# A projeção ortogonal como uma transformação linear

Nas condições do slide 171 tem-se, multiplicando à esquerda por $(A^TA)^{-1}$ ambos os membros do sistema das equações normais,

$$\begin{aligned} A^TA\bar{x} = A^Tb \quad &\Leftrightarrow \quad (A^TA)^{-1}A^TA\bar{x} = (A^TA)^{-1}A^Tb \\ &\Leftrightarrow \quad \bar{x} = (A^TA)^{-1}A^Tb, \end{aligned}$$

e portanto,

$$\mathrm{proj}_V(b) = A\bar{x} = A(A^TA)^{-1}A^Tb.$$

Logo a projeção ortogonal sobre $V$ pode ser vista como a transformação linear associada à matriz $P = A(A^TA)^{-1}A^T$,

$$\begin{array}{rcl} \mathbb{R}^m & \rightarrow & \mathbb{R}^m \\ b & \mapsto & \mathrm{proj}_V(b) = Pb, \end{array}$$

o que motiva a definição do próximo slide.



# Matriz de projeção

## Definição da matriz de projeção sobre um subespaço vetorial

Sejam $V$ subespaço vetorial de $\mathbb{R}^m$, $\{v_1, \ldots, v_n\}$ uma base para $V$ e $A = [\, v_1 \ \cdots \ v_n \,]$ a matriz dessa base. A matriz de projeção sobre $V$ é a matriz quadrada de ordem $m$,

$$\boxed{P = A(A^TA)^{-1}A^T,}$$

tendo-se, para todo o $b \in \mathbb{R}^m$,

$$\boxed{\mathrm{proj}_V(b) = P\,b.}$$

## Observação

Nas condições da definição anterior, tem-se ainda para todo o $b \in \mathbb{R}^m$,

$$\mathrm{proj}_{V^\perp}(b) = b - \mathrm{proj}_V(b) = b - Pb = (I_m - P)b,$$

onde $I_m$ denota a matriz identidade de ordem $m$, donde se conclui que $(I_m - P)$ é a matriz de projeção sobre $V^\perp$.

# Propriedades

## Propriedades da matriz de projeção

A matriz de projeção $P$ sobre um subespaço $V$ de $\mathbb{R}^m$ não depende da escolha da base de $V$ e verifica as seguintes propriedades:

- $P^T = P$ (simétrica).
- $P^2 = P$ (idempotente).

(exercício 33.6 da sebenta de exercícios)

A demonstração das seguintes propriedades decorre do facto da projeção ortogonal definir uma transformação linear associada a uma matriz e fica como exercício para os alunos.

## Linearidade da projeção ortogonal

Para todo o $u, v \in \mathbb{R}^m$ e $\alpha \in \mathbb{R}$ tem-se:

- $\text{proj}_V(u+v) = \text{proj}_V(u) + \text{proj}_V(v)$.
- $\text{proj}_V(\alpha\, u) = \alpha\, \text{proj}_V(u)$.



# Matriz de projeção ortogonal sobre uma reta (dimensão 1)

## Observação

Se $V = \langle v \rangle$ é a reta com vetor diretor $v \neq \vec{0}$, a matriz da base de $V$ reduz-se a $v$ e obtém-se a expressão muito elegante para a matriz de projeção sobre $V$,

$$\boxed{P = v(v^T v)^{-1} v^T = \frac{v\, v^T}{v^T v}.}$$

## Exemplo

A matriz de projeção sobre a reta $V = \langle(-1,-1,1)\rangle$, que passa na origem com vetor diretor $(-1,-1,1)$, é a matriz

$$P = \frac{v\, v^T}{v^T v} = \frac{\begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \begin{bmatrix} -1 & -1 & 1 \end{bmatrix}}{\begin{bmatrix} -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}} = \frac{1}{3} \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix}.$$

**TPC**: qual a projeção ortogonal do vetor genérico $b = (b_1, b_2, b_3)$ sobre $V$?

## Matriz de projeção ortogonal sobre um subespaço de dimensão 2

Se $V$ é subespaço vetorial de $\mathbb{R}^m$ de dimensão 2 e $A_{m\times 2}$ é a matriz de uma base de $V$, a matriz de projeção sobre $V$,

$$P = A(A^TA)^{-1}A^T,$$

envolve a inversa da matriz quadrada de ordem 2, $A^TA$, que pode ser facilmente obtida usando a seguinte mnemónica.

**Mnemónica para calcular a inversa de uma matriz $2\times 2$**

"Switch diagonally, negate the wings and divide by a cross" (*):

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

(*) https://www.dam.brown.edu/people/mchb/la/matrix_algebra.pdf
Nota: a matriz é invertível $\Leftrightarrow ad - bc \neq 0$, que se designa por determinante da matriz.

**Exercício na aula**

Determinar as matrizes de projeção sobre $V = \langle (1,0,1),(1,1,2)\rangle$ e $V^\perp$.



## Resolução do exercício na aula

Uma base de $V$ é $\{v_1, v_2\} = \{(1,0,1),(1,1,2)\}$ (justifique) e tem-se, designando por $A = [\, v_1 \;\; v_2 \,]$ a matriz desta base,

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

Usando a mnemónica anterior para calcular $(A^TA)^{-1}$ e determinando a matriz de projeção $P$ sobre $V$ obtém-se,

$$\begin{aligned} P &= A(A^TA)^{-1}A^T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 2 & 3 \\ 3 & 6 \end{bmatrix}^{-1}\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 2 \end{bmatrix}\frac{1}{2\times 6 - 3\times 3}\begin{bmatrix} 6 & -3 \\ -3 & 2 \end{bmatrix}\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix} \\ &= \frac{1}{3}\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 3 & -3 & 0 \\ -1 & 2 & 1 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}. \end{aligned}$$

# Resolução do exercício na aula (concl.)

A matriz de projeção sobre $V^\perp$ é $I_3 - P$, em que $I_3$ é a matriz identidade de ordem 3, e portanto vem dada por

$$\begin{aligned} I_3 - P &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \frac{1}{3}\begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \\ &= \frac{1}{3}\begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix}. \end{aligned}$$

**TPC**: calcule $V^\perp$ e confirme esta matriz com o resultado obtido no exercício do slide 178.



# Aplicação ao famoso conjunto de dados dos lírios de Fisher

Consideremos o conjunto $X_{150\times 4}$ dos dados dos 150 lírios de 3 espécies, *Setosa*, *Versicolor* e *Virginica* do slide 2 que originam uma núvem de 150 pontos em $\mathbb{R}^4$ e denotemos por $y_i \in \mathbb{R}^4$ o vetor dos comprimentos e diâmetros das pétalas e sépalas do lírio $i$ e por $Y = [\, y_1 \; y_2 \; \cdots \; y_{150}\,]_{4\times 150} = X^T$ a matriz destes vetores.

Usando certos resultados de Álgebra Linear determinamos a base $\{u, v\}$, com

$$\begin{aligned} u &= (0.36138659, -0.08452251, 0.85667061, 0.35828920), \\ v &= (0.65658877, 0.73016143, -0.17337266, -0.07548102), \end{aligned}$$

do subespaço vetorial $V$ de dimensão 2, que permite obter o melhor retrato bidimensional (projeção ortogonal sobre $V$) dos líríos, no sentido em que melhor preserva a informação (variabilidade) dos dados originais.

Os vetores $u$ e $v$ são unitários e ortogonais entre si e obtém-se $A^TA = I_2$, onde $A = [\, u \; v\,]$ denota a matriz da base $\{u, v\}$ de $V$. Logo a matriz de projeção $P = A(A^TA)^{-1}A^T$sobre $V$ vem dada simplesmente por $P = AA^T$ e tem-se,

$$P = \begin{bmatrix} 0.56170908 & 0.44887050 & 0.1957547 & 0.07992092 \\ 0.44887050 & 0.54027978 & -0.1989980 & -0.08539683 \\ 0.19575473 & -0.19899799 & 0.7639426 & 0.32002217 \\ 0.07992092 & -0.08539683 & 0.3200222 & 0.13406853 \end{bmatrix}_{4\times 4}.$$

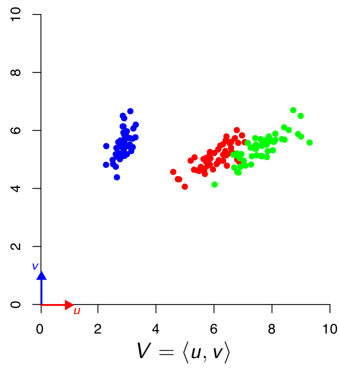
# Aplicação ao famoso iris dataset

A projeção da núvem dos 150 lírios no subespaço $V$ é então dada pela matriz,

$$Y' = PY = [y'_1 \; y'_2 \; \cdots \; y'_{150}] = \begin{bmatrix} 4.7258039 & 4.3890268 & \cdots & 5.802902 \\ 3.8845422 & 3.5246282 & \cdots & 3.100571 \\ 1.4353802 & 1.4957283 & \cdots & 5.030106 \\ 0.5835525 & 0.6102668 & \cdots & 2.088779 \end{bmatrix}_{150\times 4},$$

onde cada vetor $y'_i = Py_i$ representa a projeção do $i$-ésimo lírio no subespaço vetorial $V$. Note que $\text{car}(Y') = 2$ uma vez que $y'_i \in V = \langle u, v\rangle$ para todo o $i$.

Para cada $i = 1, \ldots, 150$, existem escalares $\alpha_i, \beta_i \in \mathbb{R}$ tais que $y'_i = \alpha_i u + \beta_i v$. Projetando no referencial ortonormado definido por $u$ e $v$, cada lírio $y_i$ no ponto de coordenadas $(\alpha_i, \beta_i)$, obtém-se o retrato abaixo da núvem dos 150 lírios.






# Conjunto ortogonal de vetores

## Definição de conjunto ortogonal de vetores

Sejam $v_1, \ldots, v_k \in \mathbb{R}^m$. Diz-se que $\{v_1, \ldots, v_k\}$ é um conjunto ortogonal se os vetores forem 2 a 2 perpendiculares entre si, isto é, se $v_i \cdot v_j = 0$, $\forall i, j = 1, \ldots, k$, $i \neq j$.

## Exemplos

- A base canónica de $\mathbb{R}^m$ é um conjunto ortogonal de vetores
- $\{v_1, v_2, v_3\} = \{(0,1,1), (1,2,-2), (4,-1,1)\}$ é um conjunto ortogonal de vetores de $\mathbb{R}^3$. De facto,

$$\begin{aligned} v_1 \cdot v_2 &= (0,1,1)\cdot(1,2,-2) = 0, && \text{isto é,} && v_1 \perp v_2, \\ v_1 \cdot v_3 &= (0,1,1)\cdot(4,-1,1) = 0, && \text{isto é,} && v_1 \perp v_3, \\ v_2 \cdot v_3 &= (1,2,-2)\cdot(4,-1,1) = 0, && \text{isto é,} && v_2 \perp v_3. \end{aligned}$$