

Módulo 08

• Note bem, a leitura destes apontamentos não dispensa de modo algum a leitura atenta da bibliografia principal da cadeira

• Chama-se à atenção para a importância do trabalho pessoal a realizar pelo aluno resolvendo os problemas apresentados na bibliografia, sem consulta prévia das soluções propostas, análise comparativa entre as suas resposta e a respostas propostas, e posterior exposição junto do docente de todas as dúvidas associadas.



Complementos e Aplicações.

[Poole 577 a 615]

Decomposição em valores singulares (SVD).

Valores singulares.

Decomposição em valores singulares.

Interpretação geométrica.

SVD e os 4 espaços fundamentais de uma matriz.

Minimização do erro quadrático.

Melhor aproximação.

Vector de erro.

Mínimo erro quadrático.

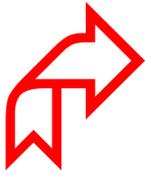
Solução de mínimo erro quadrático.

Equações normais.

Quadrados mínimos e factorização QR.

Quadrados mínimos e projecção num subespaço.

Matriz pseudo-inversa.



Valores Singulares.

1. Dada uma matriz \mathbf{A} , $m \times n$, a matriz $\mathbf{A}^T \mathbf{A}$, $n \times n$, é uma matriz simétrica com n valores próprios reais, não negativos. Designam-se por **valores singulares** da matriz \mathbf{A} as n raízes quadradas dos valores próprios da matriz $\mathbf{A}^T \mathbf{A}$, $\sigma_i = \sqrt{\lambda_i}$.



Exemplo 1.

Dada a matriz rectangular, 3×2 ,

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{2} \\ 0 & \sqrt{2} \end{bmatrix}$$

temos

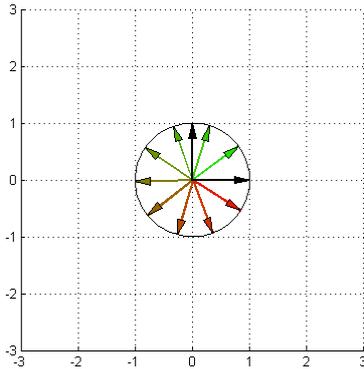
$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sqrt{2} & \sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{2} \\ 0 & \sqrt{2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$$

A matriz $\mathbf{A}^T \mathbf{A}$ tem $n = 2$ valores próprios, raízes do polinómio característico $p(\lambda) = (1 - \lambda)(4 - \lambda)$, $\lambda_1 = 4$ e $\lambda_2 = 1$, pelo que a matriz \mathbf{A} tem valores singulares

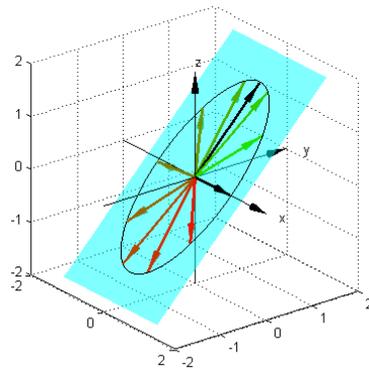
$$\sigma_1 = \sqrt{\lambda_1} = 2$$

e

$$\sigma_2 = \sqrt{\lambda_2} = 1$$



a)



b)

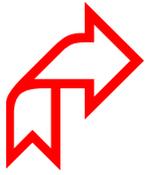
Na figuras a) a b) mostra-se, respectivamente, um conjunto de versores objecto, com extremidades sobre a circunferência de raio unitário, em \mathbb{R}^2 , e os respectivos vectores imagem, em \mathbb{R}^3 , resultantes da transformação linear $T: \mathbb{R}^2 \rightarrow \mathbb{R}^3$ a que a matriz \mathbf{A} está associada, que constituem um conjunto de vectores com extremidade sobre uma elipse em \mathbb{R}^3 .

A matriz $\mathbf{A}^T \mathbf{A}$ tem vectores próprios de norma unitária $\mathbf{v}_1 = [0 \ 1]^T$ e $\mathbf{v}_2 = [1 \ 0]^T$, assinalados a preto na figura a). As suas imagens, assinaladas a preto na figura b), estão sobre os eixos da elipse, e são vectores de comprimento $\sigma_1 = 2$ e $\sigma_2 = 1$. Na verdade, considerando um valor próprio da matriz $\mathbf{A}^T \mathbf{A}$, λ , e o correspondente vector próprio de norma unitária \mathbf{v} , temos, em geral

$$\begin{aligned} \|\mathbf{A}\mathbf{v}\|^2 &= (\mathbf{A}\mathbf{v}) \cdot (\mathbf{A}\mathbf{v}) = (\mathbf{A}\mathbf{v})^T (\mathbf{A}\mathbf{v}) = \mathbf{v}^T \mathbf{A}^T \mathbf{A} \mathbf{v} = \mathbf{v}^T (\lambda \mathbf{v}) \\ &= \mathbf{v}^T \lambda \mathbf{v} = \lambda (\mathbf{v}^T \mathbf{v}) = \lambda (\mathbf{v} \cdot \mathbf{v}) = \lambda \|\mathbf{v}\|^2 = \lambda \end{aligned}$$

e, portanto, os vectores imagem dos valores próprios de norma unitária têm comprimento igual ao valor singular associado

$$\mathbf{w} = \|\mathbf{A}\mathbf{v}\| = \sqrt{\lambda} = \sigma$$



Decomposição em Valores Singulares (SVD).

(Vimos que uma matriz quadrada A , $n \times n$, é diagonalizável, $A = PDP^{-1}$, sse possui n vectores próprios linearmente independentes, sendo P a matriz dos valores próprios e D a matriz diagonal dos valores próprios da matriz A . Se A for uma matriz simétrica, é diagonalizável, $A = QDQ^T$, por uma matriz ortogonal Q , obtida a partir dos vectores próprios.)

2. Toda a matriz A , $m \times n$, é factorizável na forma

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

, chamada **decomposição em valores singulares** da matriz A , sendo U uma matriz ortogonal $m \times m$, V uma matriz ortogonal $n \times n$, e, tendo A r valores singulares não nulos, Σ é uma matriz $m \times n$ da forma

$$\Sigma = \left[\begin{array}{cc|cc} \overbrace{}^r & \overbrace{}^{n-r} & & \\ & 0 & \dots & 0 \\ & \vdots & \ddots & \vdots \\ & 0 & \dots & 0 \\ \underbrace{0 \dots 0}_{m-r} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 0 \end{array} \right] \left. \begin{array}{l} r \\ \\ \\ m-r \end{array} \right\} , \text{ em que } D = \begin{bmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_r \end{bmatrix}$$

3. De modo análogo à decomposição espectral de uma matriz simétrica, a matriz A , $m \times n$, com r valores singulares não nulos, pode ser escrita na forma

$$A = \sigma_1 u_1 v_1^T + \dots + \sigma_r u_r v_r^T = \sum_{i=1}^r \sigma_i u_i v_i^T$$

, em que u_i são colunas de U , ditos vectores singulares à esquerda, e v_i são colunas de V , ditos vectores singulares à direita.



Exemplo 2.

Consideremos de novo a matriz

$$A = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{2} \\ 0 & \sqrt{2} \end{bmatrix}$$

Como vimos

$$A^T A = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$$

tem valores próprios $\lambda_1 = 4$ e $\lambda_2 = 1$, a que estão associados os vectores próprios unitários $v_1 = [0 \ 1]^T$ e $v_2 = [1 \ 0]^T$, pelo que a matriz A tem valores singulares $\sigma_1 = \sqrt{\lambda_1} = 2$ e $\sigma_2 = \sqrt{\lambda_2} = 1$. Podemos de imediato escrever a matriz diagonal dos valores singulares

$$D = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

, e a matriz Σ , de dimensão $m \times n = 3 \times 2$,

$$\Sigma = \begin{bmatrix} \mathbf{D} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

\mathbf{V} é uma matriz ortogonal $n \times n$, portanto da dimensão de $\mathbf{A}^T \mathbf{A}$, no caso presente, 2×2 . Para construir \mathbf{V} basta encontrar um conjunto ortonormado de vectores, próprios da matriz $\mathbf{A}^T \mathbf{A}$, $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$, associados a cada um dos n valores singulares. No caso presente temos de imediato

$$\mathbf{V} = [\mathbf{v}_1 \quad \mathbf{v}_2] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

\mathbf{U} é uma matriz ortogonal $m \times m$, cujos elementos são determinados por

$$\mathbf{u}_i = \frac{1}{\sigma_i} \mathbf{A} \mathbf{v}_i$$

Tendo \mathbf{A} r valores singulares não nulos, ficam assim encontradas r colunas da matriz \mathbf{U} . Caso tenhamos $r < m$ será necessário determinar os restantes $m - r$ versores de \mathbb{R}^m (por exemplo recorrendo à ortogonalização de Gram-Schmidt tendo por base os versores canónicos).

No caso presente temos $m \times m = 3 \times 3$ e $r = 2$, pelo que, de imediato, temos

$$\begin{aligned} \mathbf{u}_1 &= \frac{1}{\sigma_1} \mathbf{A} \mathbf{v}_1 = \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{2} \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \\ \mathbf{u}_2 &= \frac{1}{\sigma_2} \mathbf{A} \mathbf{v}_2 = \frac{1}{1} \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{2} \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \end{aligned}$$

, sendo agora necessário determinar um versor ortogonal a \mathbf{u}_1 e \mathbf{u}_2 .

Considerando a base $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{e}_3\}$, temos

$$\begin{aligned} \mathbf{u}'_3 &= \mathbf{e}_3 - \text{proj}_{\mathbf{u}_2} \mathbf{e}_3 - \text{proj}_{\mathbf{u}_1} \mathbf{e}_3 \\ &= \mathbf{e}_3 - \frac{\mathbf{e}_3 \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2} \mathbf{u}_2 - \frac{\mathbf{e}_3 \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 = \mathbf{e}_3 - (\mathbf{e}_3 \cdot \mathbf{u}_2) \mathbf{u}_2 - (\mathbf{e}_3 \cdot \mathbf{u}_1) \mathbf{u}_1 \\ &= \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - 1/\sqrt{2} \begin{bmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 1/2 \\ 1/2 \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ -1/2 \\ 1/2 \end{bmatrix} \end{aligned}$$

e

$$\mathbf{u}_3 = \frac{\mathbf{u}'_3}{\|\mathbf{u}'_3\|} = \begin{bmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

Temos então

$$U = [u_1 \quad u_2 \quad u_3] = \begin{bmatrix} 0 & 1 & 0 \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix}$$

Ficando assim determinada a decomposição em valores singulares da matriz A

$$A = U\Sigma V^T = \begin{bmatrix} 0 & 1 & 0 \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

, como podemos confirmar



```
>> x=1/sqrt(2);
>> A=[1 0; 0 sqrt(2); 0 sqrt(2)];
>> U=[0 1 0; x 0 -x; x 0 x];
>> S=[2 0; 0 1; 0 0];
>> V=[0 1; 1 0];
>> U*S*V'
ans =
    1.0000         0
         0    1.4142
         0    1.4142
```

Podemos ainda decompor A na forma

$$A = \sum_{i=1}^r \sigma_i u_i v_i^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T$$

$$= 2 \begin{bmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix} + 1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} = 2 \begin{bmatrix} 0 & 1 \\ 0 & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} \end{bmatrix} + 1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Embora possamos recorrer ao MatLab para proceder ao cálculos das matrizes U, Σ, e V, conforme acima ficou descrito, o procedimento é desnecessário, dada a existência da função **svd(A)**, que procede à decomposição em valores singulares da matriz A



```
>> A=[1 0; 0 sqrt(2); 0 sqrt(2)];
>> [U S V]=svd(A)
>> U =
         0    1.0000         0
    -0.7071         0   -0.7071
    -0.7071         0    0.7071
>> S =
    2.0000         0
         0    1.0000
         0         0
>> V =
         0     1
        -1     0
```

Exemplo 3. Interpretação Geométrica

Consideremos a transformação linear $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, em que a matriz da transformação é

$$A = \begin{bmatrix} 2 & 1 \\ 1/2 & 2 \end{bmatrix}$$

A figura a) mostra um conjunto de vectores com extremidade sobre a circunferência de raio unitário, e a figura d) as imagens resultantes da transformação, $w = Au$, estando assinalados a preto os vectores próprios de $A^T A$. Podemos verificar que A tem uma decomposição em valores singulares

$$A = U\Sigma V^T = \begin{bmatrix} 0.75 & -0.66 \\ 0.66 & 0.75 \end{bmatrix} \begin{bmatrix} 2.75 & 0 \\ 0 & 1.27 \end{bmatrix} \begin{bmatrix} 0.66 & 0.75 \\ -0.75 & 0.66 \end{bmatrix}$$

(para uma maior clareza de exposição, os valores são aproximados e são omitidos os passos intermédios). Atendendo à decomposição de A em valores singulares, a transformação T pode ser interpretada como uma composição de 3 transformações lineares $\mathbb{R}^2 \rightarrow \mathbb{R}^2$ com matrizes V^T , Σ e U . A matriz

$$V^T = \begin{bmatrix} 0.66 & 0.75 \\ -0.75 & 0.66 \end{bmatrix}$$

tem a estrutura de uma matriz de rotação em \mathbb{R}^2

$$\begin{bmatrix} \cos(\theta) & -\text{sen}(\theta) \\ \text{sen}(\theta) & \cos(\theta) \end{bmatrix}$$

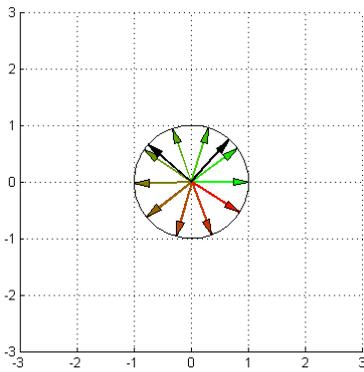
, com $\theta \approx -49^\circ$. As imagens $V^T u$ resultam de uma rotação de $\approx 49^\circ$ em sentido inverso, como se pode ver na figura b), ficando os vectores próprios de $A^T A$ alinhados com os eixos coordenados. A matriz

$$\Sigma = \begin{bmatrix} 2.75 & 0 \\ 0 & 1.27 \end{bmatrix}$$

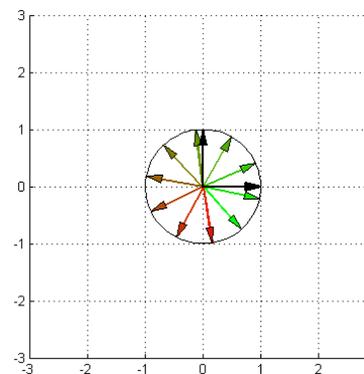
vai em seguida multiplicar cada um dos vectores próprios de $A^T A$ por $\sigma_1 = 2.75$ $\sigma_2 = 1.27$ (e cada um dos outros vectores por uma combinação linear de σ_1 e σ_2) transformando a circunferência de raio unitário numa elipse de semieixos σ_1 e σ_2 . Finalmente a matriz

$$U = \begin{bmatrix} 0.75 & -0.66 \\ 0.66 & 0.75 \end{bmatrix}$$

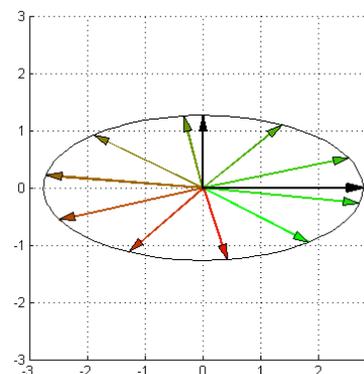
que tem a estrutura de uma matriz de rotação em \mathbb{R}^2 , com $\theta \approx 49^\circ$, provoca uma rotação em sentido directo de $\approx 49^\circ$. Todos os produtos matriciais são susceptíveis desta interpretação geométrica, dado que toda a matriz pode se decomposta na for $U\Sigma V^T$.



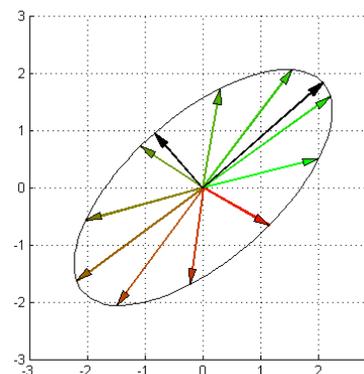
a) u



b) $V^T u$

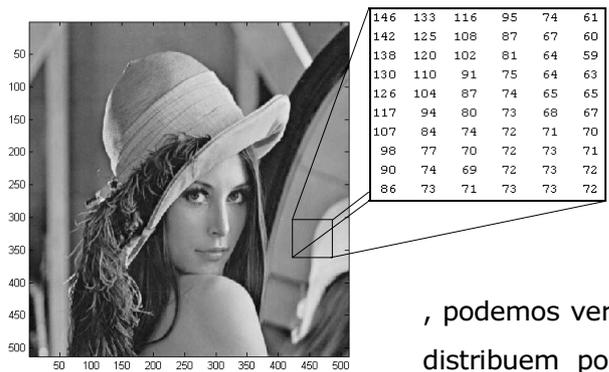


c) $\Sigma V^T u$



d) $w = Au = U\Sigma V^T u$

Exemplo 4. Aplicação: Compressão de Imagem.



Consideremos a imagem com 512×512 pixels que se reproduz na figura a). Cada um dos pixels da imagem corresponde a um número inteiro entre 0 e 255, relativo a 1 de 256 níveis de cinzento, constituindo toda a imagem uma matriz de $512 \times 512 = 262144$ números.

Decompondo a matriz em valores singulares,

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

, podemos verificar que os seus 512 valores singulares, σ_i , se distribuem por 10^6 ordens de grandeza, sendo o maior dos valores $\sigma_1 \approx 5 \times 10^4$, e o menor $\sigma_{512} \approx 2 \times 10^{-2}$, como podemos ver na figura d) (note a escala logarítmica).

Se fizermos a decomposição espectral da matriz,

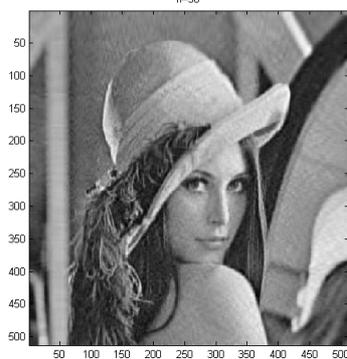
$$A = \sum_{i=1}^{512} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$$

, e sendo os coeficientes das matrizes $\mathbf{u}_i \mathbf{v}_i^T$ todos, em módulo, menores que 1, dado que os vectores próprios à esquerda e à direita, \mathbf{u}_i e \mathbf{v}_i , são vectores normalizados, resulta que cada uma das parcelas, $\sigma_i \mathbf{u}_i \mathbf{v}_i^T$, contribui de modo muito diferente para a matriz final.

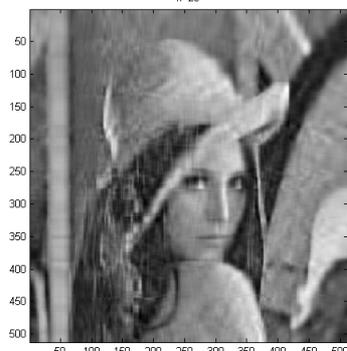
As figuras a) a c) mostram as imagens resultantes de se considerar o total das componentes espectrais, e apenas as 50 e 20 mais importantes, isto é, associadas aos 50 e 20 valores singulares de maior valor. Note que, sendo necessário $512 \times 512 = 262144$ números para reproduzir a imagem original, a imagem que utiliza apenas 50 componentes espectrais necessita de apenas 50 valores singulares mais 50 vectores \mathbf{u} e 50 vectores \mathbf{v} , cada um de dimensão 512, ou seja $50 \times (1 + 512 + 512) = 51250$ números, cerca de 5 vezes menos informação.

A decomposição de imagens em valores singulares está na base de alguns dos mais eficientes algoritmos de compressão de imagem, que estudará nas cadeiras da especialidade. O armazenamento e transmissão de dados têm um papel extremamente importante no actual panorama tecnológico, e o estudo das técnicas que permitam a sua execução com o mínimo espaço de armazenamento e o menor tempo de processamento será objecto de análise ao longo do curso. Na maioria das técnicas são extremamente relevantes os considerandos de ordem estatística e numérica, que transcendem o âmbito desta cadeira, e em que terá oportunidade de se iniciar nas cadeiras de Probabilidades e Estatística (2º semestre) e Métodos Numéricos e Optimização (3º semestre), onde deverá criar base sólidas que serão utilizadas nas cadeiras da especialidade dos semestres seguintes, nomeadamente nas que estão associadas aos ramos de produção e organização de conteúdos.

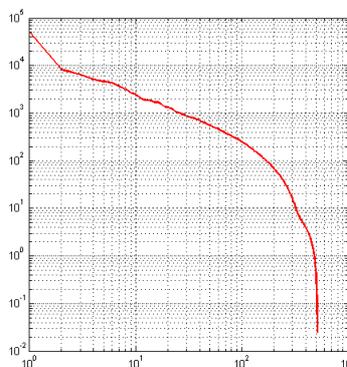
a) n=512



b) n=50



c) n=20



d) valores singulares

SVD e os 4 Espaços Fundamentais de uma Matriz.

- 4.** Sendo $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$ a decomposição em valores singulares da matriz \mathbf{A} , $m \times n$, com r valores singulares não nulos, então
1. A matriz \mathbf{A} tem característica r .
 2. $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ é uma base ortonormada de $\text{col}(\mathbf{A})$.
 3. $\{\mathbf{u}_{r+1}, \dots, \mathbf{u}_m\}$ é uma base ortonormada de $\text{Ker}(\mathbf{A}^T)$.
 4. $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ é uma base ortonormada de $\text{lin}(\mathbf{A})$.
 5. $\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$ é uma base ortonormada de $\text{Ker}(\mathbf{A})$.



Exemplo 5.

Recordemos a análise que foi feita nos exercícios 7 a 12 do módulo 7 sobre os espaços fundamentais da matriz

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 3 & 1 & 3 \\ 1 & -1 & 0 & 1 & -1 \\ -1 & 2 & 1 & -2 & 0 \\ 3 & 0 & 3 & 0 & -6 \end{bmatrix}$$

Vimos que matriz tem característica 3, sendo que os vectores

$$\mathbf{u}_1 = [1 \ 0 \ 0 \ 0]^T, \mathbf{u}_2 = [0 \ 1 \ 0 \ 6]^T \text{ e } \mathbf{u}_3 = [0 \ 0 \ 1 \ 3]^T$$

formam uma base do espaço coluna de \mathbf{A} , o vector

$$\mathbf{u}_4 = [0 \ -6 \ -3 \ 1]^T$$

forma uma base do núcleo de \mathbf{A}^T , os vectores

$$\mathbf{v}_1 = [1 \ 0 \ 1 \ 0 \ -2]^T, \mathbf{v}_2 = [0 \ 1 \ 1 \ 0 \ 3]^T \text{ e } \mathbf{v}_3 = [0 \ 0 \ 0 \ 1 \ 4]^T$$

formam uma base do espaço linha de \mathbf{A} , e os vectores

$$\mathbf{v}_4 = [-1 \ -1 \ 1 \ 0 \ 0]^T, \text{ e } \mathbf{v}_5 = [2 \ -3 \ 0 \ -4 \ 1]^T$$

formam uma base do núcleo de \mathbf{A} .

Calculemos a decomposição em valores singulares da matriz \mathbf{A}



```
>> A=[2 1 3 1 3; 1 -1 0 1 -1;-1 2 1 -2 0; 3 0 3 0 -6];
>> [U S V]=svd(A)
U =
    0.0982    -0.9910    0.0911    0.0000
   -0.1738     0.0225    0.4321   -0.8847
    0.0211   -0.0801   -0.8930   -0.4423
   -0.9796   -0.1051   -0.0867    0.1474
S =
    7.4765         0         0         0         0
         0     4.8770         0         0         0
         0         0     3.5096         0         0
         0         0         0     0.0000         0
```

$$V = \begin{bmatrix} -0.3929 & -0.4500 & 0.3554 & 0.7136 & -0.0873 \\ 0.0420 & -0.2406 & -0.6060 & 0.0811 & -0.7526 \\ -0.3508 & -0.6906 & -0.2507 & -0.4606 & 0.3534 \\ -0.0157 & -0.1658 & 0.6580 & -0.5060 & -0.5323 \\ 0.8488 & -0.4849 & 0.1031 & 0.1265 & 0.1331 \end{bmatrix}$$

Como vemos, por análise da matriz Σ , a matriz A tem $r = 3$ valores singulares não nulos, ou seja, a matriz tem característica $r = 3$.

As primeiras $r = 3$ colunas da matriz U formam uma base ortonormada de $\text{col}(A)$. Fazendo

```
>> rref(U(:,1:3)')
ans =
    1.0000         0         0   -0.0000
         0    1.0000         0    6.0000
         0         0    1.0000    3.0000
```

confirmamos que as primeiras 3 colunas de U são versores de $u_1 = [1 \ 0 \ 0 \ 0]^T$, $u_2 = [0 \ 1 \ 0 \ 6]^T$ e $u_3 = [0 \ 0 \ 1 \ 3]^T$.

A 4ª coluna de U forma uma base ortonormada de $\text{Ker}(A^T)$. Fazendo

```
>> U(:,4)
ans =
    0.0000   -0.8847   -0.4423    0.1474
>> U(:,4)'/U(4,4)
ans =
    0.0000   -6.0000   -3.0000    1.0000
```

confirmamos que a 4ª coluna de U é o versor de $u_4 = [0 \ -6 \ -3 \ 1]^T$.

As primeiras $r = 3$ colunas da matriz V formam uma base ortonormada de $\text{lin}(A)$. Fazendo

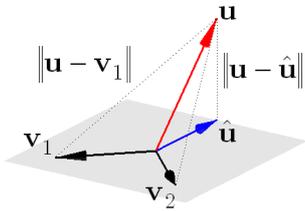
```
>> rref(V(:,1:3)')
ans =
    1.0000         0    1.0000         0   -2.0000
         0    1.0000    1.0000         0    3.0000
         0         0         0    1.0000    4.0000
```

confirmamos que as primeiras 3 colunas de V são versores de $v_1 = [1 \ 0 \ 1 \ 0 \ -2]^T$, $v_2 = [0 \ 1 \ 1 \ 0 \ 3]^T$ e $v_3 = [0 \ 0 \ 0 \ 1 \ 4]^T$.

Finalmente, poderíamos verificar que as 4ª e 5ª colunas de V forma uma base ortonormada de $\text{Ker}(A)$.



Melhor Aproximação.



Recordemos do módulo 5 que se define a distância entre dois vectores \mathbf{u} e $\mathbf{v} \in \mathbb{R}^n$ como a norma do vector entre eles

$$\text{dist}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \sqrt{\sum_{i=1}^n (u_i - v_i)^2}$$

5. Sendo \mathbb{W} um subespaço de \mathbb{R}^n e $\mathbf{u} \in \mathbb{R}^n$, designamos por **melhor aproximação** de \mathbf{u} em \mathbb{W} o vector $\hat{\mathbf{u}} \in \mathbb{W}$ tal que a distância de \mathbf{u} a $\hat{\mathbf{u}}$ é menor do que a distância de \mathbf{u} a qualquer outro vector de \mathbb{W} ,

$$\|\mathbf{u} - \hat{\mathbf{u}}\| < \|\mathbf{u} - \mathbf{v}\| \quad , \mathbf{v} \in \mathbb{R}^n$$

, correspondendo $\hat{\mathbf{u}}$ à projecção ortogonal de \mathbf{u} no subespaço \mathbb{W}

$$\hat{\mathbf{u}} = \text{proj}_{\mathbb{W}} \mathbf{u}$$

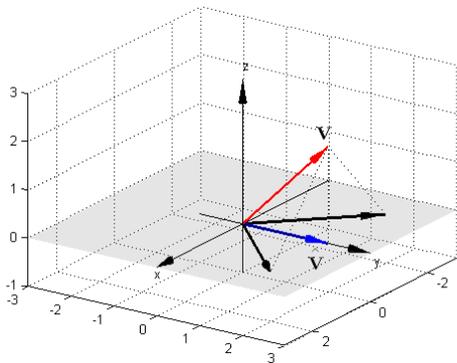
ou seja, sendo $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r$ uma base ortogonal do subespaço \mathbb{W} , temos

$$\hat{\mathbf{u}} = \text{proj}_{\mathbb{W}} \mathbf{u} = \sum_{i=1}^r \text{proj}_{\mathbf{u}_i} \mathbf{u} = \sum_{i=1}^r \frac{\mathbf{u} \cdot \mathbf{u}_i}{\mathbf{u}_i \cdot \mathbf{u}_i} \mathbf{u}_i = \sum_{i=1}^r k_i \mathbf{u}_i$$



Exemplo 6.

Dados os vectores $\mathbf{u}_1 = [2 \ 2 \ 0]^T$, $\mathbf{u}_2 = [-2 \ 2 \ 0]^T$ e $\mathbf{v} = [0 \ 2 \ 2]^T$, a melhor aproximação ao vector \mathbf{v} no espaço, \mathbb{W} , gerado por \mathbf{u}_1 e \mathbf{u}_2 , e dado que \mathbf{u}_1 e \mathbf{u}_2 são ortogonais, $\mathbf{u}_1 \cdot \mathbf{u}_2 = 0$, é



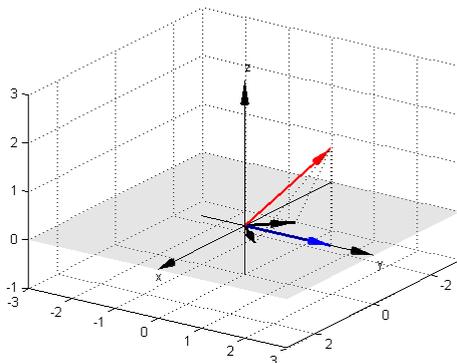
$$\begin{aligned} \hat{\mathbf{v}} &= \text{proj}_{\mathbb{W}} \mathbf{v} \\ &= \text{proj}_{\mathbf{u}_1} \mathbf{v} + \text{proj}_{\mathbf{u}_2} \mathbf{v} \\ &= \frac{\mathbf{v} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \frac{\mathbf{v} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2} \mathbf{u}_2 \\ &= \frac{1}{2} \mathbf{u}_1 + \frac{1}{2} \mathbf{u}_2 \\ &= [0 \ 2 \ 0]^T \end{aligned}$$

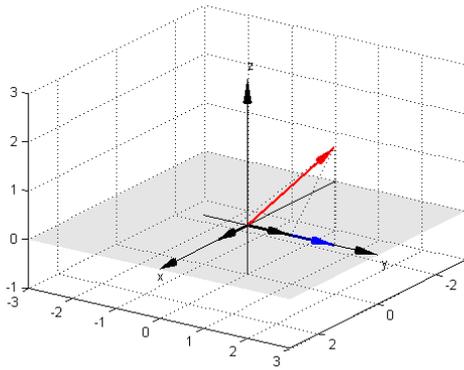
Se considerássemos os versores de \mathbf{u}_1 e \mathbf{u}_2 , que são geradores do mesmo espaço,

$$\begin{aligned} \mathbf{q}_1 &= \mathbf{u}_1 / \|\mathbf{u}_1\| = [1/\sqrt{2} \ 1/\sqrt{2} \ 0]^T \\ \mathbf{q}_2 &= \mathbf{u}_2 / \|\mathbf{u}_2\| = [-1/\sqrt{2} \ 1/\sqrt{2} \ 0]^T \end{aligned}$$

teríamos

$$\begin{aligned} \hat{\mathbf{v}} &= \text{proj}_{\mathbb{W}} \mathbf{v} \\ &= \text{proj}_{\mathbf{q}_1} \mathbf{v} + \text{proj}_{\mathbf{q}_2} \mathbf{v} \\ &= (\mathbf{v} \cdot \mathbf{q}_1) \mathbf{q}_1 + (\mathbf{v} \cdot \mathbf{q}_2) \mathbf{q}_2 \\ &= \sqrt{2} \mathbf{q}_1 + \sqrt{2} \mathbf{q}_2 \\ &= [0 \ 2 \ 0]^T \end{aligned}$$





Aliás, reconhecendo que u_1 e u_2 geram o subespaço de \mathbb{R}^3 correspondente ao plano xy , podemos tomar os versores canónicos e_1 e e_2 para base ortogonal do subespaço, resultando

$$\begin{aligned} \hat{v} &= \text{proj}_{\mathbb{W}} v \\ &= \text{proj}_{e_1} v + \text{proj}_{e_2} v \\ &= (v \cdot e_1) e_1 + (v \cdot e_2) e_2 \\ &= 2 e_2 \end{aligned}$$

, ou seja, $\hat{v} = 0 e_1 + 2 e_2 + 0 e_3$, e por isso escrevemos $\hat{v} = [0 \ 2 \ 0]^T$.

Embora a representação dependa da base ortogonal considerada, a projecção de um vector num subespaço não depende da base escolhida, correspondendo sempre ao mesmo vector

$$\hat{v} = \text{proj}_{\mathbb{W}} v = \frac{1}{2} u_1 + \frac{1}{2} u_2 = \sqrt{2} q_1 + \sqrt{2} q_2 = 2 e_2 = [0 \ 2 \ 0]^T$$

A distância do vector v aos vectores u_1 , u_2 , e \hat{v} é

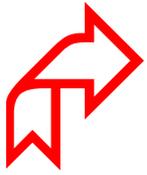
$$\begin{aligned} \text{dist}(u_1, v) &= \|u_1 - v\| = \sqrt{(u_{11} - v_1)^2 + (u_{12} - v_2)^2 + (u_{13} - v_3)^2} \\ &= \sqrt{(2 - 0)^2 + (2 - 2)^2 + (0 - 2)^2} \\ &= \sqrt{8} \end{aligned}$$

$$\begin{aligned} \text{dist}(u_2, v) &= \|u_2 - v\| = \sqrt{(u_{21} - v_1)^2 + (u_{22} - v_2)^2 + (u_{23} - v_3)^2} \\ &= \sqrt{(-2 - 0)^2 + (2 - 2)^2 + (0 - 2)^2} \\ &= \sqrt{8} \end{aligned}$$

$$\begin{aligned} \text{dist}(\hat{v}, v) &= \|\hat{v} - v\| = \sqrt{(\hat{v}_1 - v_1)^2 + (\hat{v}_2 - v_2)^2 + (\hat{v}_3 - v_3)^2} \\ &= \sqrt{(0 - 0)^2 + (2 - 2)^2 + (0 - 2)^2} \\ &= 2 \end{aligned}$$

A distância de v a \hat{v} é menor do que a distância a u_1 , u_2 , e menor que a distância a qualquer vector u pertencente ao subespaço gerado por u_1 , u_2 , \mathbb{W} , e por isso se diz que \hat{v} é a **melhor aproximação** de v em \mathbb{W}

$$\|v - \hat{v}\| < \|v - u\| \quad , u \in \mathbb{W}$$



Vector de Erro. Mínimo Erro Quadrático.

6. Sendo W um subespaço de \mathbb{R}^n , e $\hat{u} \in W$ a melhor aproximação de um vector $u \in \mathbb{R}^n$ no subespaço W , o vector

$$e = u - \hat{u}$$

designa-se por **vector de erro**, e pertence ao complemento ortogonal do subespaço W , W^\perp , sendo a sua norma,

$$\|e\| = \|u - \hat{u}\|$$

, designada por **mínimo erro quadrático**.



Exemplo 7.

Retomemos o exemplo anterior. Como vimos, a melhor aproximação do vector $v = [0 \ 2 \ 2]^T$ no subespaço W de \mathbb{R}^3 gerado pelos vectores $u_1 = [2 \ 2 \ 0]^T$ e $u_2 = [-2 \ 2 \ 0]^T$ é o vector

$$\hat{v} = \text{proj}_W v = [0 \ 2 \ 0]^T$$

O vector de erro é

$$\begin{aligned} e &= v - \hat{v} \\ &= [0 \ 2 \ 2]^T - [0 \ 2 \ 0]^T \\ &= [0 \ 0 \ 2]^T \\ &= 2e_3 \end{aligned}$$

, e pertence ao complemento ortogonal de W ,

$$\begin{aligned} e &= v - \hat{v} \\ &= v - \text{proj}_W v \\ &= \text{proj}_{W^\perp} v \end{aligned}$$

, como de imediato se reconhece, dado que W corresponde ao plano xy .

O vector \hat{v} corresponde à melhor representação que se pode fazer de v no subespaço W . Sendo

$$v = \hat{v} + e$$

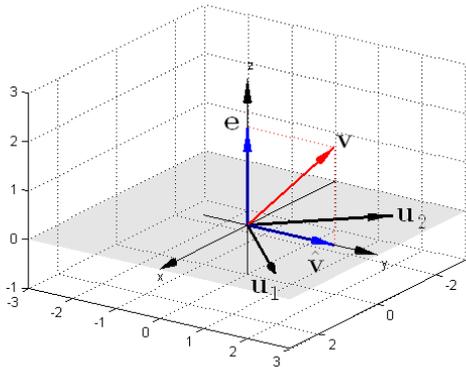
Se dissermos

$$v = \hat{v}$$

estamos a cometer um erro, o menor erro que é possível cometer, no sentido em que \hat{v} é, de todos os vectores de W , aquele que está à menor distância possível de v , sendo a norma do erro cometido

$$\begin{aligned} \|e\| &= \text{dist}(v, \hat{v}) = \|v - \hat{v}\| \\ &= \sqrt{(v_1 - \hat{v}_1)^2 + (v_2 - \hat{v}_2)^2 + (v_3 - \hat{v}_3)^2} \\ &= \sqrt{e_1^2 + e_2^2 + e_3^2} \\ &= 2 \end{aligned}$$

designada por mínimo erro quadrático. A designação resulta, como pode ver pela expressão acima, do facto de estar associado a uma soma de quadrados.



Consideremos o erro cometido ao aproximar \mathbf{v} por um qualquer dos vectores $\mathbf{u}_i \in \mathbb{W}$

$$\|\mathbf{e}\| = \text{dist}(\mathbf{v}, \mathbf{u}_i) = \|\mathbf{v} - \mathbf{u}_i\|$$

O erro depende do vector \mathbf{u}_i considerado, ou seja, é função de \mathbf{u}_i , por isso se diz ser uma **função de erro**, e, em particular, a **função erro quadrático**

$$\begin{aligned} \|\mathbf{e}\| &= \sqrt{(v_1 - u_{i1})^2 + (v_2 - u_{i2})^2 + (v_3 - u_{i3})^2} \\ &= \sqrt{e_1^2 + e_2^2 + e_3^2} \\ &= \sqrt{\sum_k e_k^2} \end{aligned}$$

, ou, de modo equivalente

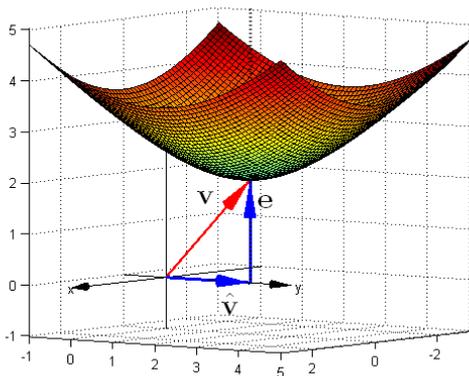
$$\|\mathbf{e}\|^2 = \sum_k e_k^2$$

, em que e_k corresponde a cada uma das coordenadas do vector erro

$$e_k = v_k - u_{ik}$$

Vimos, no exemplo anterior, que, para o vector $\mathbf{u}_1 = [2 \ 2 \ 0]^T$ resultou

$$\|\mathbf{e}\|^2 = \|\mathbf{v} - \mathbf{u}_1\|^2 = (v_1 - u_{11})^2 + (v_2 - u_{12})^2 + (v_3 - u_{13})^2 = 8$$



Se calculássemos o erro quadrático para diversos vector do subespaço \mathbb{W} , como vimos, coincidente com o plano \mathbf{xy} , obteríamos um conjunto de valores que se representa na figura.

Para facilidade de compreensão sobrepôs-se à função de erro, $\|\mathbf{e}\|$, a representa dos vectores \mathbf{v} , $\hat{\mathbf{v}}$ e \mathbf{e} .

Para todos os pontos do plano \mathbf{xy} , a distância da superfície representada ao plano é igual ao valor da função de erro.

A função de erro é parabolóide com o mínimo no ponto $(2, 2)$, as coordenadas de $\hat{\mathbf{v}}$ no plano \mathbf{xy} .



Solução de Mínimo Erro Quadrático. Equações Normais.

7. Dado o sistema $Ax = b$, sendo A uma matriz $m \times n$, e $b \in \mathbb{R}^m$, designa-se por **solução de quadrados mínimos**, ou **solução de mínimo erro quadrático**, o vector $\hat{x} \in \mathbb{R}^n$ tal que

$$\|b - A\hat{x}\| \leq \|b - Ax\| \quad , x \in \mathbb{R}^n$$

8. O vector \hat{x} é uma solução de mínimo erro quadrático do sistema $Ax = b$ sse for solução do sistema de equações

$$A^T A \hat{x} = A^T b$$

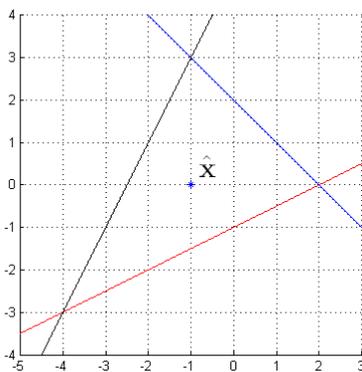
designadas por **equações normais** de \hat{x} .

9. A matriz A tem colunas linearmente independentes sse a matriz $A^T A$ for invertível, e, nesse caso, a solução de mínimo erro quadrático do sistema $Ax = b$ é única, sendo

$$\hat{x} = (A^T A)^{-1} A^T b$$



Exemplo 8.



O sistema de equações

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

, com forma matricial, $Ax = b$,

$$\begin{bmatrix} 1 & -2 \\ 1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ -5 \end{bmatrix}$$

é um sistema impossível

$$\begin{bmatrix} 1 & -2 & 2 \\ 1 & 1 & 2 \\ 2 & -1 & -5 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



```
>> A=[1 -2; 1 1; 2 -1];
>> b=[2 2 -5]';
>> rref([A b])
ans =
     1     0     0
     0     1     0
     0     0     1
```

A figura mostra as rectas correspondentes a cada uma das equações do sistema. Para que o sistema fosse possível e determinado, as rectas de que cada uma das equações do sistema é representativa deveriam cruzar-se num só ponto, a solução do sistema.

Por outras palavras, o vector b não pertence ao espaço coluna da matriz A , não sendo por isso possível determinar um vector x (objecto) cujo transformado (imagem) Ax pertença a $col(A)$.

A solução de mínimos quadrados do sistema, \hat{x} , assinalada na figura, minimiza a norma

$$\|b - Ax\|$$

, ou seja, corresponde ao vector \hat{x} cujo transformado, $A\hat{x}$, está a uma distância de b menor do que qualquer outro vector do espaço coluna da matriz.

Como a matriz A tem colunas linearmente independentes, a matriz

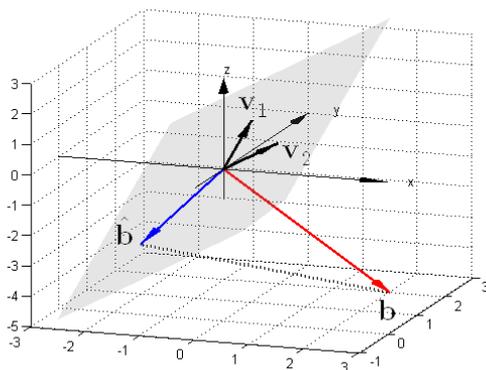
$$A^T A = \begin{bmatrix} 6 & -3 \\ -3 & 6 \end{bmatrix}$$

é invertível

$$(A^T A)^{-1} = \frac{1}{9} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

pelo que a solução de mínimo erro quadrático do sistema $Ax = b$ é única, sendo

$$\begin{aligned} \hat{x} &= (A^T A)^{-1} A^T b \\ &= \frac{1}{9} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 \\ -2 & 1 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ -5 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 0 & 1 & 1 \\ -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ -5 \end{bmatrix} \\ &= \begin{bmatrix} -1 \\ 0 \end{bmatrix} \end{aligned}$$



, sendo a imagem de \hat{x} no espaço coluna de A , ou seja, a projecção ortogonal de b no espaço coluna de A .

$$\hat{b} = A\hat{x} = \begin{bmatrix} -1 \\ -1 \\ -2 \end{bmatrix}$$

A figura mostra dois vectores, v_1 e v_2 , que constituem uma base do espaço coluna da matriz A , o espaço por eles gerado, o vector b , e a sua projecção no espaço coluna de A , \hat{b} .

Recorrendo ao MatLab, o cálculo da solução de mínimo erro quadrático, e respectiva imagem, é trivial



```
>> A=[1 -2; 1 1; 2 -1];
>> b=[2 2 -5]';
>> xs=inv(A'*A)*A'*b
xs =
    -1
     0
>> bs=A*xs
bs =
    -1
    -1
    -2
```

Exemplo 9.

Dados os pontos de \mathbb{R}^2 , $(1,2)$, $(3,2)$, e $(4,5)$, podemos determinar os parâmetros m e b da recta $y = mx + b$ (sendo m o declive da recta e $(0,b)$ o ponto em que a recta intercepta o eixo dos **yy**) que melhor se adapta ao conjunto dos ponto segundo um critério de minimização do erro quadrático cometido.

O sistema de equações

$$\begin{cases} y_1 = mx_1 + b \\ y_2 = mx_2 + b \\ y_3 = mx_3 + b \end{cases}$$

ou seja, na forma matricial,

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

é possível e determinado se os pontos forem colineares. No caso presente

$$\begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 5 \end{bmatrix}$$

, não sendo os pontos colineares, o sistema é impossível

$$\begin{bmatrix} 1 & 1 & 2 \\ 1 & 3 & 2 \\ 1 & 4 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



```
>> A=[1 1;1 3; 1 4];
>> b=[2 2 5]';
>> rref([A b])
     1     0     0
     0     1     0
     0     0     1
```

Vamos procurar a solução do sistema que minimiza o erro quadrático cometido na determinação dos parâmetros m e b , calculando $\hat{\mathbf{p}} = \begin{bmatrix} \hat{b} \\ \hat{m} \end{bmatrix}$. Temos

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 3 & 8 \\ 8 & 26 \end{bmatrix}$$

e

$$(\mathbf{A}^T \mathbf{A})^{-1} = \begin{bmatrix} 3 & 8 \\ 8 & 26 \end{bmatrix}^{-1} = \frac{1}{14} \begin{bmatrix} 26 & -8 \\ -8 & 6 \end{bmatrix}$$

pelo que

$$\hat{\mathbf{p}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}$$

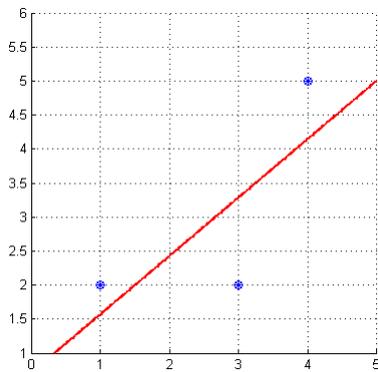
$$= \frac{1}{14} \begin{bmatrix} 26 & -8 \\ -8 & 6 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 5 \end{bmatrix} = \frac{1}{14} \begin{bmatrix} 18 & 2 & -6 \\ -5 & 1 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 5 \end{bmatrix}$$

$$\begin{bmatrix} \hat{b} \\ \hat{m} \end{bmatrix} = \begin{bmatrix} 5/7 \\ 6/7 \end{bmatrix}$$



```
>> xs=inv(A'*A)*A'*b
```

```
xs =  
  
5/7  
6/7
```



Temos então $b = 5/7$ e $m = 6/7$, ou seja, a recta que melhor se adapta ao conjunto de pontos é

$$y = \frac{6}{7}x + \frac{5}{7}$$

A figura mostra o conjunto de pontos e a recta que melhor se lhes adapta, no sentido da minimização do erro quadrático, ou seja, o somatório do quadrado das distância entre a recta e os pontos dados é menor do que para qualquer outra recta que pudéssemos definir.

Para evitar a inversão da matriz $\mathbf{A}^T \mathbf{A}$ no cálculo de

$$\hat{\mathbf{p}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}$$

podemos resolver o sistema de equações normais

$$\mathbf{A}^T \mathbf{A} \hat{\mathbf{p}} = \mathbf{A}^T \mathbf{y}$$

Teríamos então

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 3 & 8 \\ 8 & 26 \end{bmatrix}$$

e

$$\mathbf{A}^T \mathbf{y} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 5 \end{bmatrix} = \begin{bmatrix} 9 \\ 28 \end{bmatrix}$$

de onde resulta o sistema

$$\begin{bmatrix} 3 & 8 \\ 8 & 26 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{m} \end{bmatrix} = \begin{bmatrix} 9 \\ 28 \end{bmatrix}$$

, que facilmente se resolve,

$$\begin{bmatrix} 3 & 8 & 9 \\ 8 & 26 & 28 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 5/7 \\ 0 & 1 & 5/7 \end{bmatrix}$$

, e de onde resulta a solução encontrada anteriormente

$$\begin{bmatrix} \hat{b} \\ \hat{m} \end{bmatrix} = \begin{bmatrix} 5/7 \\ 6/7 \end{bmatrix}$$

Recorrendo ao MatLab temos



```
>> A=[1 1;1 3; 1 4];
>> y=[2 2 5]';
>> rref([A'*A A'*y])
ans =
      1      0      5/7
      0      1      6/7
```

Podemos ainda ter em atenção que

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 3 & 8 \\ 8 & 26 \end{bmatrix} = \begin{bmatrix} N & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{bmatrix}$$

, em que N é o número de pontos, e

$$\mathbf{A}^T \mathbf{y} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 5 \end{bmatrix} = \begin{bmatrix} 9 \\ 28 \end{bmatrix} = \begin{bmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{bmatrix}$$

, pelo que podemos escrever as equações normais

$$\mathbf{A}^T \mathbf{A} \hat{\mathbf{p}} = \mathbf{A}^T \mathbf{y}$$

na forma

$$\begin{bmatrix} N & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{bmatrix} \hat{\mathbf{p}} = \begin{bmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{bmatrix}$$



Quadrados Mínimos e Factorização QR

10. Sendo \mathbf{A} uma matriz $m \times n$, $\mathbf{b} \in \mathbb{R}^m$, e \mathbf{Q} e \mathbf{R} as matrizes resultantes da factorização **QR** de \mathbf{A} , então, a solução de mínimo erro quadrático, $\hat{\mathbf{x}}$, do sistema $\mathbf{Ax} = \mathbf{b}$ pode ser calculada por resolução do sistema

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

ou seja

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



Exemplo 10.

O cálculo da solução de mínimo de erro quadrático com base na factorização **QR**, $\mathbf{R}\hat{\mathbf{x}} = \mathbf{Q}^T \mathbf{b}$, em alternativa à resolução das equações normais, $\mathbf{A}^T \mathbf{A} \hat{\mathbf{x}} = \mathbf{A}^T \mathbf{b}$, é justificada com base em considerandos associados ao erros numéricos cometidos durante o cálculo, e será objecto de análise na cadeira de Métodos Numéricos e Optimização. Em problemas de pequena dimensão, como os que temos nos têm servido de exemplo, o recurso à factorização **QR** é uma abordagem mais complexa.

Apenas a título de exemplo didáctico, podemos analisar o exemplo 9 recorrendo à factorização **QR**. Recordemos que tínhamos o sistema de equações na forma matricial, $\mathbf{Ax} = \mathbf{b}$,

$$\begin{bmatrix} 1 & -2 \\ 1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ -5 \end{bmatrix}$$

e foi encontrada a solução de mínimos quadrados

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

Dada a matriz

$$\mathbf{A} = \begin{bmatrix} 1 & -2 \\ 1 & 1 \\ 2 & -1 \end{bmatrix}$$

, recorreremos à ortogonalização de Gram-Schmidt para obter uma base ortonormada para o subespaço gerado pelas colunas da matriz.

Temos então

$$1. \mathbf{v}_1 = \mathbf{u}_1 = [1 \ 1 \ 2]^T$$

$$2. \mathbf{v}_2 = \mathbf{u}_2 - \text{proj}_{\mathbf{v}_1} \mathbf{u}_2 = \mathbf{u}_2 - \frac{\mathbf{u}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 = \begin{bmatrix} -3/2 \\ 3/2 \\ 0 \end{bmatrix}$$

, e, normalizando os vectores,

$$3. \mathbf{q}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \begin{bmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix}, \quad \mathbf{q}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \begin{bmatrix} -\sqrt{2} \\ \sqrt{2} \\ 0 \end{bmatrix}$$

Assim sendo, temos

$$Q = [q_1 \quad q_2] = \begin{bmatrix} 1/\sqrt{6} & -\sqrt{2} \\ 1/\sqrt{6} & \sqrt{2} \\ 2/\sqrt{6} & 0 \end{bmatrix}$$

e

$$R = Q^T A = \begin{bmatrix} 1/\sqrt{6} & 1/\sqrt{6} & 2/\sqrt{6} \\ -\sqrt{2} & \sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ 1 & 1 \\ 2 & -1 \end{bmatrix} = \begin{bmatrix} \sqrt{6} & -\sqrt{6}/2 \\ 0 & 3\sqrt{2}/2 \end{bmatrix}$$

, pelo que $R\hat{x} = Q^T b$ resulta

$$\begin{bmatrix} \sqrt{6} & -\sqrt{6}/2 \\ 0 & 3\sqrt{2}/2 \end{bmatrix} \hat{x} = \begin{bmatrix} 1/\sqrt{6} & 1/\sqrt{6} & 2/\sqrt{6} \\ -\sqrt{2} & \sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ -5 \end{bmatrix}$$

$$\begin{bmatrix} \sqrt{6} & -\sqrt{6}/2 \\ 0 & 3\sqrt{2}/2 \end{bmatrix} \hat{x} = \begin{bmatrix} \sqrt{6} \\ 0 \end{bmatrix}$$

e finalmente, resolvendo o sistema

$$\begin{bmatrix} \sqrt{6} & -\sqrt{6}/2 & \sqrt{6} \\ 0 & 3\sqrt{2}/2 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

encontramos a mesma solução do exemplo 9

$$\hat{x} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

Recorrendo ao MatLab



```
>> u1=[1 1 2]';
>> u2=[-2 1 -1]';
>> A=[u1 u2];
>> b=[2 2 -5]';
>> v1=u1;
>> v2=u2-(u2'*v1)/(v1'*v1)*v1;
>> q1=v1/norm(v1);
>> q2=v2/norm(v2);
>> Q=[q1 q2];
>> R=Q'*A;
>> rref([R Q'*b])
ans =
    1.0000         0   -1.0000
         0    1.0000         0
```



Quadrados Mínimos e Projecção num Subespaço.

11. Sendo \mathbb{W} um subespaço de \mathbb{R}^m e \mathbf{A} uma matriz $m \times n$ cujas colunas formam uma base de \mathbb{W} , a projecção ortogonal de um qualquer vector $\mathbf{u} \in \mathbb{R}^m$ no subespaço \mathbb{W} é dada por

$$\text{proj}_{\mathbb{W}} \mathbf{u} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{u}$$

, ou seja, a matriz da transformação $T : \mathbb{R}^m \rightarrow \mathbb{R}^m$ que projecta ortogonalmente \mathbb{R}^m em \mathbb{W} é

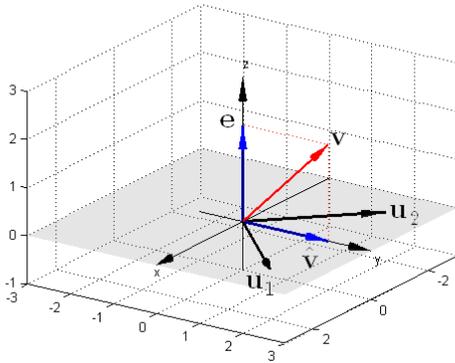
$$\mathbf{T} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$$



Exemplo 11.

Retomemos o problema 7. Como vimos, a melhor aproximação do vector $\mathbf{v} = [0 \ 2 \ 2]^T$ no subespaço \mathbb{W} de \mathbb{R}^3 gerado pelos vectores $\mathbf{u}_1 = [2 \ 2 \ 0]^T$ e $\mathbf{u}_2 = [-2 \ 2 \ 0]^T$ é o vector

$$\hat{\mathbf{v}} = \text{proj}_{\mathbb{W}} \mathbf{v} = [0 \ 2 \ 0]^T$$



Atendendo a que a matriz da transformação $T : \mathbb{R}^m \rightarrow \mathbb{R}^m$ que projecta ortogonalmente \mathbb{R}^m em \mathbb{W} é

$$\mathbf{T} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$$

, sendo

$$\mathbf{A} = [\mathbf{u}_1 \ \mathbf{u}_2] = \begin{bmatrix} 2 & -2 \\ 2 & 2 \\ 0 & 0 \end{bmatrix}$$

, temos

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} 2 & 2 & 0 \\ -2 & 2 & 0 \end{bmatrix} \begin{bmatrix} 2 & -2 \\ 2 & 2 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 8 & 0 \\ 0 & 8 \end{bmatrix}$$

$$(\mathbf{A}^T \mathbf{A})^{-1} = \begin{bmatrix} 8 & 0 \\ 0 & 8 \end{bmatrix}^{-1} = \frac{1}{8} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} = \begin{bmatrix} 2 & -2 \\ 2 & 2 \\ 0 & 0 \end{bmatrix} \frac{1}{8} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \frac{1}{8} \begin{bmatrix} 2 & -2 \\ 2 & 2 \\ 0 & 0 \end{bmatrix}$$

, e finalmente

$$\mathbf{T} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T = \frac{1}{8} \begin{bmatrix} 2 & -2 \\ 2 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 2 & 0 \\ -2 & 2 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Como tínhamos concluído nos exercício 7 e 8, os vectores \mathbf{u}_1 e \mathbf{u}_2 geram o subespaço \mathbb{W} de \mathbb{R}^3 correspondente ao plano \mathbf{xy} .

Recorde do módulo 5, que a matriz de projecção no plano **xy** corresponde à matriz que acima foi determinada

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

A matriz é única, sendo a matriz associada à transformação, não dependendo dos vectores considerados para gerar o subespaço.

Consideremos, por exemplo, os vectores $\mathbf{u}_3 = [2 \ 2 \ 0]^T$ e $\mathbf{u}_4 = [1 \ 2 \ 0]^T$, facilmente se reconhece que os vectores geram o subespaço \mathbb{W} de \mathbb{R}^3 correspondente ao plano **xy** (basta notar que os vectores existem sobre o plano e têm diferentes direcções). No entanto, ao contrário de \mathbf{u}_1 e \mathbf{u}_2 , \mathbf{u}_3 e \mathbf{u}_4 não são ortogonais,

$$\mathbf{u}_3 \cdot \mathbf{u}_4 = 6$$

Independentemente deste facto, temos

$$\mathbf{A} = [\mathbf{u}_3 \ \mathbf{u}_4] = \begin{bmatrix} 2 & 1 \\ 2 & 2 \\ 0 & 0 \end{bmatrix}$$

, e a matriz de projecção no espaço por eles gerado

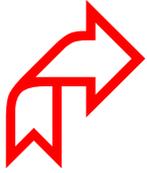
$$\mathbf{T} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$



```
>> u1=[2 2 0]';
>> u2=[1 2 0]';
>> A=[u1 u2];
T=A*inv(A'*A)*A'
T =
     1     0     0
     0     1     0
     0     0     0
```

Consideremos agora que o conjunto de vectores não é linearmente independente, por exemplo, os vectores $\mathbf{u}_5 = [2 \ 2 \ 0]^T$, $\mathbf{u}_6 = [1 \ 2 \ 0]^T$, e $\mathbf{u}_7 = [3 \ 4 \ 0]^T$. Facilmente se reconhece que $\mathbf{u}_7 = \mathbf{u}_5 + \mathbf{u}_6$. A matriz \mathbf{A} tem colunas linearmente dependentes, de onde resulta que $\mathbf{A}^T \mathbf{A}$ é singular, pelo que não podemos calcular $(\mathbf{A}^T \mathbf{A})^{-1}$.

Podemos contornar o problema, procurando uma base para o espaço gerado por \mathbf{u}_5 , \mathbf{u}_6 , e \mathbf{u}_7 (o que no presente exemplo é obviamente simples, basta ignorar \mathbf{u}_7). Veremos na próxima secção como abordar a questão sem necessidade de procurar uma base do subespaço.



Matriz Pseudo-Inversa.

12. Sendo A uma matriz $m \times n$ com colunas linearmente independentes, designa-se por **matriz pseudo-inversa** de A a matriz $n \times m$

$$A^+ = (A^T A)^{-1} A^T$$

13. A matriz pseudo-inversa de qualquer matriz A (mesmo não sendo $A^T A$ invertível) pode ser calculada a partir da decomposição em valores singulares $A = U \Sigma V^T$, sendo

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

em que Σ^+ é a matriz

$$\Sigma^+ = \begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix}$$

14. Propriedades da matriz pseudo inversa

$$A A^+ A = A$$

$$A^+ A A^+ = A^+$$

$A A^+$ e $A^+ A$ são matrizes simétricas.



Exemplo 12.

Com base na definição da matriz pseudo-inversa podemos reformular os exemplos das secções anteriores.

Podemos dizer, por exemplo, que, sendo A a matriz cujas colunas formam uma base de \mathbb{W} , a matriz da transformação $T : \mathbb{R}^m \rightarrow \mathbb{R}^m$ que projecta ortogonalmente \mathbb{R}^m em \mathbb{W} é

$$T = A A^+$$

, e ainda que, a solução de mínimo erro quadrático do sistema $Ax = b$ é

$$\hat{x} = A^+ b$$

Note que, se A fosse regular poderíamos escrever

$$x = A^{-1} b$$

Independentemente da formulação mais sintética, não deixa de estar em causa o cálculo de $A^+ = (A^T A)^{-1} A^T$, e portanto o problema da inversão da matriz $A^T A$, no entanto, recorrendo ao conceito de decomposição em valores singulares, podemos resolver a questão.

Exemplo 13.

Retomemos o conjunto de vectores $u_5 = [2 \ 2 \ 0]^T$, $u_6 = [1 \ 2 \ 0]^T$, e $u_7 = [3 \ 4 \ 0]^T$, o vector $v = [0 \ 2 \ 2]^T$, e procuremos a melhor aproximação do vector v no subespaço \mathbb{W} de \mathbb{R}^3 gerado pelos vectores u_5 , u_6 , e u_7 , que, como sabemos, corresponde à projecção ortogonal de v em \mathbb{W} .

Temos a matriz

$$A = [u_5 \ u_6 \ u_7] = \begin{bmatrix} 2 & 1 & 3 \\ 2 & 2 & 4 \\ 0 & 0 & 0 \end{bmatrix}$$

, e podemos verificar que a matriz

$$A^T A = \begin{bmatrix} 2 & 2 & 0 \\ 1 & 2 & 0 \\ 3 & 4 & 0 \end{bmatrix} \begin{bmatrix} 2 & 1 & 3 \\ 2 & 2 & 4 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 8 & 6 & 14 \\ 6 & 5 & 11 \\ 14 & 11 & 25 \end{bmatrix}$$

não é invertível

$$\begin{bmatrix} 8 & 6 & 14 & 1 & 0 & 0 \\ 6 & 5 & 11 & 0 & 1 & 0 \\ 14 & 11 & 25 & 0 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 & 0 & -11/4 & 5/4 \\ 0 & 1 & 1 & 0 & 7/2 & -3/2 \\ 0 & 0 & 0 & 1 & 1 & -1 \end{bmatrix}$$



```
>> u1=[2 2 0]';
>> u2=[1 2 0]';
>> u3=[3 4 0]';
>> A=[u1 u2 u3];
>> inv(A'*A)
Warning: Matrix is close to singular or badly scaled.
Results may be inaccurate. RCOND = 1.184238e-017.
ans =
      *          *          *
      *          *          *
      *          *          *
```

No entanto, a matriz pseudo-inversa de qualquer matriz A , mesmo não sendo $A^T A$ invertível, pode ser calculada a partir da decomposição em valores singulares da matriz A , sendo

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

em que Σ^+ é

$$\Sigma^+ = \begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix}$$

, sendo D a matriz dos valores singulares de A .

Dada a morosidade dos cálculos, vamos recorrer ao MatLab para decompor A em valores singulares.



```
>> [U S V]=svd(A);
>> S(1:2,1:2)=inv(S(1:2,1:2))
>> Am=V*S*U'
Am =
    1.0000    -0.6667     0
   -1.0000     0.8333     0
   -0.0000     0.1667     0
```

O cálculo é dado meramente a título de exemplo, dado que existe em MatLab a função **pinv(A)** que calcula a pseudo-inversa de uma matriz A ,



```
>> pinv(A)
ans =
    1.0000   -0.6667    0
   -1.0000    0.8333    0
   -0.0000    0.1667    0
```

$$\mathbf{A}^+ = \begin{bmatrix} 1 & -2/3 & 0 \\ -1 & 5/6 & 0 \\ 0 & 1/6 & 0 \end{bmatrix}$$

Dada a pseudo-inversa, podemos agora calcular a melhor aproximação do vector $\mathbf{v} = [0 \ 2 \ 2]^T$ no subespaço \mathbb{W} de \mathbb{R}^3 gerado pelos vectores \mathbf{u}_5 , \mathbf{u}_6 , e \mathbf{u}_7 , que, como sabemos, corresponde à projecção ortogonal de \mathbf{v} em \mathbb{W} .

Temos

$$\begin{aligned} \hat{\mathbf{v}} &= \mathbf{A}\mathbf{A}^+\mathbf{v} = \begin{bmatrix} 2 & 1 & 3 \\ 2 & 2 & 4 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & -2/3 & 0 \\ -1 & 5/6 & 0 \\ 0 & 1/6 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} \end{aligned}$$

Identicamente ao determinado nos exemplos anteriores $\hat{\mathbf{v}} = [0 \ 2 \ 0]^T$.



Exercícios.

DECOMPOR UMA MATRIZ EM VALORES SINGULARES.

1. Dada a matriz rectangular, 3×2 ,

$$\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 0 & -2 \\ 1 & 1 \end{bmatrix}$$

temos

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} 2 & 0 & 1 \\ 0 & -2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & -2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 1 \\ 1 & 5 \end{bmatrix}$$

A matriz $\mathbf{A}^T \mathbf{A}$ tem polinómio característico

$$p(\lambda) = \det \begin{bmatrix} 5 - \lambda & 1 \\ 1 & 5 - \lambda \end{bmatrix} = (5 - \lambda)^2 - 1 = \lambda^2 - 10\lambda + 24$$

cujas raízes são

$$\lambda_1, \lambda_2 = \frac{10 \pm \sqrt{100 - 96}}{2} = \frac{10 \pm 2}{2} = 6, 4$$

tem 2 valores próprios, $\lambda_1 = 6$ e $\lambda_2 = 4$, pelo que a matriz \mathbf{A} tem valores singulares

$$\sigma_1 = \sqrt{\lambda_1} = \sqrt{6}$$

e

$$\sigma_2 = \sqrt{\lambda_2} = 2$$

Podemos de imediato escrever a matriz diagonal dos valores singulares

$$\mathbf{D} = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} = \begin{bmatrix} \sqrt{6} & 0 \\ 0 & 2 \end{bmatrix}$$

, e a matriz Σ , de dimensão $m \times n = 3 \times 2$,

$$\Sigma = \begin{bmatrix} \mathbf{D} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \sqrt{6} & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix}$$

A matriz $\mathbf{A}^T \mathbf{A}$ tem vectores próprios normalizados:

1. Associado a $\lambda_1 = 6$

$$\begin{bmatrix} 5 - 6 & 1 \\ 1 & 5 - 6 \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} \Rightarrow x = y \Rightarrow v'_1 = k \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow v_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

2. Associado a $\lambda_2 = 4$

$$\begin{bmatrix} 5 - 4 & 1 \\ 1 & 5 - 4 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \Rightarrow x = -y \Rightarrow v'_2 = k \begin{bmatrix} -1 \\ 1 \end{bmatrix} \Rightarrow v_2 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

V é uma matriz ortogonal da dimensão de $A^T A$, no caso presente, 2×2 , constituída por um conjunto ortonormado de vectores próprios da matriz $A^T A$. No caso presente temos

$$V = [v_1 \quad v_2] = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

U é uma matriz ortogonal $m \times m$, no caso presente 3×3 , cujos elementos são determinados por

$$u_i = \frac{1}{\sigma_i} A v_i$$

Em que σ_i são os valores singulares não nulos de A e v_i os valores próprios normalizados a eles associados, pelo que temos de imediato

$$u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 2 & 0 \\ 0 & -2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} \\ -1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$$

$$u_2 = \frac{1}{\sigma_2} A v_2 = \frac{1}{2} \begin{bmatrix} 2 & 0 \\ 0 & -2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} -1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{bmatrix}$$

, sendo necessário determinar um versor ortogonal a u_1 e u_2 , recorrendo à ortogonalização de Gram-Schmidt. Considerando à base $\{u_1, u_2, e_3\}$, temos

$$\begin{aligned} u'_3 &= e_3 - \text{proj}_{u_2} e_3 - \text{proj}_{u_1} e_3 \\ &= e_3 - \frac{e_3 \cdot u_2}{u_2 \cdot u_2} u_2 - \frac{e_3 \cdot u_1}{u_1 \cdot u_1} u_1 = e_3 - (e_3 \cdot u_2) u_2 - (e_3 \cdot u_1) u_1 \\ &= \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} -1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{bmatrix} \begin{bmatrix} -1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1/\sqrt{3} \\ -1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix} \begin{bmatrix} 1/\sqrt{3} \\ -1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - 1/\sqrt{3} \begin{bmatrix} 1/\sqrt{3} \\ -1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 1/3 \\ -1/3 \\ 1/3 \end{bmatrix} \\ &= \begin{bmatrix} -1/3 \\ 1/3 \\ 2/3 \end{bmatrix} \end{aligned}$$

e

$$u_3 = \frac{u'_3}{\|u'_3\|} = \begin{bmatrix} -\sqrt{2}/2\sqrt{3} \\ \sqrt{2}/2\sqrt{3} \\ \sqrt{2}/\sqrt{3} \end{bmatrix}$$

Temos então

$$U = [u_1 \quad u_2 \quad u_3] = \begin{bmatrix} 1/\sqrt{3} & -1/\sqrt{2} & -\sqrt{2}/2\sqrt{3} \\ -1/\sqrt{3} & -1/\sqrt{2} & \sqrt{2}/2\sqrt{3} \\ 1/\sqrt{3} & 0 & \sqrt{2}/\sqrt{3} \end{bmatrix}$$

Ficando assim determinada a decomposição em valores singulares da matriz A

$$A = U\Sigma V^T = \begin{bmatrix} 1/\sqrt{3} & -1/\sqrt{2} & -\sqrt{2}/2\sqrt{3} \\ -1/\sqrt{3} & -1/\sqrt{2} & \sqrt{2}/2\sqrt{3} \\ 1/\sqrt{3} & 0 & \sqrt{2}/\sqrt{3} \end{bmatrix} \begin{bmatrix} \sqrt{6} & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

, como podemos confirmar



```
>> x=sqrt(2); y=1/sqrt(3); z=1/sqrt(2) ;
>> U=[y -z -x*y/2;-y -z x*y/2;y 0 x*y];
>> S=[sqrt(6) 0; 0 2; 0 0];
>> V=[z -z; z z];
>> U*S*V'
ans =
    2.0000    0.0000
   -0.0000   -2.0000
    1.0000    1.0000
```

Podemos ainda decompor A na forma

$$A = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T$$

$$= \sqrt{6} \begin{bmatrix} 1/\sqrt{3} \\ -1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} + 2 \begin{bmatrix} -1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{bmatrix} \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & -1 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ 0 & 0 \end{bmatrix}$$

A decomposição em valores singulares da matriz A pode ser feita em MatLab utilizando a função `svd(A)`



```
>> A=[2 0; 0 -2; 1 1]
>> [U S V]=svd(A)
U =
   -0.5774   -0.7071   -0.4082
    0.5774   -0.7071    0.4082
   -0.5774    0.0000    0.8165
S =
    2.4495         0
         0    2.0000
         0         0
V =
   -0.7071   -0.7071
   -0.7071    0.7071
```

DADA UMA SVD DETERMINAR AS BASES DOS 4 ESPAÇOS FUNDAMENTAIS DA MATRIZ

2. A matriz \mathbf{A} , $m \times n = 3 \times 2$, tem $r = 2$ valores singulares não nulos, e decomposição em valores singulares

$$\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 0 & -2 \\ 1 & 1 \end{bmatrix} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \begin{bmatrix} 1/\sqrt{3} & -1/\sqrt{2} & -\sqrt{2}/2\sqrt{3} \\ -1/\sqrt{3} & -1/\sqrt{2} & \sqrt{2}/2\sqrt{3} \\ 1/\sqrt{3} & 0 & \sqrt{2}/\sqrt{3} \end{bmatrix} \begin{bmatrix} \sqrt{6} & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

Os vectores $\mathbf{u}_1 = [1/\sqrt{3} \quad -1/\sqrt{3} \quad 1/\sqrt{3}]^T$ e $\mathbf{u}_2 = [-1/\sqrt{2} \quad -1/\sqrt{2} \quad 0]^T$, correspondentes às primeiras $r = 2$ colunas da matriz \mathbf{U} formam uma base ortonormada do espaço coluna de \mathbf{A} .

Os vectores $\mathbf{v}_1 = [1/\sqrt{2} \quad 1/\sqrt{2}]^T$ e $\mathbf{v}_3 = [-1/\sqrt{2} \quad 1/\sqrt{2}]^T$, correspondentes às primeiras $r = 2$ colunas da matriz \mathbf{V} formam uma base ortonormada do espaço linha de \mathbf{A} .

Como $n - r = 0$, o núcleo da matriz \mathbf{A} é constituído apenas pelo vector nulo, $\mathbf{0} = [0 \quad 0]^T$.

Sendo $m - r = 1$, a 3ª coluna da matriz \mathbf{U} , $\mathbf{u}_3 = [-\sqrt{2}/2\sqrt{3} \quad \sqrt{2}/2\sqrt{3} \quad \sqrt{2}/\sqrt{3}]^T$, forma uma base ortonormada do núcleo de \mathbf{A}^T .

DETERMINAR A MELHOR APROXIMAÇÃO A UM VECTOR NUM SUBESPAÇO

3. Consideremos de novo a matriz

$$\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 0 & -2 \\ 1 & 1 \end{bmatrix}$$

Os vectores que constituem as colunas da matriz, $\mathbf{u}_1 = [2 \quad 0 \quad 1]^T$ e $\mathbf{u}_2 = [0 \quad -2 \quad 1]^T$, são linearmente independentes, basta recordar do exercício anterior que $\dim(\text{col}(\mathbf{A})) = 2$.

Dado o vector $\mathbf{v} = [0 \quad 1 \quad 1]^T$, a melhor aproximação ao vector \mathbf{v} no subespaço, \mathbb{W} , gerado por \mathbf{u}_1 e \mathbf{u}_2 , ou seja, a melhor aproximação ao vector \mathbf{v} no espaço coluna de \mathbf{A} (poderíamos verificar que \mathbf{v} não pertence ao espaço coluna, ou seja, o sistema $\mathbf{A}\mathbf{x} = \mathbf{v}$ é impossível), **não pode ser determinado a partir de $\text{proj}_{\mathbf{u}_1} \mathbf{v}$ e $\text{proj}_{\mathbf{u}_2} \mathbf{v}$ porque \mathbf{u}_1 e \mathbf{u}_2 não são ortogonais.**

$$\hat{\mathbf{v}} = \text{proj}_{\mathbb{W}} \mathbf{v} \neq \text{proj}_{\mathbf{u}_1} \mathbf{v} + \text{proj}_{\mathbf{u}_2} \mathbf{v}$$

É necessário encontrar uma base ortogonal do espaço gerado por \mathbf{u}_1 e \mathbf{u}_2 , por exemplo, recorrendo à ortogonalização de Gram-Schmidt. Temos então

$$\mathbf{v}_1 = \mathbf{u}_1 = [2 \quad 0 \quad 1]^T$$

$$\mathbf{v}_2 = \mathbf{u}_2 - \text{proj}_{\mathbf{v}_1} \mathbf{u}_2 = \mathbf{u}_2 - \frac{\mathbf{u}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 = \begin{bmatrix} -2/5 \\ -2 \\ 4/5 \end{bmatrix}$$

Podemos agora determinar a melhor aproximação ao vector \mathbf{v} no espaço coluna de \mathbf{A} .

$$\begin{aligned} \hat{\mathbf{v}} &= \text{proj}_{\mathbf{W}} \mathbf{v} = \text{proj}_{\mathbf{v}_1} \mathbf{v} + \text{proj}_{\mathbf{v}_2} \mathbf{v} \\ &= \frac{\mathbf{v} \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 + \frac{\mathbf{v} \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2} \mathbf{v}_2 \\ &= [1/2 \quad 1/2 \quad 0]^T \end{aligned}$$



```
>> u1=[2 0 1]';
>> u2=[0 -2 1]';
>> A=[u1 u2];
>> v=[0 1 1]';
>> v1=u1;
>> v2=u2-(u2'*v1)/(v1'*v1)*v1;
vm=(v'*v1)/(v1'*v1)*v1+(v'*v2)/(v2'*v2)*v2
vm =
    0.5000
    0.5000
    0
```

DETERMINAR A SOLUÇÃO DE MÍNIMO ERRO QUADRÁTICO DE UM SISTEMA

4. O sistema de equações

$$\begin{cases} x = 0 \\ -2y = 1 \\ x + y = 1 \end{cases}$$

, com forma matricial, $\mathbf{Ax} = \mathbf{b}$,

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

é, como podemos verificar, um sistema impossível. Como a matriz \mathbf{A} (que é a mesma dos exercícios anteriores) tem colunas linearmente independentes, a solução de mínimo erro quadrático do sistema $\mathbf{Ax} = \mathbf{b}$ é única, sendo

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$$

Para evitar a inversão da matriz $\mathbf{A}^T \mathbf{A}$ no cálculo de $\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$, é mais prático resolver o sistema de equações normais $(\mathbf{A}^T \mathbf{A}) \hat{\mathbf{x}} = \mathbf{A}^T \mathbf{b}$. Temos

$$\begin{aligned} (\mathbf{A}^T \mathbf{A}) \hat{\mathbf{x}} &= \mathbf{A}^T \mathbf{b} \\ \begin{bmatrix} 5 & 1 \\ 1 & 5 \end{bmatrix} \hat{\mathbf{x}} &= \begin{bmatrix} 2 & 0 & 1 \\ 0 & -2 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \\ \begin{bmatrix} 5 & 1 \\ 1 & 5 \end{bmatrix} \hat{\mathbf{x}} &= \begin{bmatrix} 1 \\ -1 \end{bmatrix} \end{aligned}$$

e, resolvendo o sistema

$$\begin{bmatrix} 5 & 1 & 1 \\ 1 & 5 & -1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1/4 \\ 0 & 1 & -1/4 \end{bmatrix}$$

temos

$$\hat{\mathbf{x}} = \begin{bmatrix} 1/4 \\ -1/4 \end{bmatrix}$$



```
>> A=[2 0; 0 -2 ;1 1];
>> b=[0 1 1]';
>> xm=inv(A'*A)*A'*v
xm =
    0.2500
   -0.2500
```

Note a imagem de $\hat{\mathbf{x}}$ no espaço coluna de \mathbf{A} , corresponde à projecção ortogonal de \mathbf{b} no espaço coluna de \mathbf{A} .

$$\hat{\mathbf{b}} = \mathbf{A}\hat{\mathbf{x}} = \begin{bmatrix} 2 & 0 \\ 0 & -2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1/4 \\ -1/4 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1/2 \\ 0 \end{bmatrix}$$

, pelo que o exercício anterior pode ser resolvido por esta via, isto é, dada a matriz

$$\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 0 & -2 \\ 1 & 1 \end{bmatrix}$$

e o vector $\mathbf{v} = [0 \ 1 \ 1]^T$, a melhor aproximação ao vector \mathbf{v} no espaço coluna de \mathbf{A} é

$$\hat{\mathbf{v}} = \text{proj}_{\mathbb{W}} \mathbf{v} = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{v} = \mathbf{A}\hat{\mathbf{x}} = \begin{bmatrix} 1/2 \\ 1/2 \\ 0 \end{bmatrix}$$



```
>> A=[2 0; 0 -2 ;1 1];
>> v=[0 1 1]';
>> vm=A*inv(A'*A)*A'*v
vm =
    0.5000
    0.5000
    0
```

DETERMINAR OS PARÂMETROS DE UM MODELO LINEAR QUE MELHOR SE ADAPTA A UM CONJUNTO DE PONTOS

5. Dados os pontos de \mathbb{R}^2 , $(-1.5,2)$, $(-1,1)$, $(0,1)$, e $(0.5,1.5)$, determinar os parâmetros a_0 , a_1 , e a_2 da parábola $y = a_0 + a_1x + a_2x^2$ que melhor se adapta ao conjunto dos ponto, segundo um critério de minimização do erro quadrático cometido.

Dado o sistema de equações

$$\begin{cases} y_1 = a_0 + a_1x_1 + a_2x_1^2 \\ y_2 = a_0 + a_1x_2 + a_2x_2^2 \\ y_3 = a_0 + a_1x_3 + a_2x_3^2 \\ y_4 = a_0 + a_1x_4 + a_2x_4^2 \end{cases}$$

, na forma matricial

$$\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ 1 & x_4 & x_4^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$$

, e no caso presente

$$\begin{bmatrix} 1 & -1.5 & 2.25 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 0.5 & 0.25 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 1.5 \end{bmatrix}$$

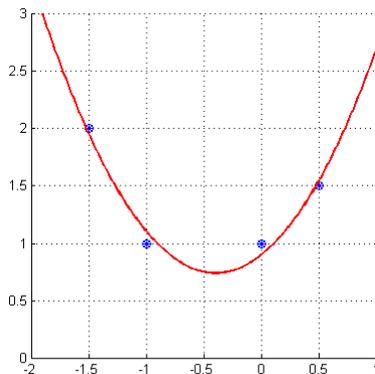
, podemos determinar os valores dos parâmetros da parábola que melhor se adapta ao conjunto de pontos, do ponto de vista da minimização do erro quadrático, resolvendo as equações normais

$$\mathbf{A}^T \mathbf{A} \hat{\mathbf{p}} = \mathbf{A}^T \mathbf{y}$$

Temos então

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1.5 & -1 & 0 & 0.5 \\ 2.25 & 1 & 0 & 0.25 \end{bmatrix} \begin{bmatrix} 1 & -1.5 & 2.25 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 0.5 & 0.25 \end{bmatrix} = \begin{bmatrix} 4 & -2 & 3.5 \\ -2 & 3.5 & -4.25 \\ 3.5 & -4.25 & 6.125 \end{bmatrix}$$

, e



$$\mathbf{A}^T \mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1.5 & -1 & 0 & 0.5 \\ 2.25 & 1 & 0 & 0.25 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 1 \\ 1.5 \end{bmatrix} = \begin{bmatrix} 5.5 \\ -3.25 \\ 5.875 \end{bmatrix}$$

, de onde resulta o sistema

$$\begin{bmatrix} 4 & -2 & 3.5 \\ -2 & 3.5 & -4.25 \\ 3.5 & -4.25 & 6.125 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 5.5 \\ -3.25 \\ 5.875 \end{bmatrix}$$

e, resolvendo o sistema, os parâmetros

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 0.9 \\ 0.8 \\ 1.0 \end{bmatrix}$$

, ou seja, a parábola de equação

$$y = 0.9 + 0.8x + x^2$$



```
>> x1=[-1.5 2]';
>> x2=[-1 1]';
>> x3=[0 1]';
>> x4=[0.5 1.5]';
>> X=[x1 x2 x3 x4];
>> A=[1 1 1 1]' X(1,:) (X(1,:).^2)';
>> b=X(2,:)';
>> p=inv(A'*A)*A'*b
p =
    0.9000
    0.8000
    1.0000
```

CALCULAR A MATRIZ DE PROJEÇÃO NUM SUBESPAÇO

6. Dada a matriz

$$A = \begin{bmatrix} 2 & 0 \\ 0 & -2 \\ 1 & 1 \end{bmatrix}$$

calcular a matriz de projecção no seu espaço coluna.

A matriz de projecção no espaço coluna de A , ou seja, a matriz de projecção no subespaço gerado pelos vectores $u_1 = [2 \ 0 \ 1]^T$ e $u_2 = [0 \ -2 \ 1]^T$ que constituem as suas colunas, é

$$T = A(A^T A)^{-1} A^T$$

Temos

$$A^T A = \begin{bmatrix} 2 & 0 & 1 \\ 0 & -2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & -2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 1 \\ 1 & 5 \end{bmatrix}$$

$$(A^T A)^{-1} = \begin{bmatrix} 5 & 1 \\ 1 & 5 \end{bmatrix}^{-1} = \frac{1}{24} \begin{bmatrix} 5 & -1 \\ -1 & 5 \end{bmatrix}$$

$$A(A^T A)^{-1} = \begin{bmatrix} 2 & 0 \\ 0 & -2 \\ 1 & 1 \end{bmatrix} \frac{1}{24} \begin{bmatrix} 5 & -1 \\ -1 & 5 \end{bmatrix} = \frac{1}{12} \begin{bmatrix} 5 & -1 \\ 1 & -5 \\ 2 & 2 \end{bmatrix}$$

, e finalmente

$$T = A(A^T A)^{-1} A^T = \frac{1}{12} \begin{bmatrix} 5 & -1 \\ 1 & -5 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 2 & 0 & 1 \\ 0 & -2 & 1 \end{bmatrix}$$

$$= \frac{1}{6} \begin{bmatrix} 5 & 1 & 2 \\ 1 & 5 & -2 \\ 2 & -2 & 2 \end{bmatrix}$$

Conhecida a matriz de projecção (ortogonal) no espaço coluna de uma matriz, calcular a imagem de qualquer vector nesse subespaço é trivial. Por exemplo, considerando de novo o vector $\mathbf{v} = [0 \ 1 \ 1]^T$, temos

$$\hat{\mathbf{v}} = \mathbf{T}\mathbf{v} = \frac{1}{6} \begin{bmatrix} 5 & 1 & 2 \\ 1 & 5 & -2 \\ 2 & -2 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1/2 \\ 0 \end{bmatrix}$$



```
>> A=[2 0; 0 -2 ;1 1];
>> T=A*inv(A'*A)*A'
T =
    5/6    1/6    1/3
    1/6    5/6   -1/3
    1/3   -1/3    1/3
>> vm=T*v
vm =
    1/2
    1/2
    0
```

CALCULAR A PSEUDO INVERSA DE UMA MATRIZ

7. Calcular a a matriz psudo-inversa da matriz

$$\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 0 & -2 \\ 1 & 1 \end{bmatrix}$$

A matriz psudo-inversa de uma matriz \mathbf{A} é, por definição,

$$\mathbf{A}^+ = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$$

No caso presente, como já vimos em exemplos anteriores, $\mathbf{A}^T \mathbf{A}$ é invertível, pelo que temos de imediato

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} 2 & 0 & 1 \\ 0 & -2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & -2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 1 \\ 1 & 5 \end{bmatrix}$$

$$(\mathbf{A}^T \mathbf{A})^{-1} = \begin{bmatrix} 5 & 1 \\ 1 & 5 \end{bmatrix}^{-1} = \frac{1}{24} \begin{bmatrix} 5 & -1 \\ -1 & 5 \end{bmatrix}$$

, e finalmente

$$\begin{aligned} \mathbf{A}^+ &= (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T = \frac{1}{24} \begin{bmatrix} 5 & -1 \\ -1 & 5 \end{bmatrix} \begin{bmatrix} 2 & 0 & 1 \\ 0 & -2 & 1 \end{bmatrix} \\ &= \frac{1}{12} \begin{bmatrix} 5 & 1 & 2 \\ -1 & -5 & 2 \end{bmatrix} \end{aligned}$$

Conhecida a matriz pseudo-inversa da matriz A , o cálculo da matriz de projecção no seu espaço coluna, $T = AA^+$, ou da solução de mínimo erro quadrático do sistema $Ax = b$, $\hat{x} = A^+ b$, é trivial. Por exemplo, dado o sistema

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

, a solução de mínimo erro quadrático do sistema é

$$\begin{aligned} \hat{x} = A^+ b &= \frac{1}{12} \begin{bmatrix} 5 & 1 & 2 \\ -1 & -5 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 1/4 \\ -1/4 \end{bmatrix} \end{aligned}$$



```
>> A=[2 0; 0 -2 ;1 1];
>> b=[0 1 1]';
>> pinv(A)
ans =
    5/12    1/12    1/6
   -1/12   -5/12    1/6
>> xm=pinv(A)*b
xm =
    1/4
   -1/4
```