenadas.

Portanto $\|U^T\mathbf{b} - \Sigma V^T\mathbf{x}\|^2 = \|\mathbf{c} - \Sigma\mathbf{y}\|^2$.

Temos que

$$\Sigma\mathbf{y} = \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_r & \\ \hline & 0 & & 0 \end{array}\right] \begin{bmatrix} y_1 \\ \vdots \\ y_r \\ y_{r+1} \\ \vdots \\ y_n \end{bmatrix},$$

sendo que Σ tem as últimas $m - r$ linhas nulas e as últimas $n - r$ colunas nulas.

Assim,

$$\Sigma\mathbf{y} = \left[\begin{array}{c|c} \Sigma_1 & 0 \\ - & - \\ 0 & 0 \end{array}\right] \begin{bmatrix} y_1 \\ \vdots \\ y_r \\ y_{r+1} \\ \vdots \\ y_n \end{bmatrix},$$

ou seja,

$$\Sigma\mathbf{y} = \begin{bmatrix} \sigma_1 y_1 \\ \vdots \\ \sigma_r y_r \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \Sigma_1\mathbf{y}_1 \\ \\ \mathbf{O} \end{bmatrix}.$$

As últimas $m - r$ coordenadas de $\Sigma\mathbf{y}$ são nulas.

Deste modo,

$$\mathbf{c} - \Sigma\mathbf{y} = \begin{bmatrix} c_1 \\ \vdots \\ c_r \\ c_{r+1} \\ \vdots \\ c_m \end{bmatrix} - \begin{bmatrix} \sigma_1 y_1 \\ \vdots \\ \sigma_r y_r \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Então,

$$\mathbf{c} - \Sigma\mathbf{y} = \begin{bmatrix} c_1 - \sigma_1 y_1 \\ \vdots \\ c_r - \sigma_r y_r \\ c_{r+1} \\ \vdots \\ c_m \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1 - \Sigma_1\mathbf{y}_1 \\ \\ \mathbf{c}_2 \end{bmatrix}.$$

Portanto, $\|\mathbf{b} - A\mathbf{x}\|^2 = \|\mathbf{c} - \Sigma\mathbf{y}\|^2$, com $\mathbf{c} - \Sigma\mathbf{y}$ dado acima.
Mas,

$$\|\mathbf{c} - \Sigma\mathbf{y}\|^2 = \sum_{i=1}^{r} (c_i - \sigma_i y_i)^2 + \sum_{i=r+1}^{m} c_i^2.$$

Logo, minimizar $\|\mathbf{b} - A\mathbf{x}\|^2$ é o mesmo que minimizar a expressão dada acima.

Como o vetor $\mathbf{c}_2$ não depende de $\mathbf{x}$, segue-se que $\|\mathbf{b} - A\mathbf{x}\|^2$ é mínima quando $\sum_{i=1}^{r} (c_i - \sigma_i y_i)^2 = 0$, ou seja, quando $\mathbf{c}_1 - \Sigma_1\mathbf{y}_1 = \mathbf{O}$.

Isto ocorre, se $y_i = \dfrac{c_i}{\sigma_i}$, $i = 1, ..., r$.

Em termos matriciais, para que $\mathbf{c}_1 - \Sigma_1\mathbf{y}_1 = \mathbf{O}$, devemos ter $\mathbf{y}_1 = (\Sigma_1)^{-1}\mathbf{c}_1$, isto é, $\mathbf{y}$ é um vetor da forma

$$\mathbf{y} = \begin{bmatrix} (\Sigma_1)^{-1}\mathbf{c}_1 \\ \mathbf{y}_2 \end{bmatrix}.$$

Logo, para que $\|\mathbf{c} - \Sigma\mathbf{y}\|$ seja mínima, $\mathbf{y}$ deve ter a forma acima; as coordenadas de $\mathbf{y}_2$ assumem qualquer valor.
Como $\mathbf{y} = V^T\mathbf{x}$, segue-se que $\mathbf{x} = V\mathbf{y}$.
Logo, uma solução de quadrados mínimos de $A\mathbf{x} = \mathbf{b}$ é da forma $\mathbf{x} = V\mathbf{y}$ com

$$\mathbf{y} = \begin{bmatrix} (\Sigma_1)^{-1}\mathbf{c}_1 \\ \mathbf{y}_2 \end{bmatrix}.$$

Fazendo $\mathbf{y}_2 = \mathbf{O}$, segue-se que

$$\mathbf{x} = V \begin{bmatrix} (\Sigma_1)^{-1}\mathbf{c}_1 \\ \mathbf{O} \end{bmatrix} = V \left[\begin{array}{c|c} (\Sigma_1)^{-1} & 0 \\ \hline 0 & 0 \end{array}\right] \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix}.$$

$$\mathbf{x} = V \left[\begin{array}{ccc|c} \frac{1}{\sigma_1} & & & \\ & \ddots & & 0 \\ & & \frac{1}{\sigma_r} & \\ \hline & 0 & & 0 \end{array}\right] \begin{bmatrix} \sigma_1 y_1 \\ \vdots \\ \sigma_r y_r \\ 0 \\ \vdots \\ 0 \end{bmatrix} = V\Sigma^+U^T\mathbf{b} = A^+\mathbf{b}.$$

Logo, $A^+\mathbf{b}$ é uma solução de quadrados mínimos que denotaremos por $\mathbf{x}^+$, ou seja,

$$\mathbf{x}^+ = A^+\mathbf{b} \ .$$

Seja $\mathbf{z}$ outra solução de quadrados mínimos. Então

$$\mathbf{z} = V\mathbf{y} = V\begin{bmatrix} (\Sigma_1)^{-1}\mathbf{c}_1 \\ \mathbf{y}_2 \end{bmatrix}, \text{ sendo que neste caso temos que fazer } \mathbf{y}_2 \neq \mathbf{O}.$$

Deste modo

$$\|\mathbf{z}\|^2 = \|V\mathbf{y}\|^2 = \|\mathbf{y}\|^2 = \left\|(\Sigma_1)^{-1}\mathbf{c}_1\right\|^2 + \|\mathbf{y}_2\|^2 > \left\|(\Sigma_1)^{-1}\mathbf{c}_1\right\|^2 = \|\mathbf{x}^+\|^2.$$

Logo, $\mathbf{x}^+$ obtida acima é a solução de quadrados mínimos de norma mínima.

$\square$

Exemplo 2.7. DETERMINAR A SOLUÇÃO ÓTIMA DE QUADRADOS MÍNIMOS DO SISTEMA

$$\begin{cases} -3x + y = 5 \\ 6x - 2y = 3 \\ 6x - 2y = 1 \end{cases}$$

Seja a matriz

$$A = \begin{bmatrix} -3 & 1 \\ 6 & -2 \\ 6 & -2 \end{bmatrix}_{3\times 2} \qquad \text{e} \qquad \mathbf{b} = \begin{bmatrix} 5 \\ 3 \\ 1 \end{bmatrix}.$$

As colunas de A são linearmente dependentes, logo $posto(A) = 1$.
Podemos também escalonar a matriz A obtendo a matriz na forma escada

$$\begin{bmatrix} 1 & -\frac{1}{3} \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \text{ isto é, } posto(A) = 1 < n = 2 \ .$$

Portanto existe uma infinidade de soluções de quadrados mínimos.

Deste modo resolveremos o problema de quadrados mínimos através da pseudoinversa da matriz A.

Seja a decomposição abaixo a decomposição em valores singulares de A.

$$A = \begin{bmatrix} -\frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \end{bmatrix} \begin{bmatrix} 3\sqrt{10} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{3}{\sqrt{10}} & -\frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{bmatrix}.$$

Seja

$$U = \begin{bmatrix} -\frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 3\sqrt{10} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{e} \quad V^T = \begin{bmatrix} \frac{3}{\sqrt{10}} & -\frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{bmatrix}.$$

Da *definição 2.8* obtemos:

$$\Sigma^+ = \begin{bmatrix} \frac{1}{3\sqrt{10}} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}_{2\times 3}.$$

Já que $A^+ = V\Sigma^+U^T$, temos que

$$A^+ = \begin{bmatrix} \frac{3}{10}\sqrt{10} & \frac{1}{10}\sqrt{10} \\ -\frac{1}{10}\sqrt{10} & \frac{3}{10}\sqrt{10} \end{bmatrix} \begin{bmatrix} \frac{1}{3\sqrt{10}} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -\frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \end{bmatrix}.$$

Logo,

$$A^+ = \begin{bmatrix} -\dfrac{1}{30} & \dfrac{1}{15} & \dfrac{1}{15} \\ \\ \dfrac{1}{90} & -\dfrac{1}{45} & -\dfrac{1}{45} \end{bmatrix}.$$

Pelo *teorema 2.9* temos que a nossa solução tem a forma $\mathbf{x}^+ = A^+\mathbf{b}$.
Assim,

$$\mathbf{x}^+ = \begin{bmatrix} -\dfrac{1}{30} & \dfrac{1}{15} & \dfrac{1}{15} \\ \\ \dfrac{1}{90} & -\dfrac{1}{45} & -\dfrac{1}{45} \end{bmatrix} \begin{bmatrix} 5 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} \dfrac{1}{10} \\ \\ -\dfrac{1}{30} \end{bmatrix}$$

é a solução ótima de quadrados mínimos, ou seja, $\mathbf{x}^+$ é a solução de norma mínima do sistema dado.

□

Teorema 2.10. Seja A uma matriz de ordem $n \times n$ inversível. Seja $posto(A) = r$. Então $A^+ = A^{-1}$,ou seja, temos um caso particular da pseudoinversa para matrizes quadradas e inversíveis.

Demonstração.

Pelo *teorema 1.29*, as colunas de A são linearmente independentes, ou seja, $posto(A) = r = n$.

Sabemos que $A = U\Sigma V^T$.

Como A é $n \times n$, temos que U, Σ e V^T são matrizes de ordem $n \times n$.

Assim, pelo *teorema 1.21*, U e V^T são matrizes ortogonais, ou seja, pela *definição 1.20*,

$$U^T = U^{-1} \quad \text{e} \quad V^T = V^{-1}.$$

Já que $det(A) \neq 0$ temos que

$$det(A) = det(U\Sigma V^T) = det(U)det(\Sigma)det(V^T) \neq 0.$$

Pelo *teorema 1.20*, $det(U) = det(V^T) = \pm 1$.
Deste modo,

$$det(A) = det(\Sigma) \neq 0, \text{ isto é, } \Sigma \text{ é uma matriz inversível.}$$

Portanto, Σ é uma matriz diagonal, tal que $\sigma_i \neq 0$, $i = 1, ..., n$.
Logo,

$$A^{-1} = \left(U\Sigma V^T\right)^{-1} = \left(V^T\right)^{-1}\Sigma^{-1}U^{-1} = V\Sigma^{-1}U^T =$$

$$= V \begin{bmatrix} \frac{1}{\sigma_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sigma_n} \end{bmatrix} U^T = A^+.$$

□

É também possível definir A^+ por suas propriedades algébricas. Essas propriedades são dadas nas condições a seguir.

Teorema 2.11. Seja A uma matriz $m \times n$. Então existe uma única matriz X de ordem $n \times m$ que satisfaz as seguintes condições:

1. $AXA = A$
2. $XAX = X$
3. $(AX)^T = AX$
4. $(XA)^T = XA$

Essas condições são chamadas de *Condições de Penrose*.

Demonstração.

Seja $X = A^+ = V\Sigma^+U^T$.

Então temos:

1. $AA^+A = U\Sigma V^T(V\Sigma^+U^T)U\Sigma V^T = U\Sigma\Sigma^+\Sigma V^T = U\Sigma V^T = A$

2. $A^+AA^+ = V\Sigma^+U^T(U\Sigma V^T)V\Sigma^+U^T = V\Sigma^+U^T = A^+$

3. $(AA^+)^T = (U\Sigma V^T(V\Sigma^+U^T))^T = (U\Sigma\Sigma^+U^T)^T = (UU^T)^T = (I)^T = I$
$AA^+ = U\Sigma V^T(V\Sigma^+U^T) = U\Sigma\Sigma^+U^T = UU^T = I$

4. $(A^+A)^T = (V\Sigma^+U^T(U\Sigma V^T))^T = (V\Sigma^+\Sigma V^T)^T = (VV^T)^T = (I)^T = I$
$A^+A = V\Sigma^+U^T(U\Sigma V^T) = V\Sigma^+\Sigma V^T = VV^T = I$

Como vimos, $X = A^+$ satisfaz as condições de Penrose.
Suponha que exista uma matriz Y que também satisfaça tais condições.
Deste modo temos:

$$Y \stackrel{2}{=} YAY \stackrel{3}{=} Y(AY)^T = YY^TA^T \stackrel{1}{=} YY^T(AXA)^T = Y(Y^TA^T)(X^TA^T) =$$

$$= Y(AY)^T(AX)^T \stackrel{3}{=} YAYAX \stackrel{1}{=} YAX$$

Por outro lado temos:

$$X \stackrel{2}{=} XAX \stackrel{4}{=} (XA)^TX = A^TX^TX \stackrel{1}{=} (AYA)^TX^TX = (A^TY^T)(A^TX^T)X =$$

$$= (YA)^T(XA)^TX \stackrel{4}{=} YAXAX \stackrel{1}{=} YAX.$$

Logo $X = Y$.
Portanto $X = Y = A^+$ é a única matriz que satisfaz as condições de Penrose.

$\square$

Conclusão

O método de quadrados mínimos é muito utilizado em várias áreas, como na Física Experimental, Astronomia, Biologia, Administração e Estatística. Já existe até mesmo um modelo de quadrados mínimos para a audição humana [1].

Para que este trabalho fosse realizado alguns livros foram pesquisados e assim, tomei conhecimento da existência de outras técnicas para a obtenção de decomposições de matrizes, visando resolver o problema de quadrados mínimos.

Essas técnicas são utilizadas para a obtenção de resultados mais precisos. Uma dessas ferramentas é o algoritmo de Golub-Reinsch, que é utilizado para obter os valores singulares após uma bidiagonalização obtida através de transformações de Householder. Entretanto, essas técnicas mais refinadas não foram estudadas aqui, e exigem um conhecimento maior de análise numérica.

Para completar, o trabalho que aqui se conclui exigiu que se aprendesse a utilizar o editor Scientific WorkPlace, o software gráfico Origin e alguns comandos do Maple, os quais contribuíram para uma melhor formação em informática. A experiência desta pesquisa trouxe benefícios principalmente pelo aprendizado na produção de relatórios matemáticos e pelo aprofundamento dos conhecimentos adquiridos em Álgebra Linear, especialmente do Método de Quadrados Mínimos.

Referências Bilbliográficas

[1] ANTON, Howard, RORRES, Chris. Elementary Linear Algebra: Applications Version. 7. ed. New York: John Wiley & Sons, 1994.

[2] EDWARDS Jr., C.H., PENNEY, David E. Introdução à Álgebra Linear. Rio de Janeiro: Prentice-Hall do Brasil, 1998.

[3] KOLMAN, Bernard. Introdução à Álgebra Linear com Aplicações. Rio de Janeiro: Prentice-Hall do Brasil, 1998.

[4] LAWSON, Terry. Álgebra Linear. São Paulo: Edgard Blucher, 1997.

[5] LAY, David C. Álgebra Linear e suas Aplicações. 2. ed. Rio de Janeiro: Livros Técnicos e Científicos, 1999.

[6] LEWIS, David W. Matrix Theory. Singapore: World Scientific, 1991.

[7] LEON, Steven J. Álgebra Linear com Aplicações. 4. ed. Rio de Janeiro: Livros Técnicos e Científicos, 1999.

[8] NOBLE, Ben, DANIEL, James. Álgebra Linear Aplicada. 2. ed. Rio de Janeiro: Prentice-Hall do Brasil, 1986.

[9] ROBINSON, Derek J. S. A Course in Linear Algebra with Applications. Singapore: World Scientific, 1991.

[10] STRANG, Gilbert. Linear Algebra and its Applications. 3. ed. Orlando: Harcourt Brace Jovanovich, 1988.

[11] WATKINS, David S. Fundamentals of Matrix Computations. New York: Jonh Wiley & Sons, 1991.

[12] SOUTO, Gilberto. Decomposição em Valores Singulares. Florianópolis: Departamento de Matemática/UFSC, 2000. 62 p. Trabalho de Conclusão de Curso.

Referências para WWW (World Wide Web)

[$W1$] BEYOND ORDINARY LEAST SQUARE. http://www.physics.csbsju.edu/stats/fitting_lines.html. Consultado em 10/04/02.