

# AULA 33

- Note bem: a leitura destes apontamentos não dispensa de modo algum a leitura atenta da bibliografia principal da cadeira
- Chama-se a 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.

## TÓPICOS

Melhor aproximação.
Vector de erro.
Mínimo erro quadrático.
Solução de mínimo erro quadrático.
Equações normais.

## 33. Mínimo erro quadrático.

### 33.1. Melhor aproximação.

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

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

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  $\mathbb{W}$ ,

$$\|\mathbf{u} - \hat{\mathbf{u}}\| \leq \|\mathbf{u} - \mathbf{v}\|, \forall \mathbf{v} \in \mathbb{W}$$

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

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

Se  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r$  for uma base ortogonal do subespaço  $\mathbb{W}$ , temos

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

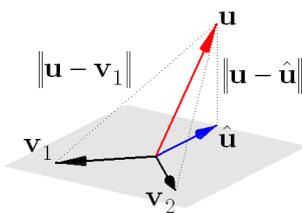
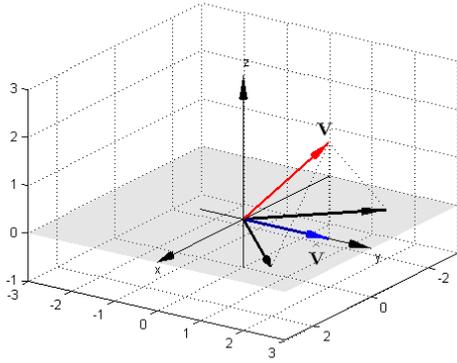


Figura 33.1

### Exemplo

1. 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 do vector  $\mathbf{v}$  no subespaço  $\mathbb{W}$ , 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  $\mathbf{u}_1$  e  $\mathbf{u}_2$  geram o subespaço de  $\mathbb{R}^3$  correspondente ao plano  $xy$ , podemos tomar os versores canónicos  $\mathbf{e}_1$  e  $\mathbf{e}_2$  para base ortogonal do subespaço  $\mathbb{W}$ , resultando

$$\begin{aligned}\hat{\mathbf{v}} &= \text{proj}_{\mathbb{W}} \mathbf{v} \\ &= \text{proj}_{\mathbf{e}_1} \mathbf{v} + \text{proj}_{\mathbf{e}_2} \mathbf{v} \\ &= (\mathbf{v} \cdot \mathbf{e}_1) \mathbf{e}_1 + (\mathbf{v} \cdot \mathbf{e}_2) \mathbf{e}_2 \\ &= 2 \mathbf{e}_2\end{aligned}$$

, ou seja,  $\hat{\mathbf{v}} = 0 \mathbf{e}_1 + 2 \mathbf{e}_2 + 0 \mathbf{e}_3$ , e por isso escrevemos  $\hat{\mathbf{v}} = [0 \ 2 \ 0]^T$ .

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

$$\hat{\mathbf{v}} = \text{proj}_{\mathbb{W}} \mathbf{v} = \frac{1}{2} \mathbf{u}_1 + \frac{1}{2} \mathbf{u}_2 = \sqrt{2} \mathbf{q}_1 + \sqrt{2} \mathbf{q}_2 = 2 \mathbf{e}_2 = [0 \ 2 \ 0]^T$$

A distância do vector  $\mathbf{v}$  aos vectores  $\mathbf{u}_1$ ,  $\mathbf{u}_2$ , e  $\hat{\mathbf{v}}$  é

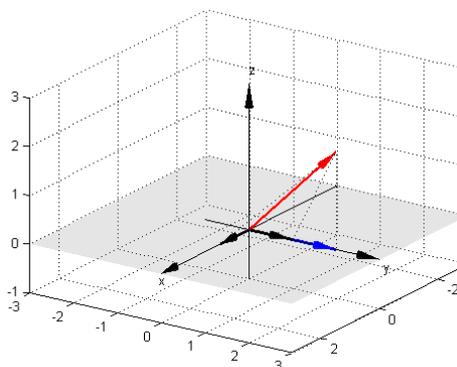
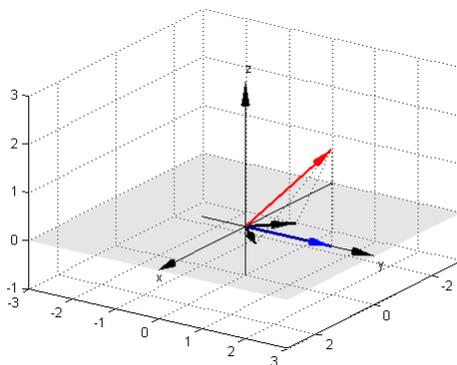


Figura 33.2

$$\begin{aligned} \text{dist}(\mathbf{u}_1, \mathbf{v}) &= \|\mathbf{u}_1 - \mathbf{v}\| = \sqrt{(u_{11} - v_1)^2 + (u_{12} - v_2)^2 + (u_{13} - v_3)^2} \\ &= \sqrt{(2 - 0)^2 + (2 - 2)^2 + (0 - 2)^2} \\ &= \sqrt{8} \end{aligned}$$

$$\begin{aligned} \text{dist}(\mathbf{u}_2, \mathbf{v}) &= \|\mathbf{u}_2 - \mathbf{v}\| = \sqrt{(u_{21} - v_1)^2 + (u_{22} - v_2)^2 + (u_{23} - v_3)^2} \\ &= \sqrt{(-2 - 0)^2 + (2 - 2)^2 + (0 - 2)^2} \\ &= \sqrt{8} \end{aligned}$$

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

A distância de  $\mathbf{v}$  a  $\hat{\mathbf{v}}$  é menor do que a distância a  $\mathbf{u}_1$ , a  $\mathbf{u}_2$ , e menor que a distância a qualquer vector  $\mathbf{u}$  pertencente ao subespaço gerado por  $\mathbf{u}_1$ ,  $\mathbf{u}_2$ ,  $\mathbb{W}$ , e por isso se diz que  $\hat{\mathbf{v}}$  é a **melhor aproximação** de  $\mathbf{v}$  em  $\mathbb{W}$

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

### 33.2. Vector de erro. Mínimo erro quadrático.

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

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

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

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

, designada por **mínimo erro quadrático**.

#### Exemplo

2. Retomemos o exemplo anterior. 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

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

O vector de erro é

$$\begin{aligned} \mathbf{e} &= \mathbf{v} - \hat{\mathbf{v}} \\ &= [0 \ 2 \ 2]^T - [0 \ 2 \ 0]^T \\ &= [0 \ 0 \ 2]^T \\ &= 2\mathbf{e}_3 \end{aligned}$$

, e pertence ao complemento ortogonal de  $\mathbb{W}$ ,

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

, como de imediato se reconhece, dado que  $\mathbb{W}$  corresponde ao plano  $xy$ .

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

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

Se dissermos

$$\mathbf{v} = \hat{\mathbf{v}}$$

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

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

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

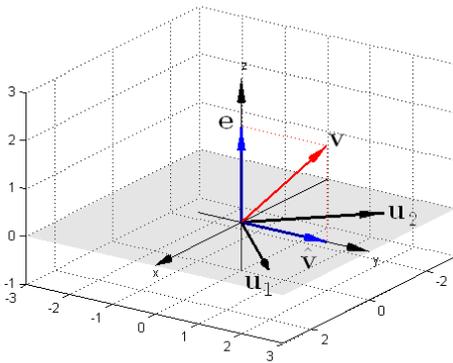


Figura 33.3

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

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

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**

$$\begin{aligned} \|\mathbf{e}_i\| &= \sqrt{(v_1 - u_{i1})^2 + (v_2 - u_{i2})^2 + (v_3 - u_{i3})^2} \\ &= \sqrt{e_{i1}^2 + e_{i2}^2 + e_{i3}^2} \\ &= \sqrt{\sum_k e_{ik}^2} \end{aligned}$$

, ou, de modo equivalente

$$\|\mathbf{e}_i\|^2 = \sum_k e_{ik}^2$$

, em que  $e_{ik}$  corresponde a cada uma das coordenadas do vector erro

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

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

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

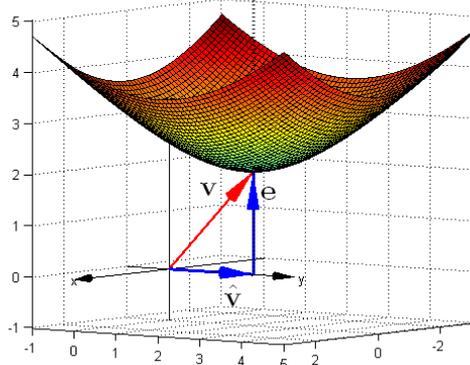


Figura 33.4

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

Para facilidade de compreensão sobrepôs-se à função de erro,  $\|\mathbf{e}_i\|$ , a representação 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 um parabolóide com o mínimo no ponto  $(2, 2)$ , as coordenadas de  $\hat{\mathbf{v}}$  no plano  $\mathbf{xy}$ .

### 33.3. Solução de mínimo erro quadrático. Equações normais.

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

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

Se  $\mathbf{Ax} = \mathbf{b}$  é possível,  $\hat{\mathbf{x}}$  é uma qualquer solução do sistema.

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

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

designadas por **equações normais** de  $\hat{\mathbf{x}}$ . Com efeito,  $\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}$  é ortogonal a  $\text{col}(\mathbf{A})$  pelo que  $\mathbf{A}^T(\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}) = 0 \Leftrightarrow \mathbf{A}^T \mathbf{b} = \mathbf{A}^T \mathbf{A} \hat{\mathbf{x}}$ .

A matriz  $\mathbf{A}$  tem colunas linearmente independentes sse a matriz  $\mathbf{A}^T \mathbf{A}$  for invertível, e, nesse caso, a solução de mínimo erro quadrático do sistema  $\mathbf{Ax} = \mathbf{b}$  é única e igual a

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

#### Exemplos

O sistema de equações

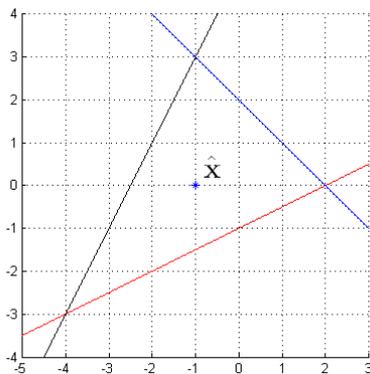


Figura 33.5

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

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

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

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  $\mathbf{b}$  não pertence ao espaço coluna da matriz  $\mathbf{A}$ , não sendo por isso possível determinar um vector  $\mathbf{x}$  (objecto) cujo transformado (imagem)  $\mathbf{Ax}$  pertença a  $\text{col}(\mathbf{A})$ .

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

$$\|\mathbf{b} - \mathbf{Ax}\|$$

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

Como a matriz  $\mathbf{A}$  tem colunas linearmente independentes, a matriz

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

é invertível

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

, pelo que a solução de mínimo erro quadrático do sistema  $\mathbf{Ax} = \mathbf{b}$  é única:

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

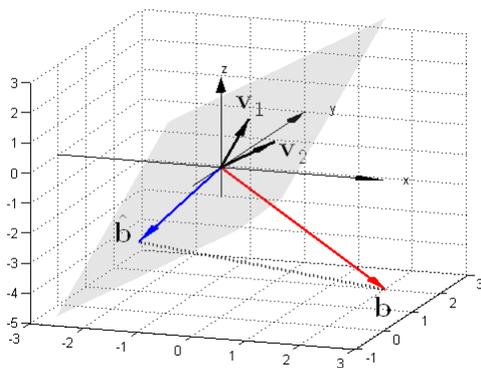


Figura 33.6

A imagem de  $\hat{\mathbf{x}}$  no espaço coluna de  $\mathbf{A}$ , ou seja, a projecção ortogonal de  $\mathbf{b}$  no espaço coluna de  $\mathbf{A}$  é

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

A figura mostra dois vectores,  $\mathbf{v}_1$  e  $\mathbf{v}_2$ , que constituem uma base do espaço coluna da matriz  $\mathbf{A}$ , o espaço por eles gerado, o vector  $\mathbf{b}$ , e a sua projecção no espaço coluna de  $\mathbf{A}$ ,  $\hat{\mathbf{b}}$ .



```
>> 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
```