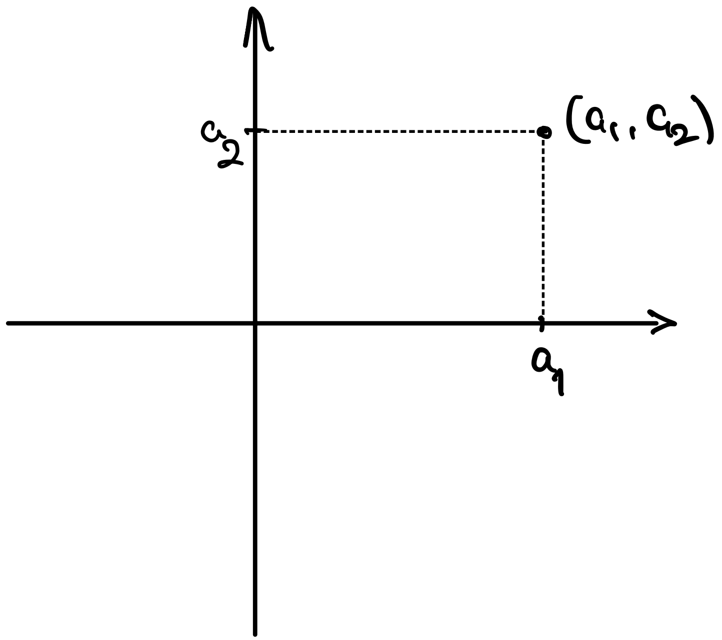
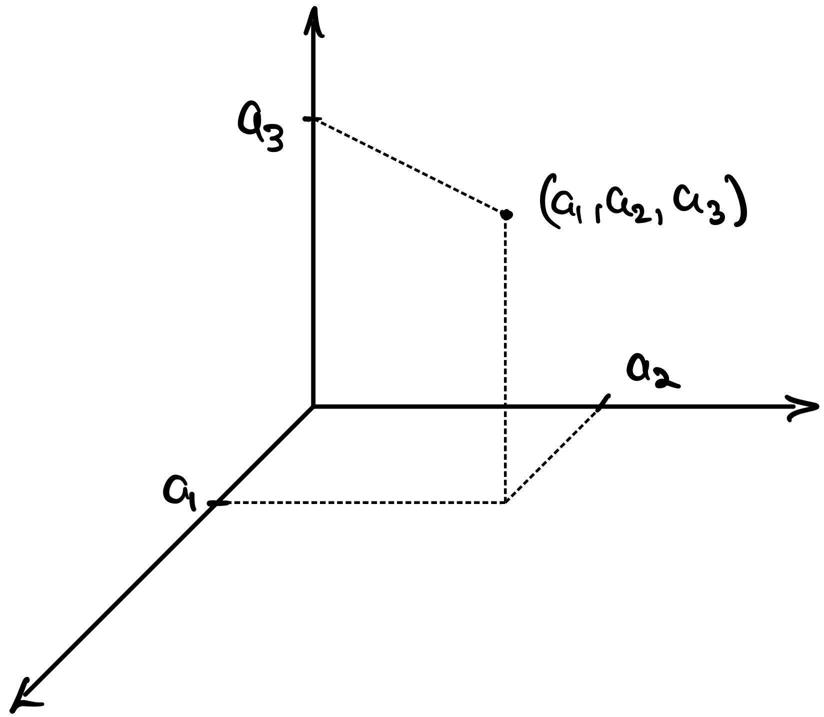
Álgebra de Vetores

$\mathbb{R} = \{\text{números reais}\}$

$\mathbb{R}^n = \{(a_1, a_2, \ldots, a_n) : \text{cada } a_i \text{ é um número real}\}$

Os exemplos mais bem conhecidos são o plano $\mathbb{R}^2$ e o espaço tridimensional $\mathbb{R}^3$.

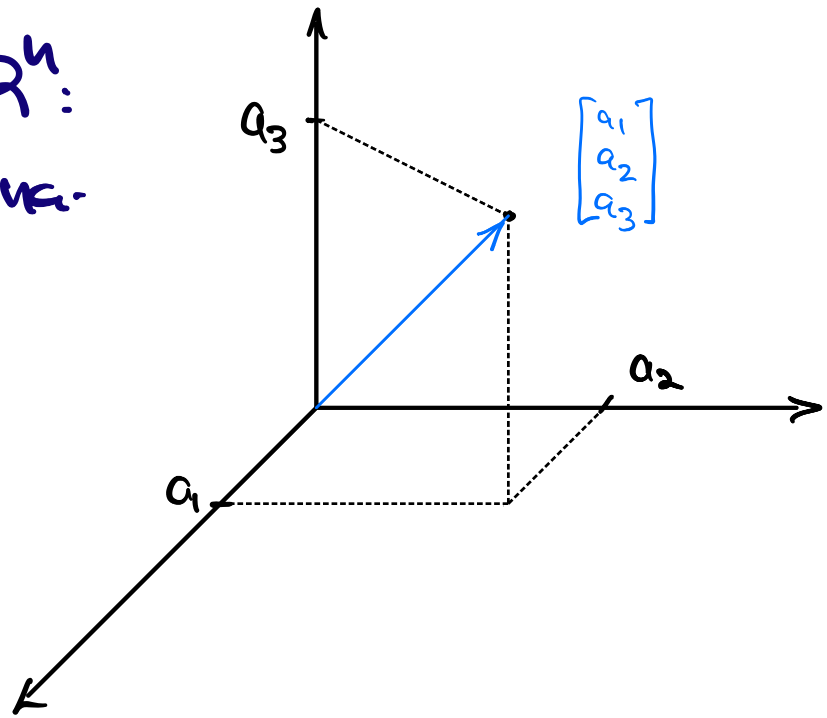




Com frequência usaremos colunas $\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$ para indicar os pontos de $\mathbb{R}^n$.

Também chamaremos os elementos de $\mathbb{R}^n$ de <u>VETORES</u>, quando pensarmos no ponto $(a_1, \ldots, a_n)$ como o ponto final de uma "setinha" que vai da <u>ORIGEM</u> $0 = (0,0,\ldots,0)$ até $(a_1, \ldots, a_n)$.

Podemos <u>SOMAR</u> elementos de $\mathbb{R}^n$: o fazemos coordenada-a-coordenada:



$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} a_1+b_1 \\ a_2+b_2 \\ a_3+b_3 \\ a_4+b_4 \end{bmatrix}$$

Se $c \in \mathbb{R}$, podemos multiplicar c por um elemento de $\mathbb{R}^n$, e chamamos isso de multiplicação por escalar (escalar é um nome que se usa, nesse contexto, para se referir a números $c \in \mathbb{R}$ e distingui-los de vetores).

$$c \cdot \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \begin{bmatrix} ca_1 \\ ca_2 \\ ca_3 \\ ca_4 \end{bmatrix}$$

Multiplicação por escalar tem a interpretação geométrica de manter o vetor na mesma reta e expandi-lo (se $c > 1$), contraí-lo (se $c \in (0,1)$), "trocá-lo de lado" (se $c < 0$), etc.

Soma tem a interpretação geométrica da regra do paralelogramo. Por favor, vejam o livro texto.

Soma e produto por escalar têm as propriedades usuais: se $u, v, w \in \mathbb{R}^n$ e $\lambda, \mu \in \mathbb{R}$, então

(i) $u+v = v+u$

(ii) $(u+v)+w = u+(v+w)$

(iii) $\lambda(\mu u) = (\lambda\mu)u$

(iv) $\lambda(u+v) = \lambda u + \lambda v$

Além disso, o vetor $0 = (0, \dots, 0)$ é o elemento neutro da adição e o escalar (número) $1 \in \mathbb{R}$ serve como elemento neutro do produto por escalar:

$$u + 0 = u, \ \forall u \in \mathbb{R}^n \quad \text{e} \quad 1 \cdot u = u \ \ \forall u \in \mathbb{R}^n.$$

E, claro, $(-1)u$ é o inverso aditivo de $u \in \mathbb{R}^n$, já que

$$(-1)u + u = (-1)u + 1u = (-1+1)u = 0u = 0$$

O PRODUTO ESCALAR OU PRODUTO INTERNO

Definimos agora um "produto", chamado pelos nomes acima, que "come" dois vetores $u, v \in \mathbb{R}^n$ e "cospe" um número, denotado $u \cdot v \in \mathbb{R}$.

<u>Observações</u>: 1) Note que o nome "produto escalar" não é a mesma coisa que o produto por um escalar que definimos anteriormente.

2) Note também que, ao contrário do produto de dois números reais que produz um número real (isto é, o que entra no produto e o que sai são objetos da mesma natureza), o produto escalar come vetores e cospe números, objetos de naturezas distintas.

Def: $u = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$, $v = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{R}^n$

$$u \cdot v = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \sum_{i=1}^{n} u_i v_i \in \mathbb{R}.$$

Ex: $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = 1\cdot(-1) + 2\cdot 0 + 3\cdot 1 = 2.$

Teorema: Se $u, v, w \in \mathbb{R}^n$ e $\lambda \in \mathbb{R}$, então:

(i) $u \cdot v = v \cdot u$

(ii) $u \cdot (v + w) = u \cdot v + u \cdot w$

(iii) $\lambda (u \cdot v) = (\lambda u) \cdot v = u \cdot (\lambda v)$

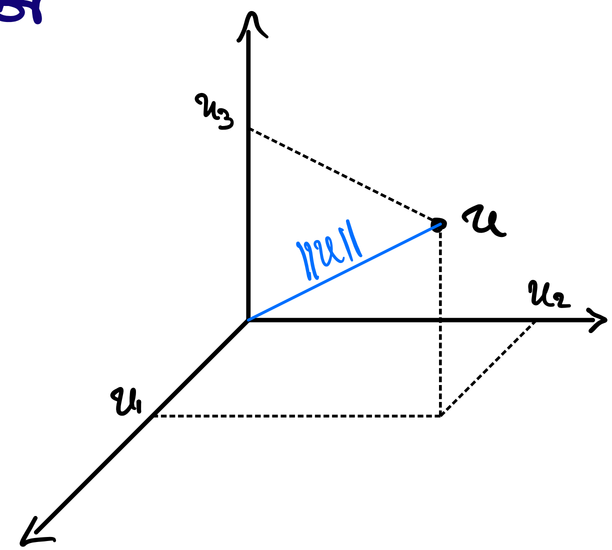
(iv) $u \cdot u > 0 \quad \forall u \neq 0$ e $0 \cdot 0 = 0$ (claro).

As propriedades (i), (ii) e (iii) são consequências diretas de propriedades análogas de números reais. Verifique-as você mesma. Se quiser escolher apenas uma delas, prove (ii).

(iv) Se $u = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$, $u \cdot u = u_1^2 + u_2^2 + \cdots + u_n^2$. Essa é uma soma de quadrados de números reais, que são, portanto, positivos ou 0. Assim a soma é ≥ 0 e só é 0 se todos os $u_i = 0$. $\blacksquare$

Definimos a <u>norma</u> de um vetor $u \in \mathbb{R}$ por

$$\|u\| = \sqrt{u \cdot u} = \left(u_1^2 + u_2^2 + \cdots + u_n^2\right)^{1/2}$$



que deve ser interpretada como a distância de u à origem.

Desigualdade de Cauchy-Schwarz

$$\forall u, v \in \mathbb{R}^n, \quad \boxed{(u \cdot v)^2 \leq (u \cdot u)(v \cdot v)} = \|u\|^2 \|v\|^2$$

Note que, elevando ambos os lado a $1/2$, obtemos

$$\boxed{|u \cdot v| \leq \|u\| \, \|v\|}$$

<u>Prova</u>: Note, para começar, que o resultado é óbvio se qualquer um dos vetores for $= 0$. Assim, podemos supor que $u, v \neq 0$. Sejam $x, y \in \mathbb{R}$ e considere o vetor $xu - yv \in \mathbb{R}^n$. Então

$$0 \leq \|xu - yv\|^2 = (xu - yv) \cdot (xu - yv) =$$

$$= x^2 (u \cdot u) + y^2 (v \cdot v) - 2xy (u \cdot v).$$

Essa desigualdade vale, claro, para quaisquer $x, y \in \mathbb{R}$ e podemos então escolhê-los como quisermos. Tome então $x = v \cdot v$ e $y = u \cdot v$. Temos então

$$0 \leq (v \cdot v)^2 (u \cdot u) + (u \cdot v)^2 (v \cdot v) - 2 (v \cdot v)(u \cdot v)^2$$

$$= (v \cdot v)^2 (u \cdot u) - (v \cdot v)(u \cdot v)^2$$

Como $(v \cdot v) > 0$, podemos dividir por $(v \cdot v)$ e obter

$$(u \cdot u)(v \cdot v) - (u \cdot v)^2 \geq 0$$

que é o que queríamos provar. ∎

Adendo à desigualdade de C·S: Igualdade ocorre somente quando um dos vetores é múltiplo do outro.

Prova: Igualdade acontece acima se e só se $\|xu - yv\| = 0$. Como vimos, isso só é possível se $xu - yv = 0 \iff xu = yv$. $\blacksquare$

Propriedades da norma

Definimos $\|u\| = (u \cdot u)^{1/2}$ e disso segue que:

(i) $\|u\| \geq 0$ e $\|u\| = 0 \iff u = 0$

(ii) $\|\lambda u\| = |\lambda| \, \|u\| \quad \forall \lambda \in \mathbb{R}$

De novo, é fácil verificar essas propriedades e as deixamos para você fazer. Prove (ii) que é bom para a saúde.

Desigualdade Triangular

$$\boxed{\|u+v\| \leq \|u\| + \|v\|}, \qquad \forall u,v \in \mathbb{R}^n$$

Prova: $\|u+v\|^2 = (u+v)\cdot(u+v) = u\cdot u + v\cdot v + 2u\cdot v$

$\leq \|u\|^2 + \|v\|^2 + 2\|u\|\|v\| = (\|u\| + \|v\|)^2$ ■

Adendo à desigualdade triangular: Igualdade ocorre se e somente se um dos vetores é múltiplo positivo (ou zero) do outro.

Prova: Para que valha a igualdade na demonstração acima é preciso que

$$u\cdot v = \|u\|\,\|v\| \quad (*)$$

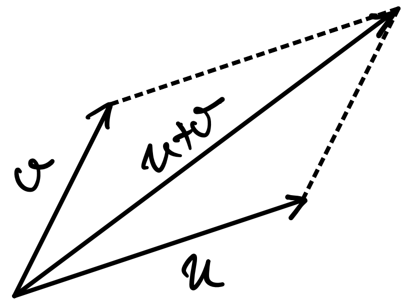
Como vimos no adendo à desigualdade C-S, $|u\cdot v| = \|u\|\|v\|$ $\iff$ $u = \lambda v$ (ou vice versa). Nesse caso, $u\cdot v = \lambda \|v\|^2$ e $\|u\| = |\lambda|\,\|v\|$. Assim $(*)$ vale $\iff \lambda \geq 0$. ■

Ortogonalidade

$$\|u+v\|^2 = \|u\|^2 + \|v\|^2 + 2\,u\cdot v$$

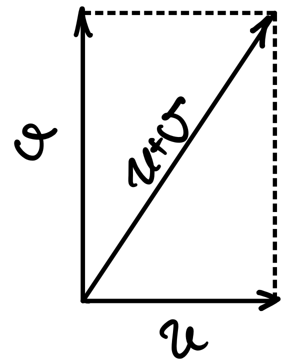
Assim, o Teorema de Pitágoras vale se e somente se $u\cdot v = 0$. Isso nos leva a declarar dois vetores u, v ortogonais se (e só se) $u\cdot v = 0$.

Ex: $\begin{bmatrix} 2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} -3 \\ 6 \end{bmatrix} = 0$



$$\|u+v\| < \|u\| + \|v\|$$

Desigualdade Triangular

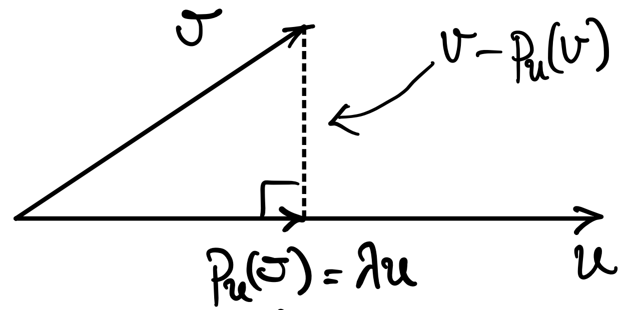


$$\|u+v\|^2 = \|u\|^2 + \|v\|^2$$

Pitágoras

Vetores ortogonais são também chamados perpendiculares.

Projeção Ortogonal



Projeção ortogonal de v na direção de u.

A projeção ortogonal $P_u(v)$ do vetor v na direção do vetor u é, por definição, um múltiplo λu de u de modo que

$$(v - \lambda u) \perp u$$

Perpendicular

Isto é, queremos que $(v - \lambda u) \cdot u = 0$, ou seja

$$u \cdot v = \lambda\, u \cdot u \iff \lambda = \frac{u \cdot v}{u \cdot u}$$

$$\boxed{P_u(v) = \left(\frac{u \cdot v}{u \cdot u}\right) u} = \left(\frac{u \cdot v}{\|u\|^2}\right) u$$

Observação IMPORTANTÍSSIMA

Se pensamos em vetores em $\mathbb{R}^n$ como colunas, o produto escalar pode — e deve — ser pensado como multiplicação matricial.

$$u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \qquad v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

$$\boxed{u \cdot v = u^T v}$$

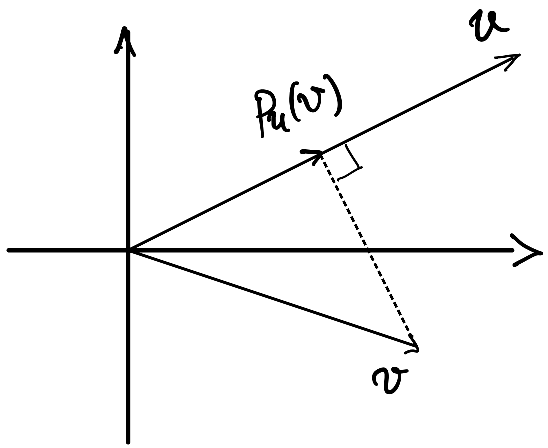
$$u \cdot v = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$$

$$u^T v = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$$

Assim, a projeção é dada por:

$$\boxed{P_u(v) = \frac{u \cdot v}{u \cdot u}\, u = \frac{u^T v}{u^T u}\, u = \frac{u^T v}{\|u\|^2}\, u}$$

<u>Exemplo</u>: $u = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$, $v = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$



$$P_u(v) = \frac{u \cdot v}{u \cdot u} \cdot u = \frac{\begin{bmatrix} 4 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \end{bmatrix}}{\begin{bmatrix} 4 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \end{bmatrix}} \cdot \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

$$= \frac{10}{20} \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Multiplicação por escalar pode ser feita à direita ou à esquerda. Portanto podemos escrever a projeção da seguinte forma (onde também usamos associatividade):

$$P_u(v) = \frac{1}{u^T u} \cdot u \cdot (u^T v) = \frac{1}{u^T u}(u u^T) \cdot v$$

<u>Exemplo</u> (anterior): $u = \begin{bmatrix} 4 \\ 2 \end{bmatrix} \Rightarrow \quad u^T u = \begin{bmatrix} 4 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \end{bmatrix} = 16 + 4 = 20$

$$\begin{bmatrix} 4 \\ 2 \end{bmatrix} \begin{bmatrix} 4 & 2 \end{bmatrix} = \begin{bmatrix} 16 & 8 \\ 8 & 4 \end{bmatrix}$$

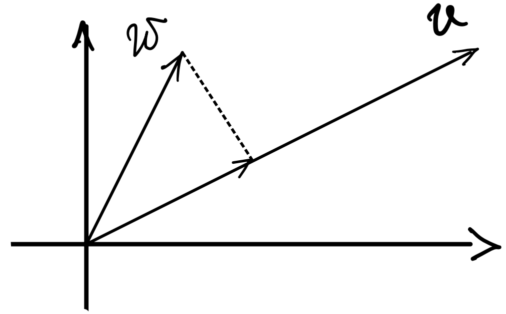
$$P_u(v) = \frac{1}{20} \begin{bmatrix} 16 & 8 \\ 8 & 4 \end{bmatrix} \begin{bmatrix} \,\cdot\, \\ 1 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 12-2 \\ 6-1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Isto é, multiplicação à esquerda pela matriz

$$\boxed{P_u = \frac{1}{u^T u}\, u u^T}$$

realiza a projeção ortogonal na direção de u. Portanto, se quisermos agora projetar outro vetor $w = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ na direção u é só multiplicá-lo, à esquerda:

$$P_u w = \frac{1}{5}\begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 8 \\ 4 \end{bmatrix} = \begin{bmatrix} 1.6 \\ 0.8 \end{bmatrix}$$



Observação: Teria sido, se não impossível, altamente improvável que notássemos que a projeção é feita por essa multiplicação matricial se nós tivéssemos observado que $u \cdot v = u^T v$.

<u>Exemplo</u>: A matriz que projeta na direção do vetor $\begin{bmatrix}1\\1\\1\end{bmatrix}$ em $\mathbb{R}^3$ é

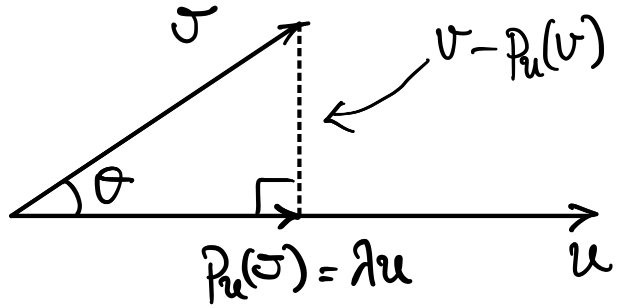
$$P = \frac{1}{u^T u} u u^T = \frac{1}{3}\begin{bmatrix}1 & 1 & 1\\1 & 1 & 1\\1 & 1 & 1\end{bmatrix}$$

Se tomamos um múltiplo de $\begin{bmatrix}1\\1\\1\end{bmatrix}$, por exemplo, $\begin{bmatrix}2\\2\\2\end{bmatrix}$, obtemos

$$\frac{1}{12}\begin{bmatrix}4 & 4 & 4\\4 & 4 & 4\\4 & 4 & 4\end{bmatrix} = \frac{1}{3}\begin{bmatrix}1 & 1 & 1\\1 & 1 & 1\\1 & 1 & 1\end{bmatrix} = P$$

Ângulos

O ângulo θ na figura é tal que



$$\cos\theta = \frac{\|P_u(v)\|}{\|v\|} = \frac{u^T v}{\|u\|^2 \|v\|} \|u\|$$

$$= \frac{u^T v}{\|u\| \|v\|}$$

ou

$$\boxed{u \cdot v = \|u\| \|v\| \cos\theta}$$

Observação: Fomos desonestos na igualdade. Por quê?
Mas a igualdade na caixinha acima é honesta.

Sugestão: Considere ângulos $> \pi/2$.

Podemos então definir o ângulo entre dois vetores $u, v \in \mathbb{R}^n$ como sendo o número θ tal que

$$\boxed{\cos\theta = \frac{u^T v}{\|u\|\,\|v\|}}$$

A desigualdade de Cauchy-Schwarz garante que a definição é de boa qualidade:

$$|u^T v| \leq \|u\|\,\|v\| \iff$$

$$-1 \leq \frac{u^T v}{\|u\|\,\|v\|} \leq 1$$

Por exemplo, dois vetores são perpendiculares $\iff u^T v = 0$

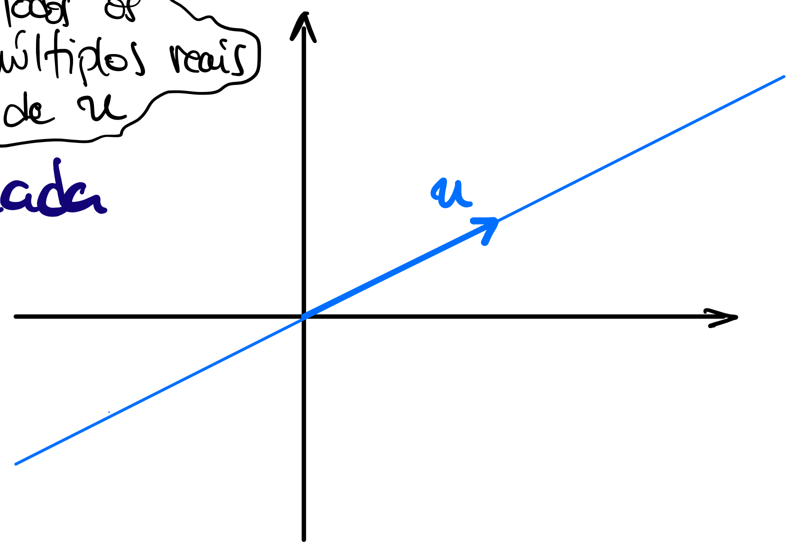
$\iff \cos\theta = 0 \iff \theta = \pm \pi/2 + 2k\pi$.

Planos e mais projeções

Quando fizemos a projeção $P_u(v)$, fizemos, na verdade, a projeção na reta determinada por u, isto é, no conjunto

$$\langle u \rangle := \{ tu : t \in \mathbb{R} \}$$

Todos os múltiplos reais de u

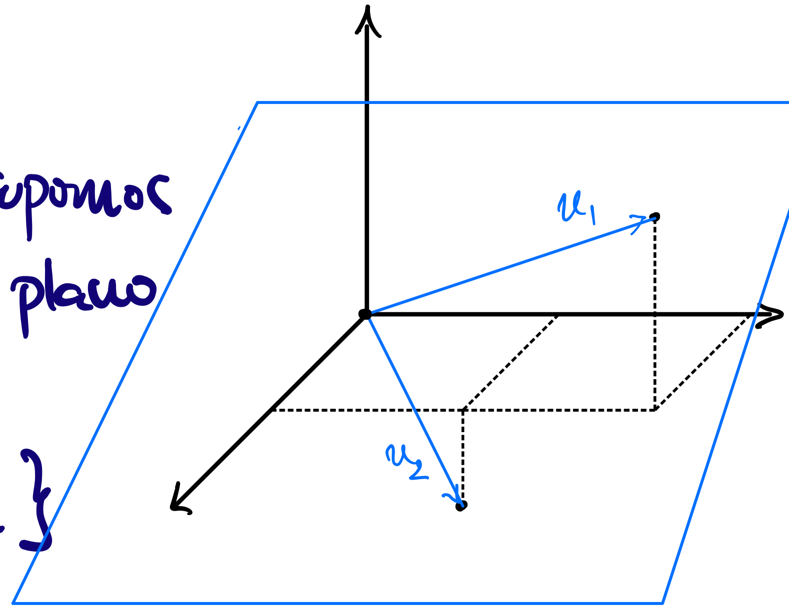


Ex: Se $u = \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix} \in \mathbb{R}^3$, a reta determinada por u é

$$\left\langle \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix} \right\rangle = \left\{ \begin{bmatrix} 3t \\ 0 \\ -t \end{bmatrix} : t \in \mathbb{R} \right\}$$

A reta determinada por u é, geometricamente, a reta que contém os pontos u e $0 \in \mathbb{R}^n$.

E um plano em $\mathbb{R}^n$ o que será?



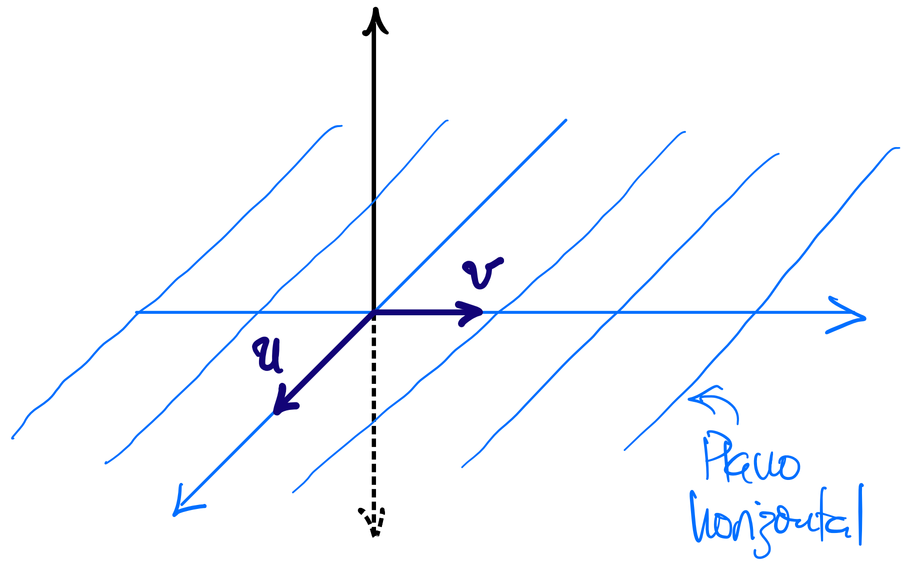
Dados dois vetores $u_1, u_2 \in \mathbb{R}^n$, que supomos não serem múltiplos um do outro, o plano determinado por eles é

$$\langle u, v \rangle = \{ xu + yv : x, y \in \mathbb{R} \}$$

isto é, é o conjunto de todas as "combinações lineares" dos vetores u_1 e u_2.

<u>Ex</u>: $u = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $v = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

$$\langle u, v \rangle = \left\{ \begin{bmatrix} x \\ y \\ 0 \end{bmatrix} : x, y \in \mathbb{R} \right\}$$



Ex: $u = \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}$ $v = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$ $\Rightarrow \langle u, v \rangle = \left\{ \begin{bmatrix} 3x+y \\ -y \\ -x+2y \end{bmatrix} : x, y \in \mathbb{R} \right\}$

Há uma outra forma de pensar em todas as combinações lineares de dois vetores u, v que já vimos antes:

$$\begin{bmatrix} | & | \\ u & v \\ | & | \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = x \begin{bmatrix} | \\ u \\ | \end{bmatrix} + y \begin{bmatrix} | \\ v \\ | \end{bmatrix}$$

Isto é, o plano $\langle u_1, u_2 \rangle$ é o conjunto de todos os produtos

$$\langle u, v \rangle = \left\{ \begin{bmatrix} | & | \\ u & v \\ | & | \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} : \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 \right\}$$

Ex: $\begin{bmatrix} 3 & 1 \\ 0 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = x \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix} + y \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3x+y \\ -y \\ -x+2y \end{bmatrix}$

Projeção ortogonal em um plano

Vamos mudar a notação um pouquinho, mas é por uma boa causa. Suponha que são dados dois vetores $a_1, a_2 \in \mathbb{R}^n$. O plano determinado por eles é, como acabamos de ver,

$$\langle a_1, a_2 \rangle = \left\{ \begin{bmatrix} | & | \\ a_1 & a_2 \\ | & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} : \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2 \right\}$$

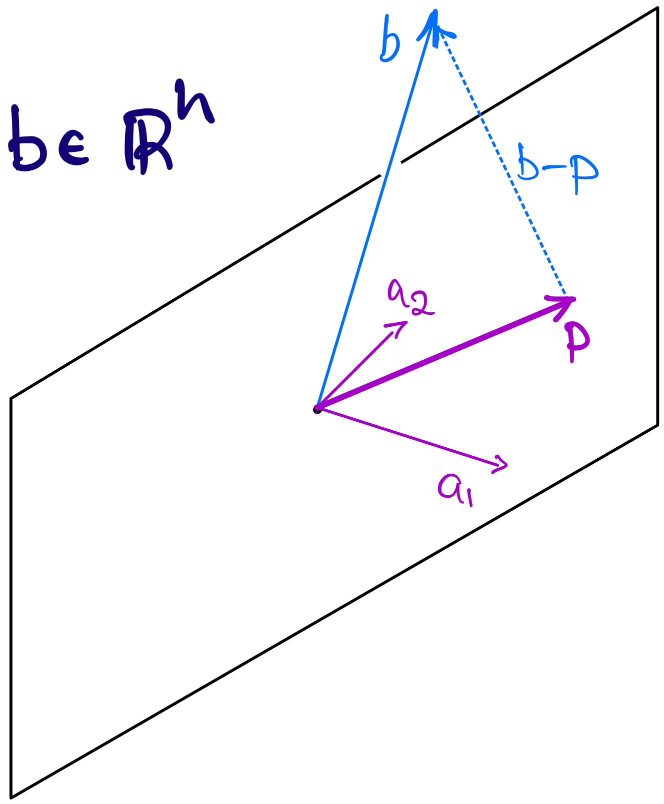
Denotemos então por A a matriz com duas colunas

$$A = \begin{bmatrix} | & | \\ a_1 & a_2 \\ | & | \end{bmatrix}_{n \times 2}$$

e por x o vetor de duas coordenadas $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. Assim

$$\langle a_1, a_2 \rangle = \{ Ax : x \in \mathbb{R}^2 \}$$

Suponha agora que nos é dado um vetor $b \in \mathbb{R}^n$ e queremos projetá-lo ortogonalmente no plano $\langle a_1, a_2 \rangle$.



Precisamos encontrar um vetor p nesse plano tal que $(b-p)$ seja perpendicular ao plano. Todos os vetores no plano são da forma Ax para algum x e portanto $p = Ax$, para algum x.

Queremos então encontrar x de forma que $(b - Ax)$ seja perpendicular ao plano gerado por a_1 e a_2. Isto quer dizer que $(b-Ax)$ deve ser perpendicular a todos os vetores desse plano, em particular, a a_1 e a_2:

$$\begin{cases} a_1^T(b-Ax) = 0 \\ a_2^T(b-Ax) = 0 \end{cases} \iff$$

$$\begin{cases} a_1^T Ax = a_1^T b \\ a_2^T Ax = a_2^T b \end{cases} \iff$$

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

<u>Ex</u>: $a_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$ $\quad a_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$ $\quad b = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$

$$A = \begin{bmatrix} 1 & -1 \\ 2 & 1 \\ 0 & 0 \end{bmatrix} \qquad A^T A = \begin{bmatrix} 5 & 1 \\ 1 & 2 \end{bmatrix} \qquad A^T b = \begin{bmatrix} 7 \\ -1 \end{bmatrix}$$

Temos então que resolver o sistema

$$\begin{bmatrix} 5 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 7 \\ -1 \end{bmatrix} \iff \begin{cases} 5x_1 + x_2 = 7 \\ x_1 + 2x_2 = -1 \end{cases}$$

Esse sisteminha 2×2 você resolve como bem entender. Uma forma de fazê-lo é multiplicando o lado direito pela inversa da matriz 2×2 à esquerda (que é, afinal A^TA):

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 2 & -1 \\ -1 & 5 \end{bmatrix}\begin{bmatrix} 7 \\ -1 \end{bmatrix} = \begin{bmatrix} 15/9 \\ -12/9 \end{bmatrix}$$

Mas $p = Ax \Rightarrow p = \begin{bmatrix} 1 & -1 \\ 2 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 15/9 \\ -12/9 \end{bmatrix} = \begin{bmatrix} (15+12)/9 = 3 \\ (30-12)/9 = 2 \\ 0 = 0 \end{bmatrix}$

isto é,

$$p = \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix}.$$

Surpreendente? Ou perfeitamente natural?

Em resumo: para projetar ortogonalmente no plano gerado pelas colunas da matriz

$$A = \begin{bmatrix} | & | \\ a_1 & a_2 \\ | & | \end{bmatrix}$$

devemos resolver o sistema $\underline{2 \times 2}$

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

Uma vez encontrada a soluçã $\bar{x}$ do sistema, a projeçã é

$$p = A\bar{x}.$$

E, como a soluçã $\bar{x}$ do sistema é $\bar{x} = (A^TA)^{-1}A^Tb$, segue que

$$\boxed{p = A(A^TA)^{-1}A^Tb}$$

Projeções e método dos mínimos quadrados

Suponha que queiramos resolver um sistema com apenas uma variável e três equações

$$\begin{cases} 2x = b_1 \\ 3x = b_2 \\ 4x = b_3 \end{cases}$$

A menos que $b_1 : b_2 : b_3$ estejam nas proporções $2:3:4$ (por exemplo, $b_1 = 4$, $b_2 = 6$, $b_3 = 8$, caso em que $x = 2$), não haverá solução exata para o sistema. Sistemas como esse, com mais equações equações ("exigências") do que variáveis ("graus de liberdade") são chamados SOBREDETERMINADOS. Tipicamente, tais sistemas não têm solução. No entanto, eles aparecem na prática e devem ser "resolvidos" de alguma forma.

O que fazemos nesses casos é "minimizar o erro médio das equações" e a forma mais útil de fazer isso (que poderia ser feito de várias maneiras diferentes) é definir o erro como uma soma de quadrados. No exemplo acima, definimos:

$$E^2(x) = (b_1 - 2x)^2 + (b_2 - 3x)^2 + (b_3 - 4x)^2$$

Se houver uma solução exata para o sistema, há uma escolha $\bar{x}$ de forma que $E^2(\bar{x}) = 0$. Na hipótese mais provável de que não haja um tal $\bar{x}$, podemos encontrar o valor $\bar{x}$ para o qual $E^2(x)$ é mínimo, já que $E^2(x)$ é uma parábola com coeficiente de x^2 igual a $2^2 + 3^2 + 4^2 > 0$. Para encontrar o mínimo dessa parábola, encontramos seu único ponto crítico, isto é, a solução $\bar{x}$ da equação $\frac{d}{dx}E^2(x) = 0$:

$$E^2(x) = (b_1 - 2x)^2 + (b_2 - 3x)^2 + (b_3 - 4x)^2$$

$$\frac{d}{dx}E^2(x) = 2\left[(2x - b_1)\,2 + (3x - b_2)\,3 + (4x - b_3)\,4\right] = 0$$

Resolvendo a equação, obtemos

$$\boxed{\bar{x} = \frac{2b_1 + 3b_2 + 4b_3}{2^2 + 3^2 + 4^2}}$$

= !

Mudando de assunto (será?), suponha agora que queiramos encontrar a projeção ortogonal de um vetor $b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$ sobre a reta determinada por $a = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$.

Mas sabemos como fazer isso:

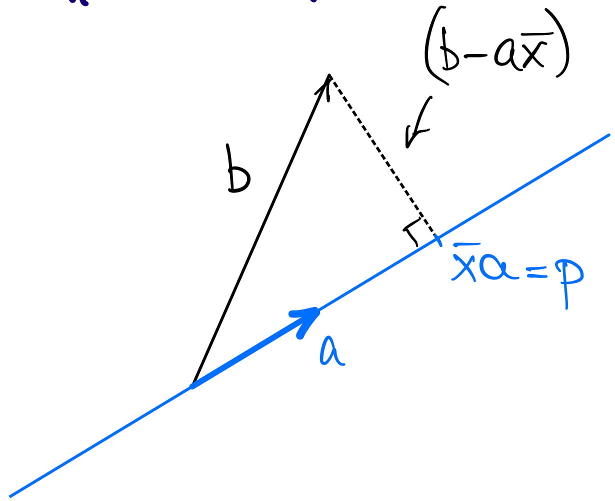
$$p = \bar{x}a \quad \text{onde} \quad \bar{x} = \frac{a^Tb}{a^Ta} = \frac{\begin{bmatrix} 2 & 3 & 4 \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}}{\begin{bmatrix} 2 & 3 & 4 \end{bmatrix}\begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}} = \boxed{\frac{2b_1 + 3b_2 + 4b_3}{2^2 + 3^2 + 4^2}}$$

O que aconteceu exatamente?

Definimos o erro quadrático por

$$E^2(x) = (b_1 - 2x)^2 + (b_2 - 3x)^2 + (b_3 - 4x)^2$$

$$= \left\| \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} - x \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} \right\|^2 = \| b - ax \|^2$$



onde $b = [b_1\ b_2\ b_3]^T$ e $a = [2\ 3\ 4]^T$.
O que vimos acima é que minimizar o erro quadrático é a mesma coisa que minimizar $\|b - ax\|^2$, e que, por sua vez, se $\bar{x}$ minimiza essa $(\text{norma})^2$, o ponto $p = a\bar{x}$ é a projeção ortogonal de $b = [b_1\ b_2\ b_3]^T$ reta determinada por $a = [2\ 3\ 4]$.

O caso geral é análogo:

Dados $b = [b_1\, b_2 \dots b_m]^T$ e $a = [a_1\, a_2 \dots a_m]^T$:

Podemos minimizar o erro quadrático no sistema

$$\begin{cases} a_1 x = b_1 \\ a_2 x = b_2 \\ \vdots \qquad \vdots \\ a_m x = b_m \end{cases}$$

$$E^2(x) = \| b - ax \|^2$$
$$= \sum_{i=1}^{m} (b_i - a_i x)^2$$

$$\frac{d}{dx} E^2(x) = 2 \sum_{i=1}^{m} a_i(a_i x - b_i) = 0 \iff \boxed{\bar{x} = \frac{\sum a_i b_i}{\sum a_i^2}}$$

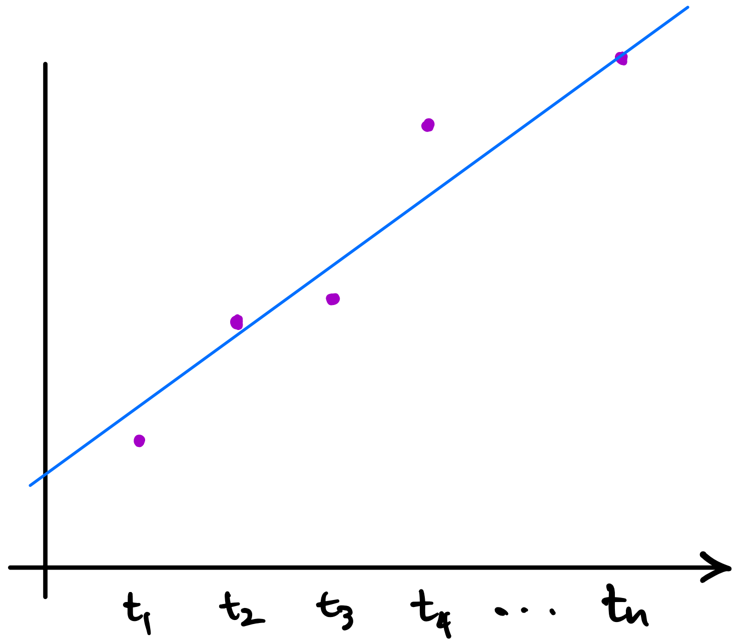
Ou podemos encontrar a projeção de b sobre a reta determinada por a:

$$p = \bar{x}\, a \quad \text{onde} \quad \boxed{\bar{x} = \frac{a^T b}{a^T a} = \frac{\sum a_i b_i}{\sum a_i^2}}$$

Regressão linear

Suponha que um sistema físico, por exemplo, é modelado por uma função linear $\boxed{f(t) = k + \sigma t}$ e que fazemos diversas medidas em vários instantes de tempo e obtemos uma lista. Como se trata de um experimento, é pouco provável que as medidas feitas estejam exatamente alinhadas em uma reta e queremos então encontrar a reta que "melhor aproxima" a tabela.

t	t_1	t_2	t_3	$\dots$	t_n
$f(t)$	b_1	b_2	b_3	$\dots$	b_n



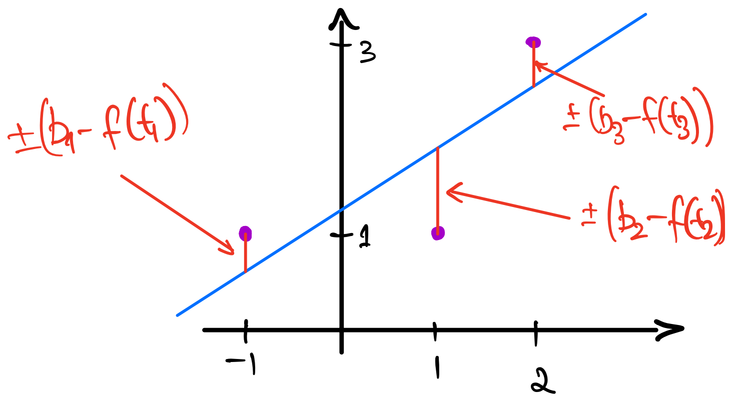
Mas o que quer dizer "melhor aproxima"?

Há muitas formas de definir essa expressão, mas a mais útil e mais utilizada é encontrar a solução que "**minimiza o erro quadrático**".

<u>Exemplo</u>: Suponha que nos é dada a tabela.

t	-1	1	2
b	1	1	3

Queremos encontrar uma reta que "aproxima bem" a tabela.



Definimos o **ERRO QUADRÁTICO** por

$$\boxed{\begin{aligned} E(k,v) &= \sum_{i=1}^{n} \left(b_i - f(t_i)\right)^2 \\ &= \sum_{i=1}^{n} \left[b_i - (k + v t_i)\right]^2 \end{aligned}}$$

Note que, ao elevar os números ao quadrado, não temos que nos preocupar com o sinal.

Queremos encontrar o mínimo da função $E(k,v)$. Para encontrar máximos e mínimos de funções de uma variável, procuramos seus "pontos críticos", isto é, os pontos nos quais a derivada se anula. Para funções de duas variáveis, fazemos o mesmo com cada uma delas, isto é, tratamos $E(k,v)$ como função de k apenas, calculamos a derivada e igualamos a zero; em seguida, fazemos o mesmo com v.

Derivadas Parciais de E em relação a k e v

$$\frac{\partial}{\partial k}E(k,v) = (-2)\sum_{i=1}^{n}\left[b_i-(k+vt_i)\right] = 0$$

$$\frac{\partial}{\partial v}E(k,v) = (-2)\sum_{i=1}^{n} t_i\left[b-(k+vt_i)\right] = 0$$

$$\Longleftrightarrow \boxed{\begin{cases} nk+\left(\sum t_i\right)v = \sum b_i \\ \left(\sum t_i\right)k+\left(\sum t_i^2\right)v = \sum t_ib_i \end{cases}}$$

<u>Exemplo</u>: $E(k,v) = (1-k+v)^2 + (1-k-v)^2 + (3-k-2v)^2$

$$\begin{cases} \dfrac{\partial E}{\partial k} = \cancel{-2}(1-k+v) \cancel{-2}(1-k-v) \cancel{-2}(3-k-2v) = 0 \\[2ex] \dfrac{\partial E}{\partial v} = \cancel{2}(1-k+v) \cancel{-2}(1-k-v) \overset{(-2)}{\cancel{-4}}(3-k-2v) = 0 \end{cases} \iff$$

Dividindo ambas por 2

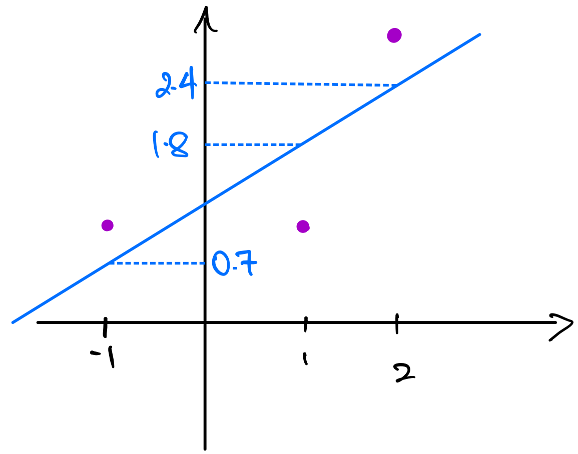
$$\begin{cases} 3k + 2v = 5 \\ 2k + 6v = 6 \end{cases} \iff \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} k \\ v \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \end{bmatrix} \iff$$

$$\begin{bmatrix} k \\ v \end{bmatrix} = \frac{1}{14} \begin{bmatrix} 6 & -2 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 5 \\ 6 \end{bmatrix} \iff \begin{cases} \bar{k} = 9/7 \cong 1.3 \\ \bar{v} = 4/7 \cong 0.6 \end{cases}$$

Isto é, $\underline{f(t) = \frac{9}{7} + \frac{4}{7}t}$.

Em particular, $f(-1) = 5/7$, $f(1) = 13/7$, $f(2) = 17/7 \cong 2{,}4$

$\cong 0.7$ $\cong 1{,}8$



ELIMINAÇÃO GAUSSIANA

1º pivô

$$\begin{cases} 2u + v + w = 5 \\ 4u - 6v = -2 \\ -2u + 7v + 2w = 9 \end{cases}$$

(a) Subtraia $2\times$(1ª eq) da 2ª eq.
(b) Subtraia $(-1)\times$(1ª eq.) da 3ª eq.

2º pivô

$$\begin{cases} 2u + v + w = 5 \\ \quad -8v - 2w = -12 \\ \quad\ \ 8v + 3w = 14 \end{cases}$$

(c) Subtraia $(-1)\times$(2ª eq.) da 3ª eq.

$$\begin{cases} 2u + v + w = 5 \\ \quad -8v - 2w = -12 \\ \qquad\quad w = 2 \end{cases}$$

3º pivô = 1

Essa parte do processo, que "elimina" variáveis de equações abaixo do pivô, subtraindo certos múltiplos de uma equação das equações abaixo dela, é chamado ELIMINAÇÃO.

Do sistema equivalente

$$\begin{cases} 2u + v + w = 5 \\ \quad -8v - 2w = -12 \\ \qquad\qquad w = 2 \end{cases}$$

é agora fácil encontrar a solução: substituindo $w = 2$ na segunda equação, obtemos

$$v = \frac{1}{-8}(-12 + 4) = 1$$

e substituindo $w = 2$ e $v = 1$ na primeira equação, obtemos

$$u = \frac{1}{2}(5 - 1 - 2) = 1$$

Assim, $u = 1$, $v = 1$, $w = 2$ é a solução do sistema. Essa segunda parte do processo é chamada SUBSTITUIÇÃO REVERSA.

Matriz

As variáveis u, v, w só são úteis no processo de eliminação para "marcar lugar". O que realmente importa é o que acontece com os coeficientes. Podemos então economizar tinta e poluição visual representando o que acontece assim:

$$\begin{bmatrix} 2 & 1 & 1 & 5 \\ 4 & -6 & 0 & -2 \\ -2 & 7 & 2 & 9 \end{bmatrix} \longrightarrow \begin{bmatrix} 2 & 1 & 1 & 5 \\ 0 & -8 & -2 & -12 \\ 0 & 8 & 3 & 14 \end{bmatrix} \longrightarrow \begin{bmatrix} 2 & 1 & 1 & 5 \\ 0 & -8 & -2 & -12 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

Essa setinha são passos (a) e (b) acima...

... e essa é o passo (c).

Os números que aparecem na diagonal no final do processo de eliminação são os <u>PIVÔS</u>.
Pivôs são, por def., $\neq 0$

O que pode dar errado

Caso 1: Curável

$$\begin{cases} u + v + w = \sim \\ 2u + 2v + 5w = \sim \\ 4u + 6v + 8w = \sim \end{cases}$$

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & \sim \\ 2 & 2 & 5 & \sim \\ 4 & 6 & 8 & \sim \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & 1 & 1 & \sim \\ 0 & 0 & 3 & \sim \\ 0 & 2 & 4 & \sim \end{array}\right] \longleftrightarrow \begin{cases} u + v + w = \sim \\ w = \sim \\ 2v + 4w = \sim \end{cases}$$

$$\Updownarrow$$

Trocando a 2ª e 3ª equações obtemos um sistema triangular novamente e podemos fazer substituição reversa

$$\begin{cases} u + v + w = \sim \\ 2v + 4w = \sim \\ w = \sim \end{cases}$$

Caso 2: Inversível

$$\begin{cases} u+v+w = \sim \\ 2u+2v+5w = \sim \\ 4u+4v+8w = \sim \end{cases} \rightsquigarrow \begin{cases} u+v+w = \sim \\ 3w = \sim \\ 4w = \sim \end{cases}$$

Nesse caso, nós encontramos um conjunto completo de (3) pivôs: lembre-se que, por definição, pivôs são diferentes de 0. Nesse caso, o sistema pode:

(i) Não ter nenhuma solução: por exemplo, se as duas últimas equações forem $3w = 5$ e $4w = 17$.

(ii) Ter um número infinito delas: por exemplo, se $3w = 6$ e $5w = 10$, encontramos $w = 2$, mas substituindo isso na 1ª equação não nos permite decidir o que são u e v.

DECOMPOSIÇÃO LU

Vejamos novamente o que fizemos para resolver o sistema linear:

$$\begin{cases} 2u + v + w = 5 \\ 4u - 6v = -2 \\ -2u + 7v + 2w = 9 \end{cases}$$

(a) Subtraia $\underline{2} \times$ (1ª eq) da 2ª eq.

(b) Subtraia $\underline{(-1)} \times$ (1ª eq.) da 3ª eq.

$$\begin{cases} 2u + v + w = 5 \\ -8v - 2w = -12 \\ 8v + 3w = 14 \end{cases}$$

(c) Subtraia $\underline{(-1)} \times$ (2ª eq.) da 3ª eq.

$$\begin{cases} 2u + v + w = 5 \\ -8v - 2w = -12 \\ w = 2 \end{cases}$$

Os números $\frac{4}{2} = 2$, $\frac{-2}{2} = -1$, $\frac{8}{-8} = -1$ são chamados <u>MULTIPLICADORES</u> e serão importantes no que faremos hoje.

Usando matrizes, o primeiro passo é

$$\left[\begin{array}{ccc|c} 2 & 1 & 1 & 5 \\ 4 & -6 & 0 & -2 \\ -2 & 7 & 2 & 9 \end{array}\right] \xrightarrow{(a)} \left[\begin{array}{ccc|c} 2 & 1 & 1 & 5 \\ 0 & -8 & -2 & -12 \\ -2 & 7 & 2 & 9 \end{array}\right]$$

isto é, mantemos a primeira e terceira linhas e subtraímos $2 \times$ (1ª linha) da 2ª linha.

Se lembrarmos que multiplicação à esquerda por uma matriz E usa os coeficientes das linhas de E para fazer combinações lineares das linhas da matriz sendo multiplicada, vemos que multiplicar pela matriz E a seguir realiza (a) acima:

$$E = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{array}{l} \longleftarrow \text{Mantenha a 1ª linha} \\ \longleftarrow \text{Some } (-2)\times(\text{1ª linha}) \text{ a } 1\times(\text{2ª linha}). \\ \longleftarrow \text{Mantenha a 3ª linha} \end{array}$$

$$\begin{bmatrix} 1 & & \\ -2 & 1 & \\ & & 1 \end{bmatrix} \left[\begin{array}{ccc|c} 2 & 1 & 1 & 5 \\ 4 & -6 & 0 & -2 \\ -2 & 7 & 2 & 9 \end{array}\right] = \left[\begin{array}{ccc|c} 2 & 1 & 1 & 5 \\ 0 & -8 & -2 & -12 \\ -2 & 7 & 2 & 9 \end{array}\right]$$

Da mesma maneira, se multiplicarmos à esquerda pelas matrizes F e G abaixo, vamos realizar os passos (b) e (c):

$$F = \begin{bmatrix} 1 & & \\ & 1 & \\ 1 & & 1 \end{bmatrix} \leftarrow$$ Mantenha linhas 1 e 2 e some linha 1 à linha 3.

$$G = \begin{bmatrix} 1 & & \\ & 1 & \\ & 1 & 1 \end{bmatrix} \leftarrow$$ Mantenha linhas 1 e 2 e some linha 2 à linha 3.

$$\begin{bmatrix} 1 & & \\ & 1 & \\ 1 & & 1 \end{bmatrix} \cdot \left[\begin{array}{ccc|c} 2 & 1 & 1 & 5 \\ 0 & -8 & -2 & -12 \\ -2 & 7 & 2 & 9 \end{array}\right] = \left[\begin{array}{ccc|c} 2 & 1 & 1 & 5 \\ 0 & -8 & -2 & -12 \\ 0 & 8 & 3 & 14 \end{array}\right]$$

$$\begin{bmatrix} 1 & & \\ & 1 & \\ & 1 & 1 \end{bmatrix} \left[\begin{array}{ccc|c} 2 & 1 & 1 & 5 \\ 0 & -8 & -2 & -12 \\ 0 & 8 & 3 & 14 \end{array}\right] = \left[\begin{array}{ccc|c} 2 & 1 & 1 & 5 \\ 0 & -8 & -2 & -12 \\ 0 & 0 & 1 & 2 \end{array}\right]$$

É importante observar que podemos separar a matriz A e o vetor b à direita do sistema original:

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix} \quad \text{e} \quad b = \begin{bmatrix} 5 \\ -2 \\ 9 \end{bmatrix}.$$

E que multiplicar ambos à esquerda por E, F e G, nesta ordem, produz o resultado acima, separado em uma matriz 3x3 e um vetor coluna

$$U = \begin{bmatrix} 2 & 1 & 1 \\ & -8 & -2 \\ & & 1 \end{bmatrix} \quad \text{e} \quad c = \begin{bmatrix} 5 \\ -12 \\ 2 \end{bmatrix}$$

Isto é,

$$GFEA = U \quad \text{e} \quad GFEb = c.$$

Mas mais relevante do que saber que $GFEA = U$ é saber como voltar de U para A. Para ir de A a U, primeiro "fazemos" E, depois fazemos F e, por fim, G. Assim, para voltar de U a A, devemos começar "desfazendo" G, depois desfazendo F e, por fim, desfazendo E. Mas o que "desfazer" o que essas matrizes fazem? E é responsável por somar $2\times$(linha 1) à linha 2. Desfazer esse processo é subtrair $2\times$(linha 1) da linha 2:

$$E^{-1} = \begin{bmatrix} 1 & & \\ 2 & 1 & \\ & & 1 \end{bmatrix} \qquad E = \begin{bmatrix} 1 & & \\ -2 & 1 & \\ & & 1 \end{bmatrix}$$

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

Do mesmo modo, obtemos as inversas de F e G:

$$F^{-1} = \begin{bmatrix} 1 & & \\ & 1 & \\ -1 & & 1 \end{bmatrix} \quad \text{e} \quad G^{-1} = \begin{bmatrix} 1 & & \\ & 1 & \\ & -1 & 1 \end{bmatrix}$$

Repetindo o que já dissemos, para voltar de U até A, devemos fazer com U o inverso do que fizemos com A, na ordem reversa, isto é, esperamos que

$$U = GFEA \iff A = E^{-1}F^{-1}G^{-1}U$$

Verifique que isso, de fato, acontece, isto é, calcule.

Mas vale a pena observar que os números que aparecem fora da diagonal nas matrizes E^{-1}, F^{-1}, G^{-1} são exatamente os três multiplicadores que usamos na eliminação.

Além disso, calculemos

$$F^{-1}G^{-1} = \begin{bmatrix} 1 & & \\ & 1 & \\ -1 & & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & & \\ & 1 & \\ & -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & & \\ & 1 & \\ -1 & -1 & 1 \end{bmatrix}$$

$$E^{-1}(F^{-1}G^{-1}) = \begin{bmatrix} 1 & & \\ 2 & 1 & \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & & \\ & 1 & \\ -1 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & & \\ 2 & 1 & \\ -1 & -1 & 1 \end{bmatrix} = L$$

A matriz L é triangular inferior, com 1's na diagonal e com os multiplicadores abaixo da diagonal, na ordem em que foram usados.

Isto é, acabamos de mostrar que

$$\boxed{A = L \cdot U}$$

onde L é triangular inferior com as propriedades acima e U é triangular superior, obtida ao final da eliminação.

Além disso, como $GFEb = c$, $E^{-1}F^{-1}G^{-1}c = b$, isto é $\boxed{Lc = b.}$

Isso quer dizer que a decomposição $A = LU$ nos permite trocar o sistema $Ax = b$ por dois sistemas triangulares:

$$\boxed{Lc = b \quad \text{e} \quad Ux = c}$$

já que as duas equações acima implicam

$$b = Lc = LUx = Ax.$$

<u>Exemplo</u>:

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

$$Ax = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \iff$$

$$Lc = b \quad : \quad \begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & -1 & 1 & \\ & & -1 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \Rightarrow \begin{array}{l} c_1 = 1 \\ c_2 = 2 \\ c_3 = 3 \\ c_4 = 4 \end{array} \iff c = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

$$Ux = c \quad : \quad \begin{bmatrix} 1 & -1 & & \\ & 1 & -1 & \\ & & 1 & -1 \\ & & & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \Rightarrow \begin{array}{l} x_1 = 10 \\ x_2 = 9 \\ x_3 = 7 \\ x_4 = 4 \end{array} \iff x = \begin{bmatrix} 10 \\ 9 \\ 7 \\ 4 \end{bmatrix}.$$