ALGEBRA LINEAR

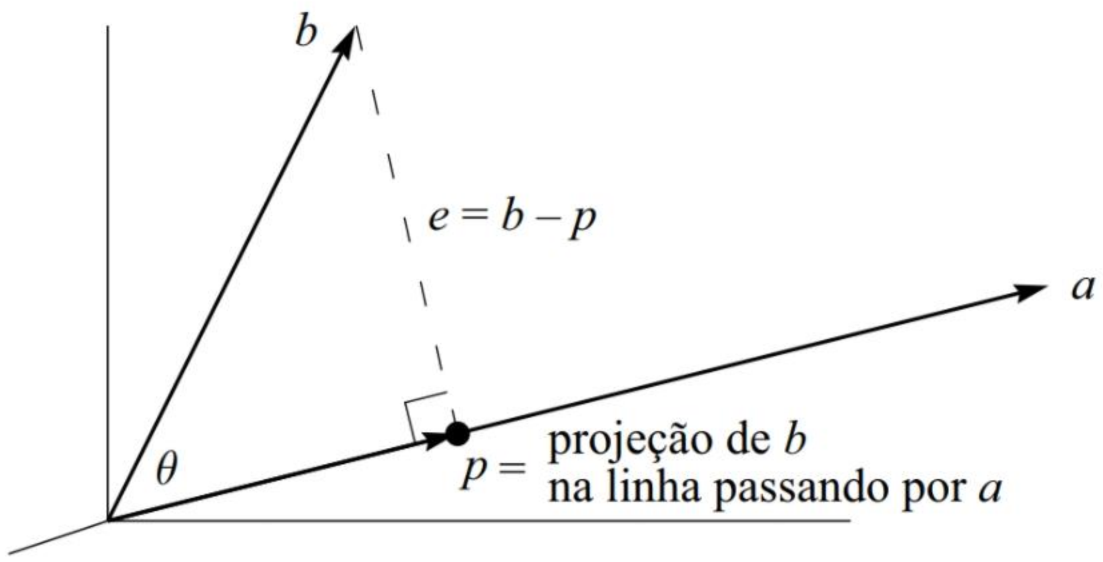
Prof. Dr. Patricio R. Impinnisi

Aula 5

Projeções e mínimos quadrados

Projeções em retas

Vamos ver agora **projeções de vetores**.
Vamos supor que queremos descobrir a MENOR distância de um ponto b até a reta na direção do vetor a. Ou seja, procuramos nessa reta pelo ponto p que esteja mais próximo de b.
A geometria nos diz que a menor distância será a reta “e” (linha pontilhada na figura) que é perpendicular à direção do vetor a e que conecta b a p.
Desta forma, a projeção do vetor b na direção do vetor a nos permite descobrir o ponto p mais próximo de b. Como vemos, o problema da menor distância automaticamente invoca a ortogonalidade...



Projeções em retas

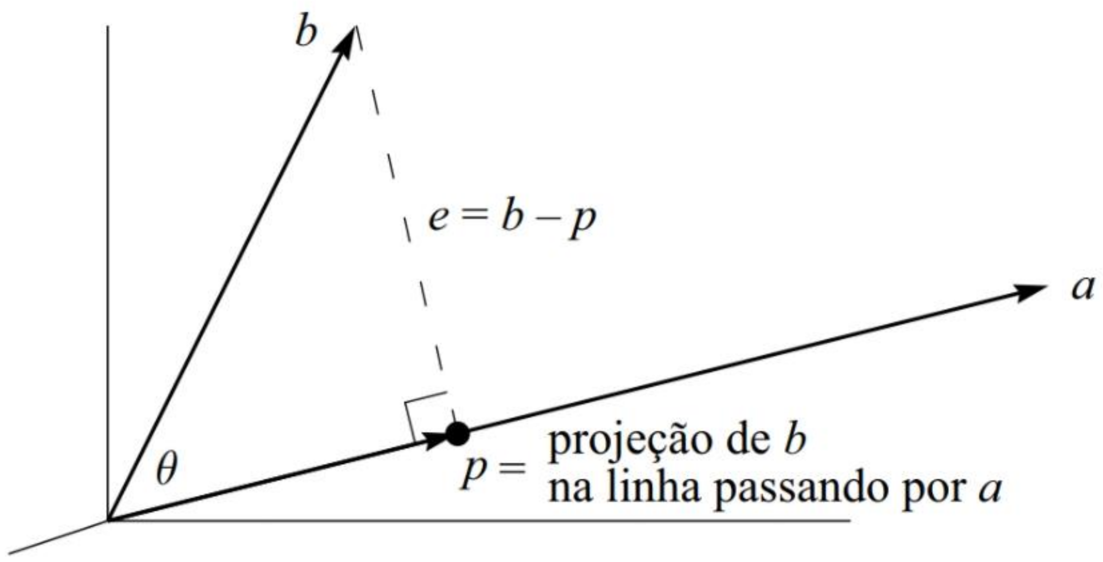
Vamos ver agora **projeções de vetores**.

A situação é a mesma quando se considera um plano, ou qualquer outro subespaço S, no lugar da reta a da figura.

O problema será sempre descobrir, nesse subespaço, o ponto p que esteja mais próximo de b.

E este ponto p será sempre obtido pela projeção de b nesse subespaço (através de uma linha perpendicular ao subespaço S que passa por b)

Por que é importante estudar o problema das projeções?



Projeções em retas

Porque este é, exatamente, o problema das soluções com mínimos quadrados para um sistema **sobredimensionado** (mais equações do que incógnitas).

Vejamos um exemplo simples...

Imagine 3 medidas experimentais (que contem erros devido a ruídos, etc.) e queremos encontrar a linha reta que melhor representa (fitting) essas três medidas...

Procuramos pela reta da figura...

... com o menor erro possível (barras verticais vermelhas), ou seja, com a menor distância aos pontos experimentais.

A linha reta é representada pela função:

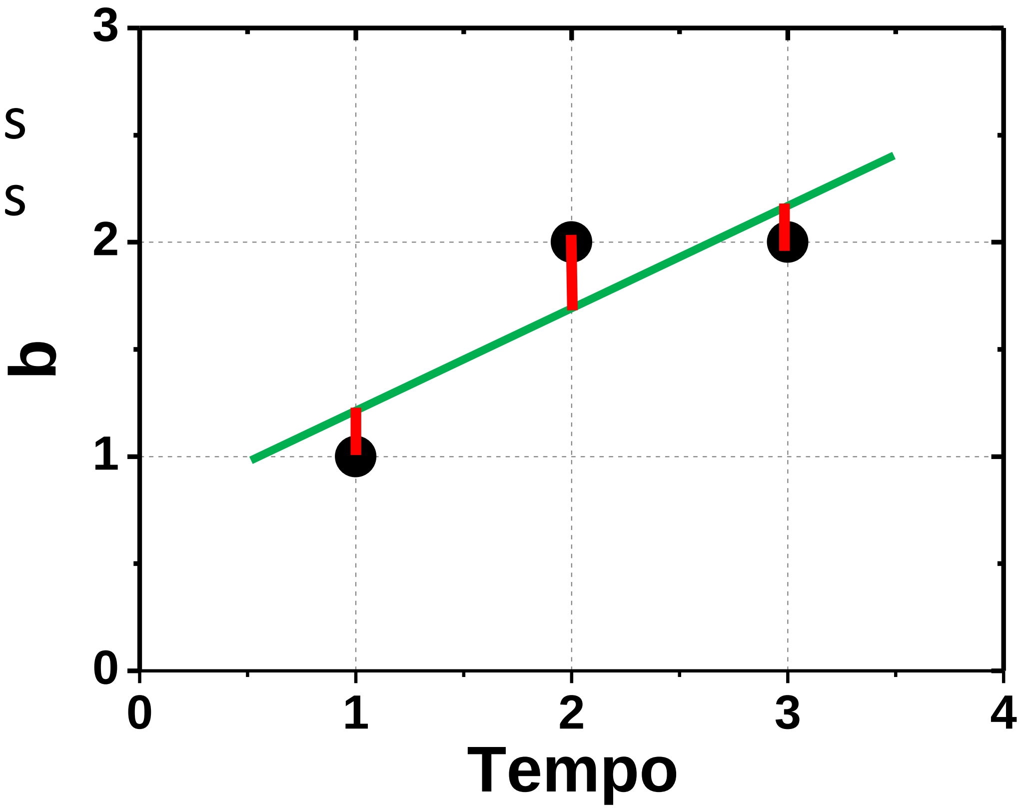
$$\mathbf{b} = C + \mathbf{t}\, D$$

Precisamos encontrar os números C e D para o qual temos 3 equações (dos três pontos)

$$\mathbf{1} = C + \mathbf{1}\, D$$

$$\mathbf{2} = C + \mathbf{2}\, D$$

$$\mathbf{2} = C + \mathbf{3}\, D$$



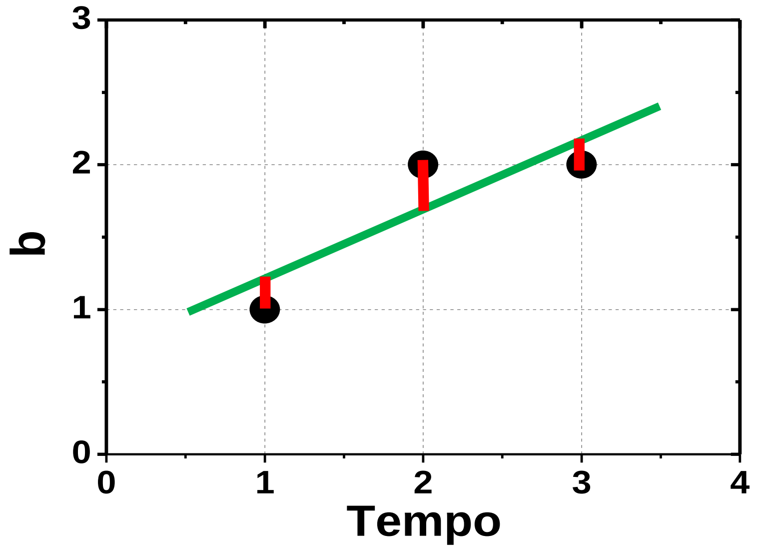
Projeções em retas

Em termos de matrizes teremos:

C + **1** D = 1
C + **2** D = 2
C + **3** D = 2

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

A x b



Evidentemente este sistema não tem solução! (b não está no espaço coluna de A, lembre os casos em que m>n, que em geral não tem solução!)

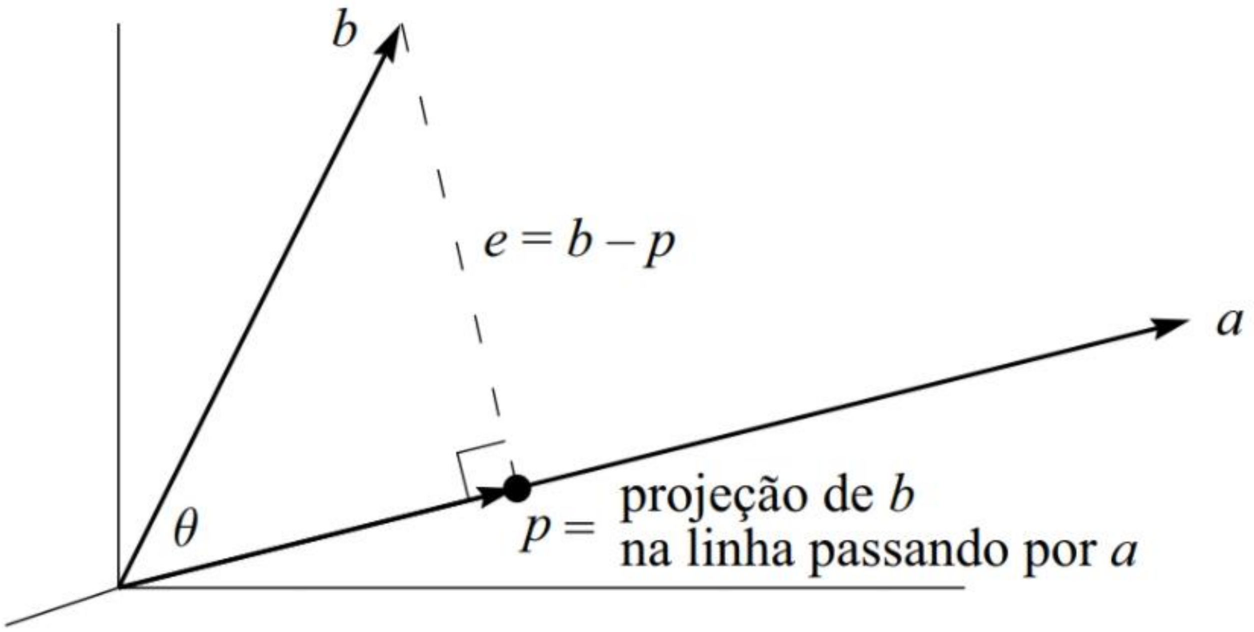
O sistema não tem solução pois nenhuma reta passa pelos três pontos, mas podemos encontrar a melhor solução possível, com o menor erro (que é a reta verde da figura)

Vamos desenvolver a teoria para resolver este tipo de problema muito comum (por exemplo em circuitos elétricos aplicando as leis de Kirchoff) e depois retornaremos a nosso problema dos três pontos...

Projeções em retas

O ponto central para desenvolver a teoria é que o menor erro (ou seja o vetor "e" da figura) **é ortogonal** ao subespaço determinado pelo vetor "a" (que representa a reta que vai aproximar os pontos experimentais).

Esta condição de perpendicularidade é que vai fornecer a equação que vai resolver o problema!



Vemos pela figura que p é um múltiplo de a, ou seja, podemos escrever $p = x\,a$ em que x aqui é um número.

O fato fundamental vem da ortogonalidade entre "e" e "a", ou seja podemos escrever:

$$a^T(b-xa)=0$$

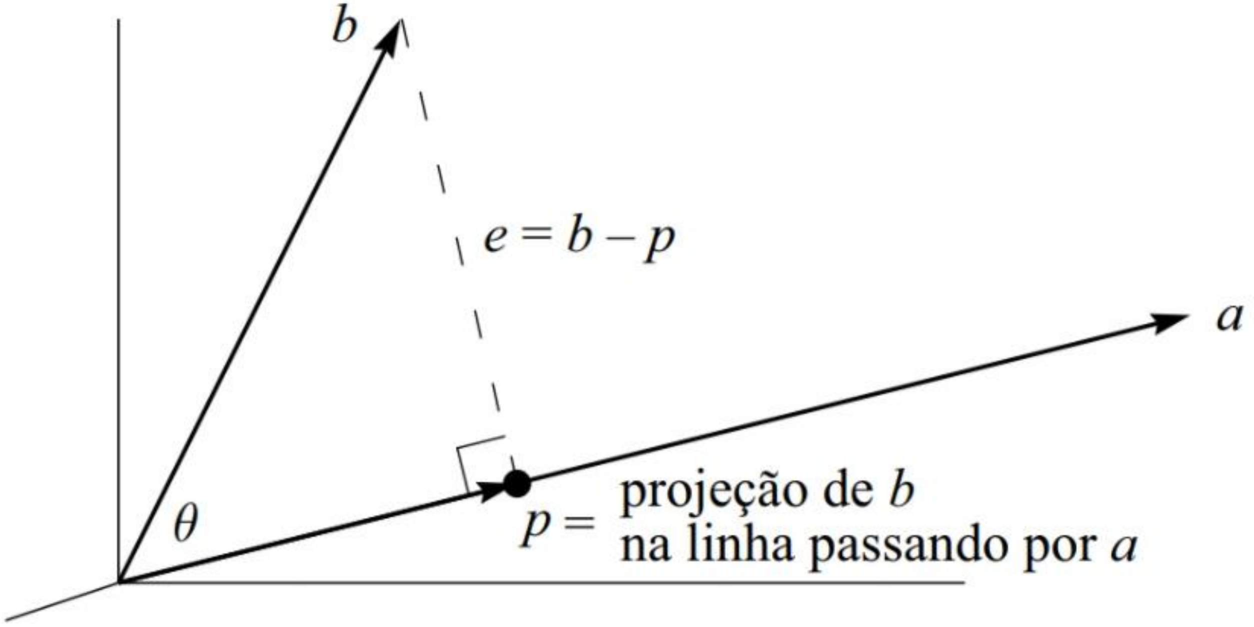
Rearranjando os termos podemos escrever (lembrando que x é um número):

$$x\,a^Ta = a^Tb$$

Projeções em retas

Evidentemente na equação $x\, a^T a = a^T b$, o termo $a^T a$ é um número (o módulo do vetor "a" elevado ao quadrado) e $a^T b$ também é um número (produto escalar de a e b) ...desta equação podemos encontrar x:

$$x = \frac{a^T b}{a^T a}$$



Desta forma, a projeção $p = a\,x$ fica:

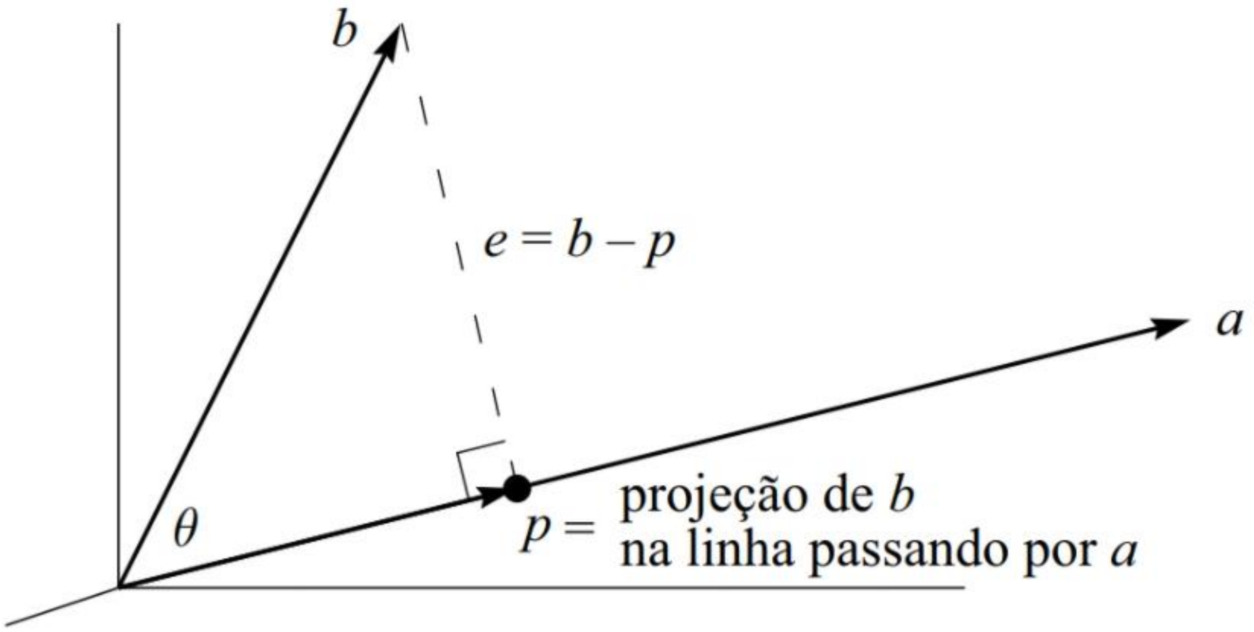
$$p = a\frac{a^T b}{a^T a}$$

O que acontece se duplicamos o vetor b?...a projeção duplica (da para ver isso na equação)!
Se duplicamos o vetor e nada acontece com a projeção ("e" nem figura na equação)...
E se duplicamos o vetor a? (olhar a equação) qualquer mudança de cancela ("a" aparece no numerado e no denominador) e nada muda com a projeção.

Projeções em retas

Se analisamos a equação da projeção...

$$p = a\frac{a^T b}{a^T a}$$



vemos que ela é fruto da operação de uma matriz! que chamamos *P (matriz projeção)*

$$p = P\,b$$

Em que a matriz projeção é: $\boxed{P = \frac{a\,a^T}{a^T a}}$

Evidentemente no numerador temos uma matriz (lembrar produto de vetor coluna vezes vetor linha) e no denominador temos um número, por tanto P é uma matriz.
Vamos ver as propriedades desta matriz...

Projeções em retas

$$P = \frac{a\,a^T}{a^T a}$$

Qual é o espaço coluna da matriz P?

Lembrar quer quando multiplicamos uma matriz A por qualquer coisa sempre ficamos no espaço coluna dela (pois por definição Ax é o espaço coluna de A, CL das colunas de A!)
No nosso caso, se multiplicamos P vezes b estaremos no espaço coluna de P que é...(em qual espaço sempre terminamos "aterrizando" ao projetar?)
...terminamos sempre no espaço definido pela reta que passa pelo ponto "a"....então

C(P)= linha que passa pelo ponto "a"

Qual é o posto (Rank) dessa matriz P?

Se o espaço coluna é uma linha reta só pode ser 1!!!

Posto (P) = 1

A matriz P é simétrica? (por que?)

$$\left(\frac{a\,a^T}{a^T a}\right)^T = \frac{a\,a^T}{a^T a} \qquad (P)^T = P$$

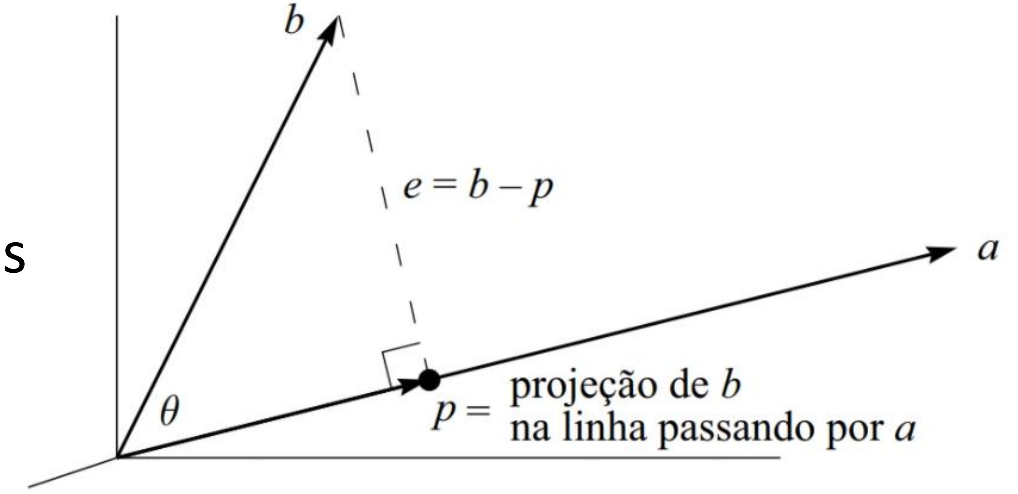
Projeções em retas

$$P = \frac{a\,a^T}{a^T a}$$

A última propriedade é a mais importante!!!

O que acontece se aplicamos a projeção duas vezes...?

$P^2 = P$



Resumindo, para projeções em linhas, as três fórmulas mais importantes são:

$$p = ax \qquad x = \frac{a^T b}{a^T a} \qquad P = \frac{a\,a^T}{a^T a}$$

Antes de retornar a nosso problema dos mínimos quadrados do inicio desta aula, vamos obter estas três formulas no caso de espaços com mais dimensões do que a reta que vimos neste caso (reta “a”).

Exercícios

Livro Álgebra Linear e suas aplicações. Gilbert Strang. 4 edição

Página 157

Conjunto de problemas 3.2

Resolver: 2; 5; 8; 11; 19; 20; 21; 24; 26

Projeções

Vamos projetar agora em planos (R^2), no espaço R^3 ou no espaço R^n em geral...

Vamos lembrar que as projeções vão ajudar a resolver sistemas de equações Ax=b que não tem solução (quando b não está no espaço coluna de A) e queremos a melhor solução possível (como no exemplo do inicio da aula)
O que faremos é projetar b no espaço coluna (que é Ax) ou seja, vamos resolver:

$$A\hat{x} = p$$

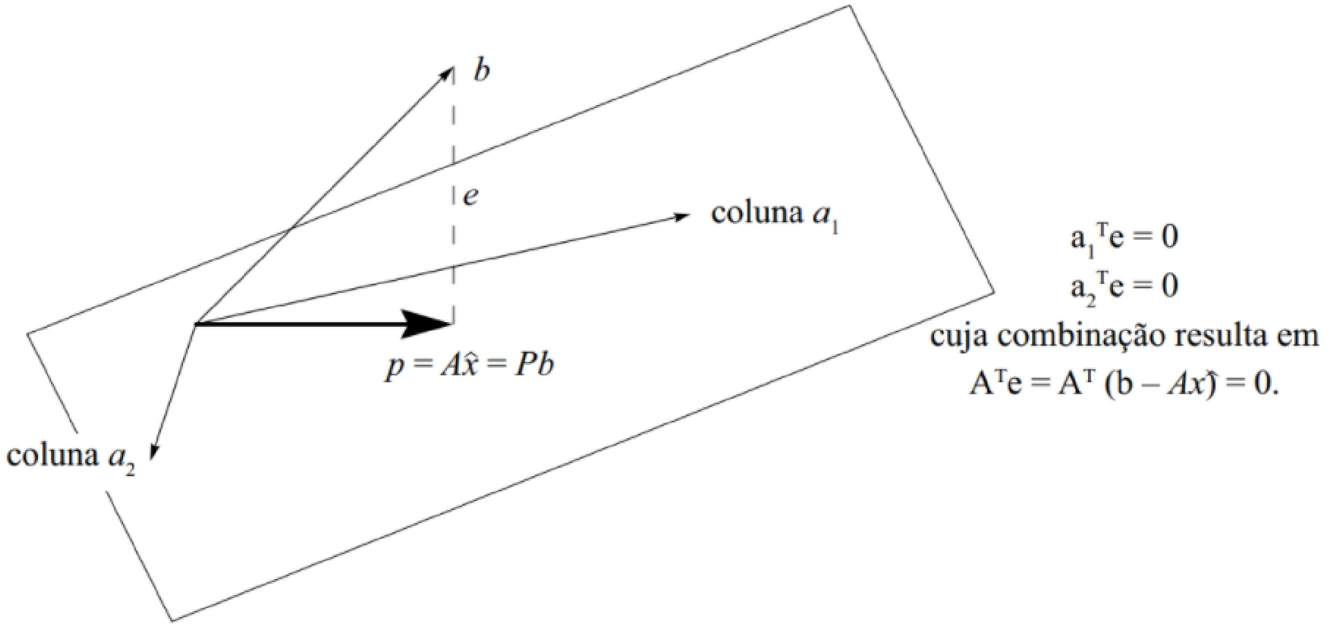
no lugar de Ax=b!
Os "x" não são os mesmos (o x de Ax=b nem existe) por isso diferenciamos eles.
Desta forma meu problema é encontrar a projeção de b no espaço coluna de A (que é o vetor p)

Vamos aplicar o procedimento num exemplo tridimensional...

Projeções

Vamos supor um espaço tridimensional, com um plano (gerado pelos dois vetores coluna independentes a_1 e a_2) sendo o espaço coluna de A e um vetor b que não está neste plano...e vamos projetar b nesse plano.

Primeiro temos que definir o espaço coluna ...ele é um plano definido por a_1 e a_2 (uma base qualquer do plano)



Qual é a matriz que tem este espaço coluna? $A = [a_1 \quad a_2]$

Logo saberíamos montar a matriz A para n dimensões (utilizando uma base desse espaço)

Agora temos que projetar b neste espaço e encontrar o vetor p

Ao utilizar o vetor p no lugar de b teremos um erro dado pelo vetor e = b – p que é perpendicular ao plano

Qual é o vetor p?

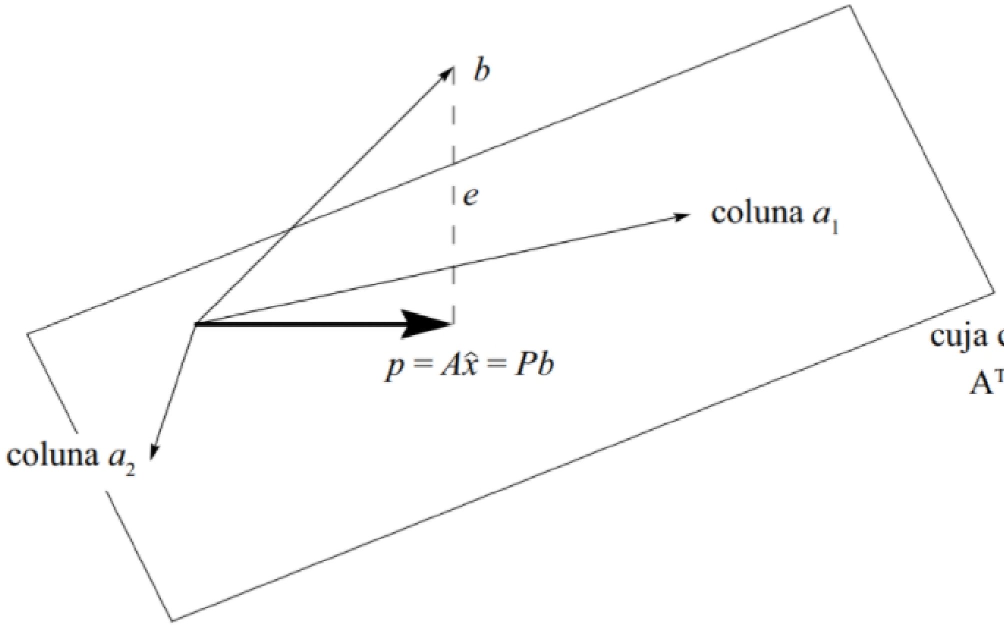
Projeções

Sabemos que p vai ser alguma CL das bases do plano (x_1e x_2 são números), ou seja:

$$p = \hat{x}_1 a_1 + \hat{x}_2\, a_2$$

ou escrito na forma matricial teremos:

$$p = \hat{x}_1\, a_1 + \hat{x}_2\, a_2 = A\hat{x}$$



$a_1^T e = 0$
$a_2^T e = 0$
cuja combinação resulta em
$A^T e = A^T(b - A\hat{x}) = 0.$

Nosso problema é encontrar $\hat{x}$

...ou seja, a CL certa de a_1 e a_2 para que o erro seja perpendicular ao plano (menor erro possível), assim:

(b—A$\hat{x}$) é perpendicular ao plano

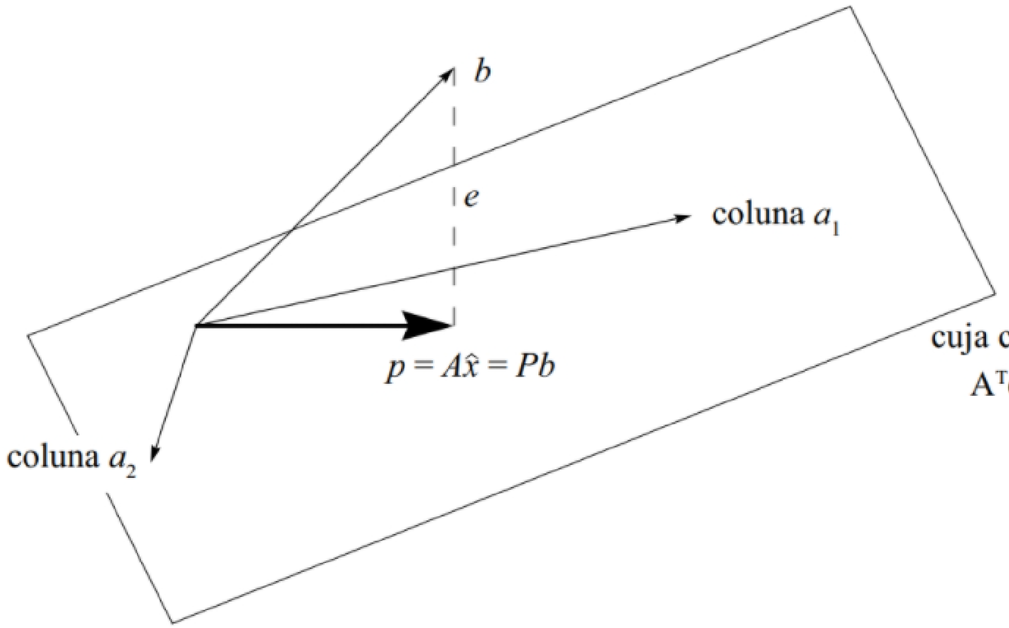
Temos duas incógnitas... a_1 e a_2 ... e quantas equações?
Temos também duas equações, pois b tem que ser perpendicular a “a_1” e a “a_2” de onde obtemos as duas equações que precisamos...

Projeções

Da ortogonalidade entre $(b - A\hat{x})$ (vetor "e")
e o plano gerado por a_1 e a_2 obtemos:

$$a_1^T(b - A\hat{x}) = 0$$

$$a_2^T(b - A\hat{x}) = 0$$



Na forma matricial teremos:

$$\begin{bmatrix} a_1^T \\ a_2^T \end{bmatrix} (b - A\hat{x}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Ou, o que é o mesmo: $\boxed{A^T(b - A\hat{x}) = 0}$

O termo $(b - A\hat{x})$ (que é o vetor e) é um dos quatro subespaços fundamentais! Qual?
O espaço nulo à esquerda ou o $N(A^T)$
Finalmente reescrevemos nossa equação da seguinte forma: $\boxed{A^T A\hat{x} = A^T b}$

Projeções

Nesta equação $A^T A\hat{x} = A^T \mathrm{b}$, no exemplo unidimensional da reta, no lugar da matriz A tínhamos um único vetor a e o produto $a^T a$ era um número, $a^T b$ também era um número e x era a razão entre o vetor a e o vetor p (outro número)...mas agora mudou...

Vamos obter as três formulas como no caso unidimensional...e ver como ficam...

Lembrando que agora A é uma matriz nxn, ...como solucionar? Como deixar $\hat{x}$ em evidência?

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

E a projeção? $p = A\hat{x} = A(A^T A)^{-1} A^T \mathrm{b}$

No caso unidimensional tínhamos, no lugar de $A(A^T A)^{-1} A^T$ a expressão $aa^T/a^T a$, ou seja, em casos multidimensionais, no lugar de um número como é ($a^T a$) tivemos que utilizar a inversa $(A^T A)^{-1}$

Neste caso a matriz projeção é: P = $A(A^T A)^{-1} A^T$

Pergunta...se aplicamos a regra da inversa ao produto, teremos que $(A^T A)^{-1} = A^{-1}(A^T)^{-1}$...

e substituindo teremos P = $AA^{-1}(A^T)^{-1} A^T$ que é a matriz I (identidade)!!!

Alguma coisa está errada? Pois desta forma não projetamos nada...o que está errado?

Projeções

Vamos conferir as propriedades que esperamos da matriz P (simetria e $P^2=P$)

No caso da simetria temos que verificar que $P^T = P$

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

...pois A^TA é simétrica e não se altera ao fazer a transposta e permanece na mesma posição

A segunda propriedade é que $P^2=P$

Vejamos.... $P^2 = A(A^TA)^{-1}\cancel{A^TA(A^TA)}^{-1}A^T \;=\; A(A^TA)^{-1}A^T$

Qual é a projeção de um vetor em C(A), se ele é perpendicular ao espaço coluna?

Aplicando o operador projeção teremos:

$$Pb= A(A^TA)^{-1}\,A^Tb = 0$$

pois se b é perpendicular ao C(A) ele está no espaço nulo da A^T (nulo à esquerda) que é justamente definido como $A^Tb=0$!!!)

E se o vetor já está no C(A), o que deveria dar sua projeção?

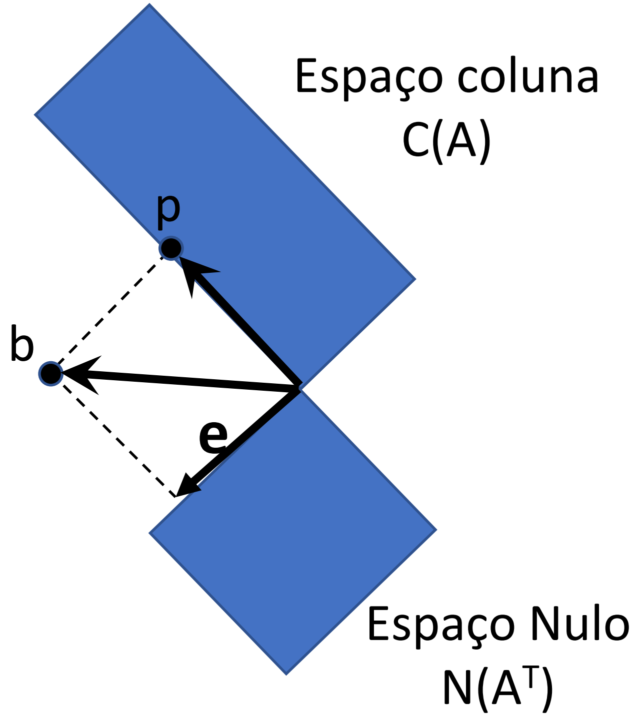
Se um vetor está no espaço coluna então sua projeção deveria dar o próprio vetor, vejamos:

$Pb = A(A^TA)^{-1}\,A^Tb$ = $= A(A^TA)^{-1}\,A^TAx = Ax = b$ (pois se b está no C(A) ele pode ser escrito como Ax!!! Logo a inversa se cancela e dá o próprio vetor, como esperado)...

Vejamos isto graficamente...

Projeções

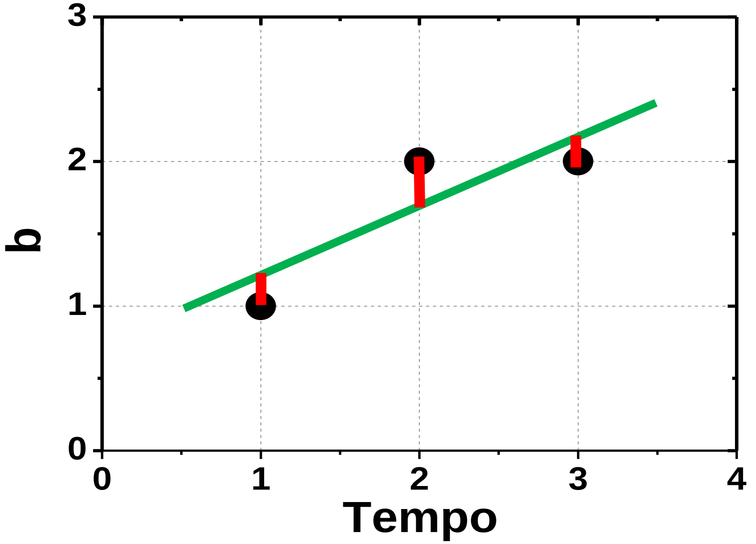
A partir da matriz A obtemos C(A) e N(A^T)
Em geral b não está no C(A) (sem solução)
Então decompomos b em “p” e “e” de forma que p+e=b
Isto é o que a matriz projeção P faz, Pb=p
O vetor “e” também é uma projeção, qual é a matriz projeção que da o vetor “e”?
É (I-P) pois I da o próprio vetor, e P da a projeção em C(A) logo a parte que falta é dada por esta matriz diferença.



Ok, assim completamos as três fórmulas para o caso multidimensional...

$p = A\hat{x}$ $\qquad \hat{x} = (A^TA)^{-1}A^T\mathrm{b}$ $\qquad P = A(A^TA)^{-1}A^T$

Vamos agora retomar a nosso problema inicial...dos quadrados mínimos...tínhamos medidas experimentais que queremos aproximar por uma reta com o menor erro possível...e tínhamos chegado a um sistema de equações que é reescrito a seguir:

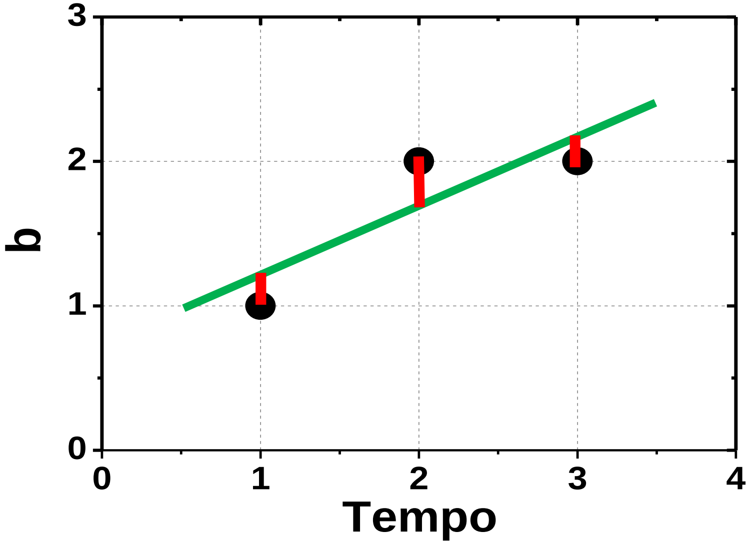


Projeções

Em termos de matrizes tínhamos:

C + **1** D = 1
C + **2** D = 2
C + **3** D = 2

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$



Como sabemos este sistema não tem solução....e queremos a melhor reta que aproxima estes pontos (minimizar o erro).

O substituir Ax=b, (que não tem solução), pela equação A$\hat{x}$=p estaremos cometendo um erro...o erro é (A$\hat{x}$-b) que é o vetor "e" (com seus componentes, neste caso três componentes e_1, e_2 e e_3), que é o que queremos minimizar.

Minimizar é diminuir seu tamanho, ou seja, minimizar ‖ A$\hat{x}$ -b ‖, que de fato significa minimizar seu quadrado (pois os erros sempre se somam, nunca são negativos), ou seja vamos minimizar ‖ A$\hat{x}$ -b ‖^2

Para nosso caso com 3 pontos teremos 3 erros (3 distâncias verticais) e o erro total será a soma dos quadrados de cada um, ou seja:

$$\| A\hat{x} - b \| = e_1^2 + e_2^2 + e_3^2$$

Projeções

Para resolver esta regressão linear vamos escrever o erro da seguinte forma...sabemos que b_1 b_2 e b_3 são as medidas experimentais, e p_1 p_2 e p_3 a suas projeções na reta.

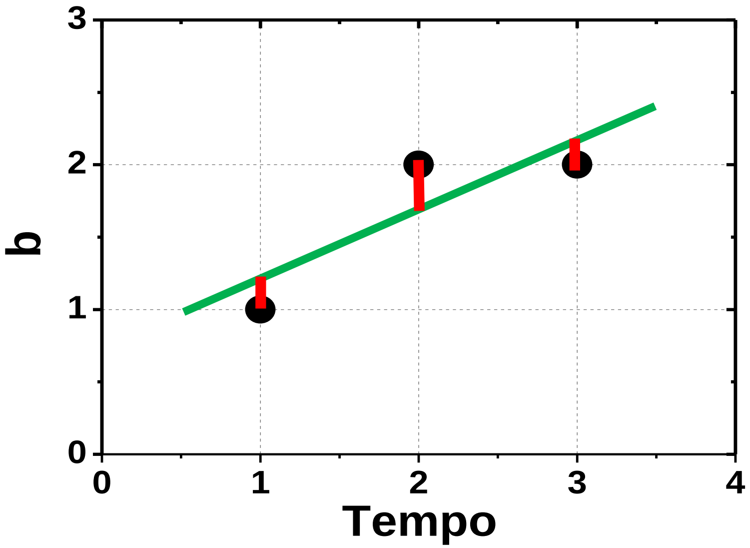
Desta forma o erro é a diferença entre cada b_i e cada p_i.

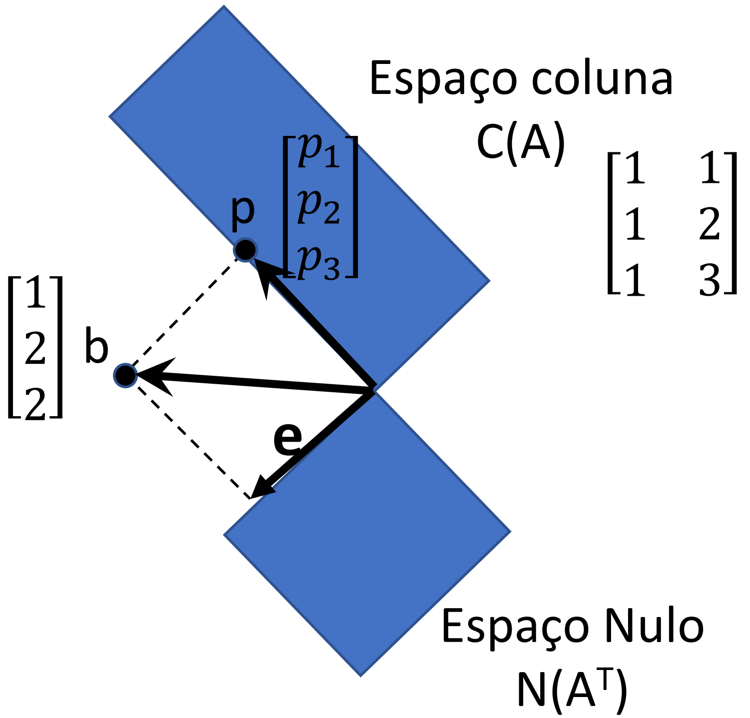
Cada ponto p_i (de fato vetor p_i) está no espaço coluna de A (pois agora podemos resolver a equação $A\hat{x}$=p) e esta CL dos p_i é o ponto mais próximo (menor erro) entre o espaço coluna de A e o vetor b original para o qual não há solução (ver figura)

No nosso caso teremos nosso b e nosso C(A)

E queremos encontrar esse vetor p

Portanto precisamos encontrar: $\hat{x} = \begin{bmatrix} C \\ D \end{bmatrix}$ e p = $\begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}$





Vamos começar com $\hat{x}$ utilizando a equação (que desenvolvemos) e depois vamos obter o vetor p...

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

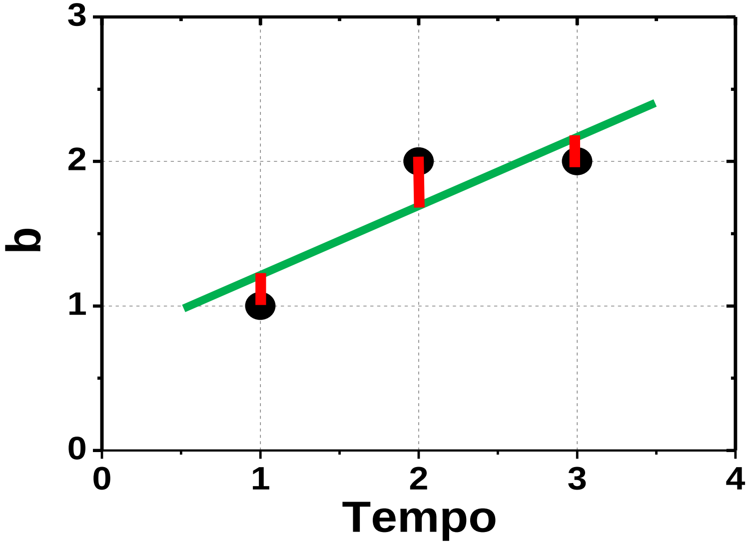
Projeções

Vamos resolver $A^T A\,\hat{x} = A^T$b

Vamos escrever essas matrizes para nosso caso e executar as operações...

$$A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}$$

$$A^T b = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 11 \end{bmatrix}$$

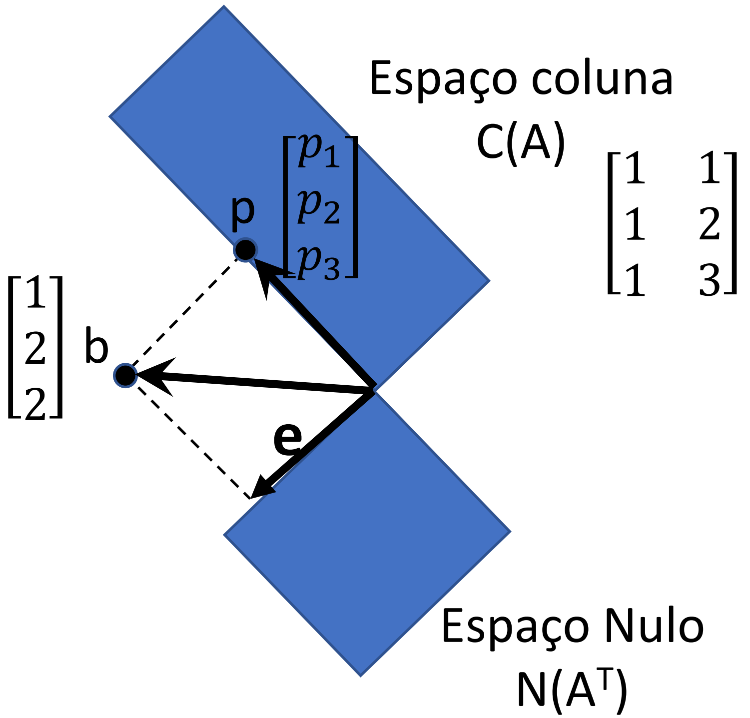


Portanto teremos as seguintes equações (chamadas normais) que minimizam o erro:

3C + 6D = 5
6C + 14D = 11

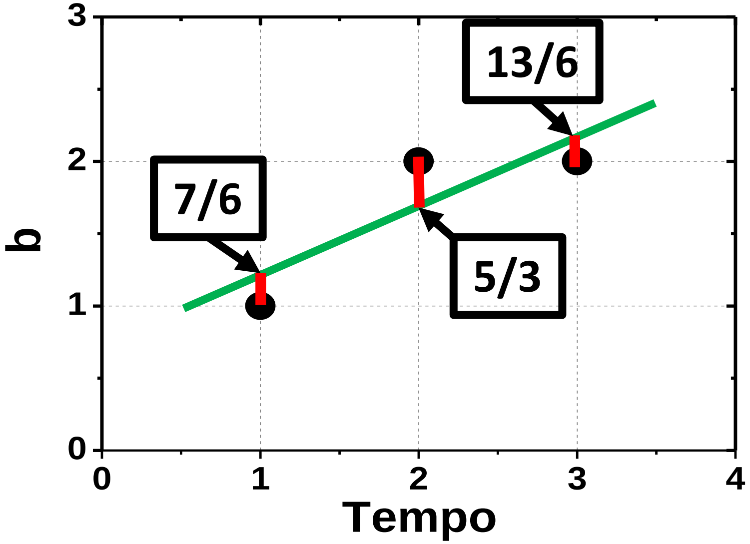
Podem checar se estas equações que a álgebra fornece de fato minimizaram o erro derivando respeito de C e D as equações explicitas dos erros (abaixo) e igualando a zero...

$\| A\hat{x} - b \| = e_1^2 + e_2^2 + e_3^2 = (C + D - 1)^2 + (C + 2D - 2)^2 + (C + 3D - 2)^2$



Projeções

Solucionando $\begin{aligned} 3C + 6D &= 5 \\ 6C + 14D &= 11 \end{aligned}$ Obtemos C = 2/3 e D=1/2



Portanto a melhor linha reta será $y = 2/3 + t/2$

Agora podemos encontrar o p_1, p_2 e p_3 substituindo as coordenadas nesta equação da linha reta...

$p_1 = 2/3 + 1/2 = 7/6$

$p_2 = 2/3 + 2/2 = 5/3$

$p_3 = 2/3 + 3/2 = 13/6$

Este vetor p está aqui!

Nossos vetores erro (linhas vermelhas) são:

$e_1 = b_1 - p_1 = 1 - 7/6 = -1/6$

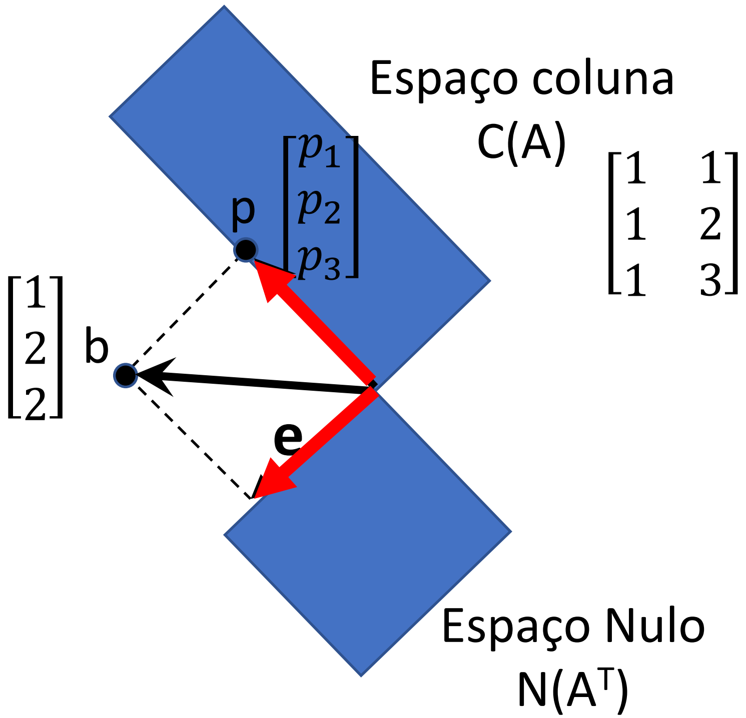
$e_2 = b_2 - p_2 = 2 - 5/3 = 2/3$

$e_3 = b_3 - p_3 = 2 - 13/6 = -1/6$

Este vetor erro está aqui!

Vamos checar se, como esperado, p + e = b

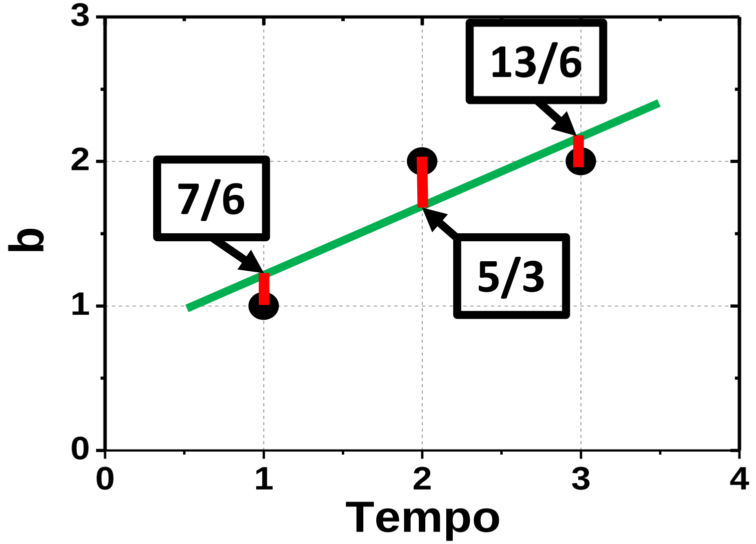
Vejamos....



Projeções

De fato a soma de p + e dá b...

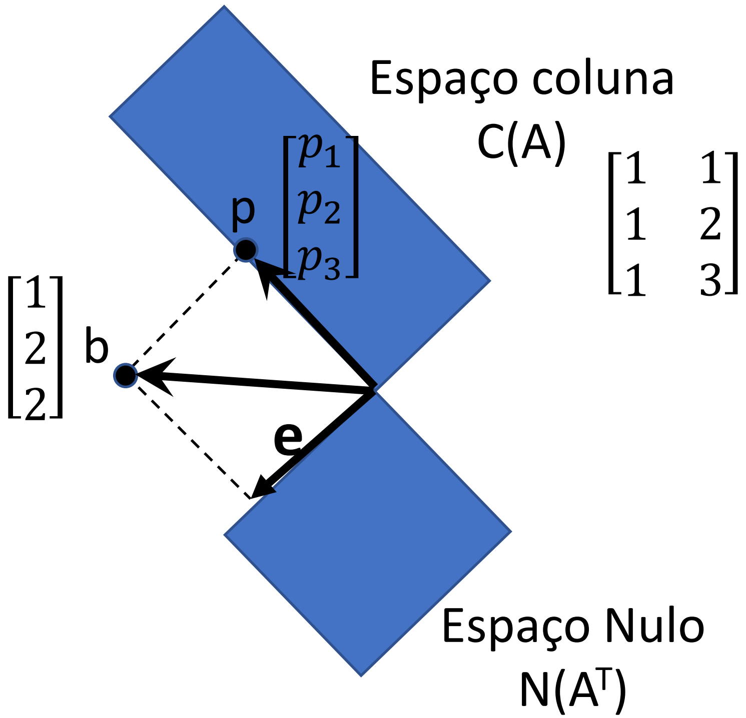
$$\begin{bmatrix}1\\2\\2\end{bmatrix} = \begin{bmatrix}7/6\\5/3\\13/6\end{bmatrix} + \begin{bmatrix}-1/6\\2/6\\-1/6\end{bmatrix}$$



Além dos vetores p e e somarem para dar b...Os vetores p e e tem que ser ortogonais...e de fato são!

$$\begin{bmatrix}7/6 & 10/6 & 13/6\end{bmatrix}\begin{bmatrix}-1/6\\2/6\\-1/6\end{bmatrix} = 0$$

O vetor "e" além de ser perpendicular ao vetor p, ele é perpendicular a todo vetor no C(A), inclusive aos vetores base deste espaço (1,1,1) e (1,2,3) ...podem verificar...



Projeções

O momento crucial deste desenvolvimento foi quando (implicitamente) assumimos que a matriz A^TA possui inversa!!! (senão não poderíamos ter resolvido)

No nosso exemplo foi a matriz

$$A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}$$

A questão é que: *sempre que uma matriz A tem colunas independentes, a matriz produto A^TA possui inversa*

É por isto que tudo funcionou bem...e **por ser importante** **vamos demonstrar** esta afirmação

Demonstração:

Para demostrar essa afirmação, precisamos demonstrar que o espaço nulo de A^TA (matriz quadrada) possui um único elemento, o vetor x=0.

Então temos que demonstrar que $A^TAx=0$ implica que x=0, se A tem colunas independentes!

Vamos multiplicar ambos lados da igualdade pelo vetor x^T

$$x^TA^TAx=0$$

Mas isso é $(Ax)^TAx=0$ e esse produto é o quadrado do módulo do vetor Ax ....e para ser zero implica que o vetor Ax=0 ....e como as colunas de A são independentes só é possível se x=0

Exercícios

Livro Álgebra Linear e suas aplicações. Gilbert Strang. 4 edição

Página 170

Conjunto de problemas 3.3

Resolver: 1; 4; 5; 9; 10; 12; 17; 19; 20; 25; 28; 34