Álgebra Linear

Um Livro Colaborativo

1 de Setembro de 2017

Organizadores

Diego Marcon Farias - UFRGS

Rafael Rigão Souza - UFRGS

Pedro Henrique de Almeida Konzen - UFRGS

Licença

Nota dos organizadores

Este livro é parte de nosso projeto de escrita colaborativa de recursos educacionais abertos. Nosso objetivo é produzir um material didático no nível de graduação de excelente qualidade e de acesso livre pela colaboração entre professores e alunos de universidades, institutos de educação e demais interessados em tópicos de álgebra linear.

O sucesso deste projeto depende da colaboração! Edite você mesmo o livro, dê sugestões ou nos avise de erros e imprecisões. Toda a colaboração é bem vinda. Saiba mais visitando o repositório oficial do livro:

https://fithub.com/livroscolaborativos/AlgebraLinear

Desejamos-lhe ótimas colaborações!

Prefácio

Em construção ... Caso deseja colaborar com a escrita deste livro, veja como em: https://github.com/livroscolaborativos/AlgebraLinear

Conteúdo

1

Semana 1

1.1 Sistemas Lineares

Começamos nosso estudo diretamente com um exemplo:

Exemplo 1. Considere o sistema linear que consiste de duas equações e duas variáveis

$$\begin{cases} x + 3y &= 1 \\ 2x - y &= -2 \end{cases} . \tag{1.1}$$

Nosso objetivo é descobrir o valor (ou todos os valores) das variáveis x e y que satisfazem a ambas equações.

Solução 1. Ao multiplicar toda a primeira equação por -2, não alteramos a solução do sistema, pois, caso quiséssemos voltar ao sistema original, bastaria dividir a primeira equação do sistema abaixo por -2. Ficamos assim com o sistema

$$\begin{cases} -2x - 6y &= -2 \\ 2x - y &= -2 \end{cases} .$$

Em seguida, somamos as duas equações e mantemos uma delas.

$$\begin{cases} -2x - 6y &= -2 \\ -7y &= -4 \end{cases} .$$

Este sistema também tem as mesmas soluções do anterior, pois podemos a qualquer hora retornar ao sistema original fazendo (linha 2) – (linha 1).

Em verdade, manter a equação da primeira linha multiplicada por -2 não ajuda em nada na resolução do sistema e voltaremos à equação original. Além disso, na linha dois já podemos isolar a variável y e descobrir o seu valor. Ficamos então com o sistema:

$$\begin{cases} x + 3y &= 1 \\ -7y &= -4 \end{cases} . \tag{1.2}$$

Observamos, finalmente, que a solução do último sistema (que é a mesma do sistema original) já está muito fácil de ser obtida. Sabemos o valor de y pela segunda equação e, para descobrir o valor de x, basta usar a primeira equação:

$$y = \frac{4}{7} \implies x + 3 \cdot \frac{4}{7} = 1 \implies x = 1 - \frac{12}{7} = -\frac{5}{7}.$$

Podemos escrever a solução do sistema como

$$\begin{cases} x &= -5/7 \\ y &= 4/7 \end{cases} \quad \triangleleft$$

O símbolo $\triangleleft$ acima nas nossas notas indica o fim de uma solução, exemplo ou observação. Em seguida, fazemos algumas observações.

Observação 2. O sistema linear como o acima possui duas equações e duas variáveis e por isto é conhecido como um sistema linear 2×2 (lê-se dois por dois). Mais geralmente, um sistema linear com m equações e n variáveis é conhecido como um sistema linear $m \times n$ (lê-se m por n).

Observação 3. Sistemas 2×2 são dos mais simples de resolver e por isso o método acima pode parecer desnecessariamente complicado. Mas, como veremos, este método pode ser aplicado para qualquer sistema. Logo, é essencial que se entenda completamente todos os seus passos para um bom acompanhamento do que está por vir.

Vamos analisar a solução apresentada acima mais de perto. Para chegar do sistema linear em (1.1) para o sistema em (1.2), fizemos uma sequência de operações que:

- *não alteram a solução do sistema linear original e que;*
- resumidamente, podem ser descritas como *adicionar um múltiplo da linha um na linha dois.*

Que a solução permanece a mesma pode ser justificado tal qual está na "Solução 1" acima.

Operações nas linhas de um sistema que não alteram o conjunto de soluções são chamadas de **operações elementares**. São as seguintes:

1. Multiplicar uma linha por uma constante;
2. Trocar duas linhas de lugar;
3. Somar um múltiplo de uma linha a outra linha.

Um artifício que torna todo o processo mais "automático" é o de formar uma tabela de números com os coeficientes do sistema:

$$\left\{\begin{array}{rcl} x+3y &=& 1 \\ 2x-y &=& -2 \end{array}\right. \quad \rightsquigarrow \quad \left[\begin{array}{cc|c} 1 & 3 & 1 \\ 2 & -1 & -2 \end{array}\right].$$

A tabela de números da direita é conhecida como a **matriz aumentada associada ao sistema linear**. Vamos nos referir às linhas da matriz associada como ℓ_1 e ℓ_2. Deste modo, "adicionar um múltiplo da linha 1 na linha dois" corresponde a fazer:

$$\left[\begin{array}{cc|c} 1 & 3 & 1 \\ 2 & -1 & -2 \end{array}\right] \xrightarrow{-2\ell_1+\ell_2 \text{ em } \ell_2} \left[\begin{array}{cc|c} 1 & 3 & 1 \\ 0 & -7 & -4 \end{array}\right].$$

Para este tipo de operação, olhamos para as colunas da matriz associada, fizemos mentalmente (ou em uma folha de rascunho) os cálculos envolvidos e preenchemos os valores na respectiva linha da nova matriz que está à direita:

$$(-2)\cdot 1 + 2 = 0, \qquad (-2)\cdot 3 + (-1) = -7, \qquad (-2)\cdot 1 + (-2) = -4.$$

Exemplo 4. Consideramos agora um sistema com três equações e três variáveis;

$$\left\{\begin{array}{l} x+2y+z=12 \\ x-3y+5z=1 \\ 2x-y+3z=10 \end{array}\right. .$$

Vamos diretamente escrever a matriz aumentada associada ao sistema, que é a matriz com os coeficientes da variável x na primeira coluna, os coeficientes da variável y na segunda coluna, os coeficientes de z na terceira coluna e os coeficientes independentes na última coluna:

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 12 \\ 1 & -3 & 5 & 1 \\ 2 & -1 & 3 & 10 \end{array}\right].$$

Em seguida, utilizamos operações elementares nas linhas como no exemplo anterior. Ao substituir $-\ell_1+\ell_2$ em ℓ_2, obtemos

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 12 \\ 0 & -5 & 4 & -11 \\ 2 & -1 & 3 & 10 \end{array}\right].$$

Ou seja, eliminamos o valor 1 da primeira posição da segunda linha (em outras palavras, eliminamos a variável x da segunda equação). Prosseguindo com a eliminação dos valores abaixo da diagonal principal da matriz, obtemos:

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 12 \\ 0 & -5 & 4 & -11 \\ 2 & -1 & 3 & 10 \end{array}\right] \xrightarrow{-2\ell_1+\ell_3 \text{ em } \ell_3} \left[\begin{array}{ccc|c} 1 & 2 & 1 & 12 \\ 0 & -5 & 4 & -11 \\ 0 & -5 & 1 & -14 \end{array}\right] \xrightarrow{-\ell_2+\ell_3 \text{ em } \ell_3} \left[\begin{array}{ccc|c} 1 & 2 & 1 & 12 \\ 0 & -5 & 4 & -11 \\ 0 & 0 & -3 & -3 \end{array}\right].$$

Voltando à notação original do sistema, obtemos

$$\begin{cases} x & +2y & +z & = 12 \\ & -5y & +4z & = -11 \\ & & -3z & = -3 \end{cases}$$

Por causa do seu formato, este sistema é dito estar em **forma triangular** ou ainda que o sistema é **triangular**.

Como já tínhamos percebido no nosso primeiro exemplo, sistemas triangulares são fáceis de resolver. Pela última equação concluimos que $z = 1$. Subsituindo este valor na segunda equação, obtemos

$$-5y + 4 = -11 \implies y = 3.$$

Finalmente, substituindo estes dois valores na primeira equação, obtemos

$$x + 2 \cdot 3 + 1 = 12 \implies x = 5.$$

Portanto, a solução para o sistema é $(x, y, z) = (5, 3, 1)$.$\triangleleft$

Esta técnica de resolver sistemas lineares é importante por ser aplicável a sistemas de qualquer ordem e é conhecida como **escalonamento** ou **eliminação Gaussiana**.

1.2 Formas escalonadas e formas escalonadas reduzidas

Vimos dois exemplos de como resolver sistemas lineares por escalonamento. Vamos agora introduzir uma terminologia, esperando tornar o método mais sistemático.

Dizemos que uma matriz está em **forma escalonada** quando

1. As linhas que contém apenas zeros estão abaixo das demais.
2. O primeiro elemento não nulo de uma linha, conhecido como **elemento líder**, está em uma coluna à direita do elemento líder da linha acima.

Estas duas condições *implicam* que os elementos que estão abaixo de elementos líder são todos iguais a zero.

Por exemplo,

$$\begin{bmatrix} 1 & a_0 & a_1 & a_2 & a_3 \\ 0 & 0 & 2 & a_4 & a_5 \\ 0 & 0 & 0 & -1 & a_6 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{1.3}$$

é uma matriz em forma escalonada. Notamos que os coeficientes a_1, a_2, a_3, a_4, a_5 e a_6 podem ser quaisquer números sem alterar o fato de a matriz estar em forma escalonada.

Outros exemplos, retirados livro do David C. Lay:

$$\begin{bmatrix} \blacksquare & a_0 & a_1 & a_2 \\ 0 & \blacksquare & a_3 & a_4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & \blacksquare & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 \\ 0 & 0 & 0 & \blacksquare & a_9 & a_{10} & a_{11} & a_{12} & a_{13} & a_{14} \\ 0 & 0 & 0 & 0 & \blacksquare & a_{15} & a_{16} & a_{17} & a_{18} & a_{19} \\ 0 & 0 & 0 & 0 & 0 & \blacksquare & a_{20} & a_{21} & a_{22} & a_{23} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \blacksquare & a_{24} \end{bmatrix}.$$

Quando uma matriz está em forma escalonada, as posições marcadas com $\blacksquare$ são chamadas de **posições de pivô**. Observe que, caso a matriz não esteja em forma escalonada, o elemento líder de uma linha pode não estar na posição de pivô. Dizemos também que uma coluna é uma **coluna pivô** quando a coluna possui uma posição de pivô. Por exemplo, na primeira matriz acima, as duas primeiras colunas são colunas pivô enquanto a terceira e a quarta não são.

Uma matriz estará em **forma escalonada reduzida** quando:

1. Está na forma escalonada;

2. Todos os elementos líder são iguais a 1 e são os únicos elementos não nulos das suas colunas.

Por exemplo,

$$\begin{bmatrix} 1 & 0 & b_1 & 0 & b_2 \\ 0 & 1 & b_3 & 0 & b_3 \\ 0 & 0 & 0 & 1 & b_4 \end{bmatrix}$$

é uma forma escalonada reduzida da matriz da fórmula (1.3) acima.

Quanto aos outros exemplos, temos

$$\begin{bmatrix} 1 & 0 & b_1 & b_2 \\ 0 & 1 & b_3 & b_4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & b_1 & 0 & 0 & 0 & b_5 & b_6 & 0 & b_8 \\ 0 & 0 & 0 & 1 & 0 & 0 & b_{11} & b_{12} & 0 & b_{14} \\ 0 & 0 & 0 & 0 & 1 & 0 & b_{16} & b_{17} & 0 & b_{19} \\ 0 & 0 & 0 & 0 & 0 & 1 & b_{20} & b_{21} & 0 & b_{23} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & b_{24} \end{bmatrix}.$$

Observação 5. A forma escalonada reduzida ainda não apareceu na seção anterior quando estávamos resolvendo sistemas, mas notamos que poderia ter sido utilizada para automatizar ainda mais o processo de resolução do sistema. Vamos retomar o Exemplo 4. Fazendo mais algumas operações elementares, é possível transformar a matriz aumentada do sistema em uma matriz em forma escalonada reduzida:

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 12 \\ 0 & -5 & 4 & -11 \\ 0 & 0 & -3 & -3 \end{array}\right] \xrightarrow{\ell_2 \div (-5) \text{ e } \ell_3 \div (-3)} \left[\begin{array}{ccc|c} 1 & 2 & 1 & 12 \\ 0 & 1 & -4/5 & 11/5 \\ 0 & 0 & 1 & 1 \end{array}\right]$$

$$\xrightarrow{\frac{4}{5}\ell_3 + \ell_2 \text{ em } \ell_2} \left[\begin{array}{ccc|c} 1 & 2 & 1 & 12 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 1 \end{array}\right] \xrightarrow{-\ell_3 + \ell_1 \text{ em } \ell_1} \left[\begin{array}{ccc|c} 1 & 2 & 0 & 11 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 1 \end{array}\right] \xrightarrow{-2\ell_2 + \ell_1 \text{ em } \ell_1} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 1 \end{array}\right].$$

A última matriz acima está em forma escalonada reduzida e está associada ao sistema

$$\begin{cases} x & & & = 5 \\ & y & & = 3 \\ & & z & = 1 \end{cases}$$

que, na verdade, já é a solução explícita do sistema original. $\lhd$

1.3 Algoritmo de escalonamento

Sistematicamente, encontramos a forma escalonada de uma matriz aplicando os seguintes passos:

1. Analisar a primeira coluna pivô – da esquerda para a direita, esta é a primeira coluna que possui algum elemento diferente de zero – e, se necessário, aplicar operações elementares para que o elemento da primeira linha (esta é a posição de pivô!) passe a ser diferente de zero;

2. A partir de operações elementares, eliminar todos elementos que estão abaixo do elemento na posição de pivô que obtivemos no Passo 1;

3. Desconsiderando por um momento a primeira linha, procuramos pela próxima coluna pivô – aquela que tem algum elemento não nulo *abaixo da primeira linha*. Se necessário, aplicar operações elementares para que, na segunda linha, o elemento passe a ser diferente de zero;

4. A partir de operações elementares, eliminar todos elementos que estão abaixo do elemento na posição de pivô que obtivemos no Passo 3;

5. Desconsiderando por um momento a primeira e a segunda linhas, procuramos pela próxima coluna pivô – aquela que tem algum elemento não nulo *abaixo da primeira linha e da segunda linha*. Se necessário, aplicar operações elementares para que, na segunda linha, o elemento passe a ser diferente de zero;

6. A partir de operações elementares, eliminar todos elementos que estão abaixo do elemento na posição de pivô que obtivemos no Passo 5 e ssim por diante.

Essencialmente, são apenas dois passos que se repetem até que se esgotem as colunas que possuem posição de pivô. Vejamos um exemplo.

Exemplo 6. Considere a matriz

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & -1 & 8 \\ 0 & 0 & -3 & -1 & 2 & -11 \\ 0 & 0 & 6 & 2 & -4 & 22 \\ 0 & 0 & -2 & 1 & 2 & -3 \end{bmatrix}.$$

Passo 1. A primeira coluna pivô é a terceira. Escolhemos um elemento não nulo da terceira coluna para ocupar a posição de pivô. Por exemplo, a segunda linha. Trocando a primeira e a segunda linhas de lugar (esta é uma operação elementar), obtemos:

$$\begin{bmatrix} 0 & 0 & 2 & 1 & -1 & 8 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -3 & -1 & 2 & -11 \\ 0 & 0 & 6 & 2 & -4 & 22 \\ 0 & 0 & -2 & 1 & 2 & -3 \end{bmatrix}.$$

Passo 2. Eliminar os elementos abaixo do 2 que está na posição de pivô da terceira coluna:

$$\begin{bmatrix} 0 & 0 & 2 & 1 & -1 & 8 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -3 & -1 & 2 & -11 \\ 0 & 0 & 6 & 2 & -4 & 22 \\ 0 & 0 & -2 & 1 & 2 & -3 \end{bmatrix} \xrightarrow{\frac{3}{2}\ell_1+\ell_3 \text{ em } \ell_3} \begin{bmatrix} 0 & 0 & 2 & 1 & -1 & 8 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 & 1 \\ 0 & 0 & 6 & 2 & -4 & 22 \\ 0 & 0 & -2 & 1 & 2 & -3 \end{bmatrix}$$

$$\xrightarrow{-3\ell_1+\ell_4 \text{ em } \ell_4} \begin{bmatrix} 0 & 0 & 2 & 1 & -1 & 8 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 & 1 \\ 0 & 0 & 0 & -1 & -1 & -2 \\ 0 & 0 & -2 & 1 & 2 & -3 \end{bmatrix} \xrightarrow{\ell_1+\ell_5 \text{ em } \ell_5} \begin{bmatrix} 0 & 0 & 2 & 1 & -1 & 8 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 & 1 \\ 0 & 0 & 0 & -1 & -1 & -2 \\ 0 & 0 & 0 & 2 & 1 & 5 \end{bmatrix}.$$

Passo 3. A partir de agora, vamos essencialmente repetir o processo já realizado até agora. Notamos que a próxima coluna que contém elementos não nulos (sem olhar para a primeira linha!) é a quarta coluna. Portanto, esta é uma coluna pivô e vamos posicionar um elemento não nulo na segunda linha. Por exemplo, podemos trocar a segunda e a terceira linhas.

$$\begin{bmatrix} 0 & 0 & 2 & 1 & -1 & 8 \\ 0 & 0 & 0 & 1/2 & 1/2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 & -2 \\ 0 & 0 & 0 & 2 & 1 & 5 \end{bmatrix}.$$

Antes de prosseguir, podemos ainda simplificar alguns cálculos multiplicando linhas por escalares (fazer isso é realizar um operação elementar!):

$$\begin{bmatrix} 0 & 0 & 2 & 1 & -1 & 8 \\ 0 & 0 & 0 & 1/2 & 1/2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 & -2 \\ 0 & 0 & 0 & 2 & 1 & 5 \end{bmatrix} \xrightarrow{2\cdot\ell_3 \text{ e } (-1)\cdot\ell_4} \begin{bmatrix} 0 & 0 & 2 & 1 & -1 & 8 \\ 0 & 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 2 & 1 & 5 \end{bmatrix}$$

Passo 4. Prosseguimos como no Passo 2 para eliminar todos os elementos que estão abaixo da posição de pivô da quarta coluna:

$$\begin{bmatrix} 0 & 0 & 2 & 1 & -1 & 8 \\ 0 & 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 2 & 1 & 5 \end{bmatrix} \xrightarrow{-\ell_2+\ell_4 \text{ em } \ell_4} \begin{bmatrix} 0 & 0 & 2 & 1 & -1 & 8 \\ 0 & 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 5 \end{bmatrix}$$

$$\xrightarrow{-2\ell_2+\ell_5 \text{ em } \ell_5} \begin{bmatrix} 0 & 0 & 2 & 1 & -1 & 8 \\ 0 & 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix}$$

Passo 5. Finalmente, identificamos a coluna 5 como coluna pivô e obtemos uma matriz em forma escalonada ao deixar um elemento não nulo na posição de pivô:

$$\begin{bmatrix} 0 & 0 & 2 & 1 & -1 & 8 \\ 0 & 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \lhd$$

Agora, para obter a forma escalonada reduzida *a partir da forma escalonada*, primeiramente aplicamos os passos acima para obter uma matriz em forma escalonada. Em seguida, aplicamos outros passos iterativos:

1. Começar pela posição de pivô mais à *direita* e eliminar os elementos não nulos *acima* da posição de pivô;
2. Se necessário, dividir a linha pelo valor do elemento líder (que está na posição de pivô) para que o elemento líder fique igual a 1;
3. Repetir os primeiros passos para o elemento líder na próxima (da direita para a esquerda) coluna pivô.

Observamos que poderíamos ter primeiro realizado o Passo 2 e depois o Passo 1 se julgássemos que seria mais simples para efetuar os cálculos.

Exemplo 7. Voltamos ao Exemplo 6. Identificando novamente as posições de pivô:

$$\begin{bmatrix} 0 & 0 & \mathbf{2} & 1 & -1 & 8 \\ 0 & 0 & 0 & \mathbf{1} & 1 & 2 \\ 0 & 0 & 0 & 0 & \mathbf{-1} & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

O mais da direita é o -1 da quinta coluna. Eliminando os termos não nulos acima, obtemos:

$$\begin{bmatrix} 0 & 0 & 2 & 1 & -1 & 8 \\ 0 & 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{\ell_3+\ell_2 \text{ em } \ell_2} \begin{bmatrix} 0 & 0 & 2 & 1 & -1 & 8 \\ 0 & 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{-\ell_3+\ell_1 \text{ em } \ell_1} \begin{bmatrix} 0 & 0 & 2 & 1 & 0 & 7 \\ 0 & 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

A próxima posição de pivô mais à direita é o 1 na linha 2, coluna 4. Eliminando o termo não nulo acima, obtemos:

$$\xrightarrow{-\ell_2+\ell_1 \text{ em } \ell_1} \begin{bmatrix} 0 & 0 & 2 & 0 & 0 & 4 \\ 0 & 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Finalmente, dividimos a primeira linha por 2 e a terceira linha por -1 para chegar na forma escalonada reduzida da matriz inicial:

$$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

1.4 Existência e unicidade de solução

1.4.1 Sistemas que possuem apenas uma solução

Os sistemas que vimos até agora possuiam apenas uma solução. Esta propriedade, como veremos, nem sempre é válida. No entanto, é fácil de identificar quando um sistema possui solução única analisando a forma escalonada da matriz associada: **quando todas as colunas referentes às variáveis da equação possuirem posição de pivô**. Por exemplo,

$$\left[\begin{array}{cc|c} 1 & 3 & 1 \\ 0 & -7 & -4 \end{array}\right], \left[\begin{array}{ccc|c} 1 & 2 & 1 & 12 \\ 0 & -5 & 4 & -11 \\ 0 & 0 & -3 & -3 \end{array}\right], \left[\begin{array}{cccc|c} 2 & 0 & 0 & 0 & 10 \\ 0 & 1 & 0 & 0 & -12 \\ 0 & 0 & 2 & 0 & 6 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right].$$

De fato, todas as matrizes com esta propriedade tem em sua forma escalonada reduzida a matriz identidade do lado esquerdo da barra que separa as variáveis e os valores do lado direito da igualdade do sistema:

$$\left[\begin{array}{cc|c} 1 & 0 & -5/7 \\ 0 & 1 & 4/7 \end{array}\right], \left[\begin{array}{ccc|c} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 1 \end{array}\right], \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 5 \\ 0 & 1 & 0 & 0 & -12 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right].$$

Escrevendo o sistema associado as estas últimas matrizes, é fácil de ver que a solução será única.

1.4.2 Sistemas impossíveis ou inconsistentes

Nem todo sistema linear possui solução. Um exemplo simples deste fato é

$$\left\{ \begin{array}{l} x + y = 1 \\ x + y = 2 \end{array} \right. . \tag{1.4}$$

É claro que $x + y$ não pode ser ao mesmo tempo igual a 1 e igual a 2; portanto, o sistema não possui solução. O fato de o sistema não possuir solução – ser **inconsistente** – nem sempre é tão fácil de ser identificado apenas olhando para o sistema. Mas esta propriedade salta aos olhos quando analisamos a forma escalonada da matriz associada ao sistema. De fato, na matriz escalonada de um sistema inconsistente aparece uma linha da forma

$$\left[\begin{array}{ccccc|c} 0 & 0 & 0 & 0 & 0 & a \end{array}\right],$$

com uma constante $a \neq 0$. Esta linha representa no sistema uma equação do tipo $0 = a \neq 0$, que é impossível de ser satisfeita.

No sistema da fórmula (1.4) acima, temos a matriz aumentada

$$\left[\begin{array}{cc|c} 1 & 1 & 1 \\ 1 & 1 & 2 \end{array}\right] \xrightarrow{-\ell_1+\ell_2 \text{ em } \ell_2} \left[\begin{array}{cc|c} 1 & 1 & 1 \\ 0 & 0 & 1 \end{array}\right].$$

Logo, nosso sistema é equivalente ao sistema impossível

$$\left\{ \begin{array}{ll} x + y & = 1 \\ \quad 0 & = 1 \end{array} \right. .$$

1.4.3 Sistemas com mais de uma solução

Com exceção da subseção anterior, todos os sistemas que resolvemos anteriormente possuiam apenas uma solução. Vejamos um exemplo onde este não é o caso.

Exemplo 8. Resolver

$$\begin{cases} & 3x_2 & -6x_3 & +6x_4 & +4x_5 & = & -5 \\ 3x_1 & -7x_2 & +8x_3 & -5x_4 & +8x_5 & = & 9 \\ 3x_1 & -9x_2 & +12x_3 & -9x_4 & +6x_5 & = & 15 \end{cases}.$$

A matriz aumentada associada a este sistema é

$$\left[\begin{array}{ccccc|c} 0 & 3 & -6 & 6 & 4 & -5 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 3 & -9 & 12 & -9 & 6 & 15 \end{array}\right].$$

Transformando esta matriz em sua forma escalonada reduzida (faça todos os cálculos!), obtemos

$$\left[\begin{array}{ccccc|c} 1 & 0 & -2 & 3 & 0 & -24 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{array}\right],$$

que é equivalente ao sistema linear

$$\begin{cases} x_1 - 2x_3 + 3x_4 & = & -24 \\ x_2 - 2x_3 + 2x_4 & = & -7 \\ x_5 & = & 4 \end{cases}$$

Neste sistema x_5 possui um valor fixo igual a 4, mas as outras variáveis tem dois graus de liberdade. De fato, atribuindo valores quaisquer para x_3 e x_4, obtemos os valores de x_1 e x_2 pelas duas primeiras equações acima e temos uma solução do sistema. Diferentes escolhas dos parâmetros x_3 e x_4 geram soluções diferentes.

Uma forma sistemática de analisar a situação é a seguinte: As variáveis correspondentes a colunas que possuem posição de pivô são chamadas de **variáveis dependentes** e as demais são chamadas de **variáveis livres**. Assim, podemos escrever

$$\begin{cases} x_1 = 2x_3 - 3x_4 - 24 \\ x_2 = 2x_3 - 2x_4 - 7 \\ x_3, x_4 \text{ livres} \\ x_5 = 4 \end{cases}$$

e o sistema possui infinitas soluções.$\triangleleft$

1.4.4 Conclusões

Ao analisar a matriz aumentada associada a um sistema linear, concluiremos *inevitavelmente* que uma das seguintes situações é válida:

- Todas as colunas referentes às variáveis do sistema possuem posição de pivô, *e.g*,

$$\left[\begin{array}{ccc|c} 1 & 3 & -6 & 6 \\ 0 & -8 & 11 & \pi \\ 0 & 0 & 12 & -9 \end{array}\right].$$

  Neste caso, vimos que o sistema possui **apenas uma solução**. Nota que *esta possibilidade apenas é possível se tivermos um sistema com o número de equações igual ao número de variáveis.*

- Caso a coluna referentes à alguma variável do sistema não possua posição de pivô, teremos duas possibilidades

(i) Alguma coluna não é coluna pivô mas não existem linhas inconsistentes, *e.g*,

$$\left[\begin{array}{ccccc|c} 1 & 3 & 2 & -6 & 2 & 6 \\ 0 & -8 & 4 & 11 & -1 & -3 \\ 0 & 0 & 0 & 12 & 1 & -9 \end{array}\right].$$

Neste caso, as variáveis referentes a colunas que não são pivô (na matriz acima, colunas 3 e 5) podem ser consideradas variáveis livres e o sistema **possui infinitas soluções**.

(ii) Alguma coluna não é coluna pivô mas existem linhas que são inconsistentes, *e.g*,

$$\left[\begin{array}{ccccc|c} 1 & 3 & 2 & -6 & 2 & 6 \\ 0 & -8 & 4 & 11 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & -9 \end{array}\right].$$

Neste caso, **não existem soluções** para o sistema, pois uma das equações é impossível de ser satisfeita.

Concluimos assim que *um sistema linear ou não possui solução, ou tem apenas uma solução ou então possui infinitas soluções*. São estes três todos os casos possíveis e podemos decidir em qual estamos apenas olhando para a forma escalonada da matriz aumentada associada ao sistema.

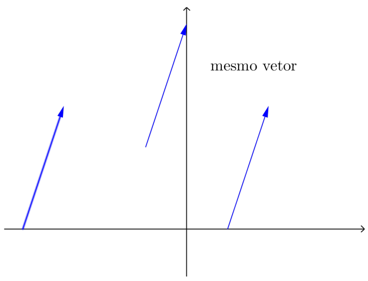
2

Semana 2

Existem maneiras diferentes de representar e de interpretar elementos de Álgebra Linear. Nesta parte, vemos algumas das principais.

2.1 Vetores

Um **vetor** no espaço tridimensional é um objeto com magnitude, direção e sentido bem definidos (não importa o ponto inicial).

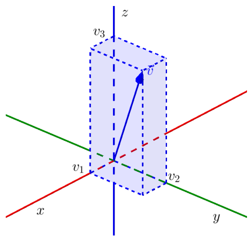


Fixando o sistema de coordenadas Cartesiano usual e usando a origem como referência, podemos introduzir coordenadas:

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = v_1\vec{e_1} + v_2\vec{e_2} + v_3\vec{e_3},$$

onde

$$\vec{e_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{e_2} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \text{e} \quad \vec{e_3} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$



No livro "Álgebra Linear e suas Aplicações", de David Lay, vetores são representados por letras em negrito. Nestas notas, utilizamos uma setinha em cima da letra para indicar um vetor.

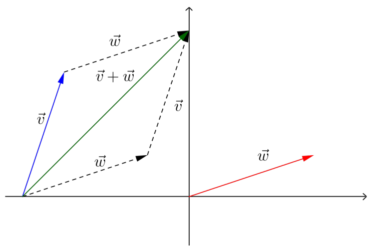
O conjunto de todos os vetores com três componentes como acima é chamado de $\mathbb{R}^3$ (leia-se "erre três"). Em coordenadas, se tivermos

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \quad \text{e} \quad \vec{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix},$$

podemos definir a **soma de vetores** componente a componente:

$$\vec{v} + \vec{w} := \begin{bmatrix} v_1 + w_1 \\ v_2 + w_2 \\ v_3 + w_3 \end{bmatrix}.$$

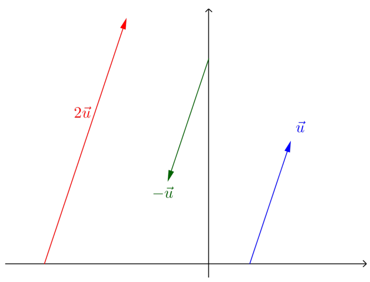
Geometricamente, isto pode ser interpretado como segue:



A **multiplicação** do vetor $\vec{v}$ **por escalar** $k \in \mathbb{R}$ é definida como:

$$k\vec{v} := \begin{bmatrix} kv_1 \\ kv_2 \\ kv_3 \end{bmatrix}.$$

Geometricamente,



Estas considerações que fizemos também são válidas para outro número de componentes, com a possível perda de visualização geométrica, no caso de quatro ou mais componentes.

Mais geralmente, um vetor $\vec{v}$ do conjunto $\mathbb{R}^n$ pode ser pensado como um objeto com n componentes:

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_{n-1} \\ v_n \end{bmatrix} = v_1\vec{e}_1 + v_2\vec{e}_2 + \cdots + v_{n-1}\vec{e}_{n-1} + v_n\vec{e}_n,$$

onde

$$\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \quad \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \cdots, \vec{e}_{n-1} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \end{bmatrix}, \quad \vec{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

A soma de vetores é dada em coordenadas por:

$$\vec{v} + \vec{w} := \begin{bmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \vdots \\ v_{n-1} + w_{n-1} \\ v_n + w_n \end{bmatrix}.$$

A multiplicação por escalar por:

$$k\vec{v} := \begin{bmatrix} kv_1 \\ kv_2 \\ \vdots \\ kv_{n-1} \\ kv_n \end{bmatrix}.$$

2.1.1 Representação matricial para sistemas Lineares

Na linguagem da seção anterior, nós dizemos que dois vetores são iguais quando todas as suas componentes são iguais. Desta forma, podemos interpretar as equações de um sistema linear

$$\begin{cases} x + 3y &= 1 \\ 2x - y &= -2 \end{cases}$$

como uma igualdade entre vetores de $\mathbb{R}^2$, isto é, de vetores com duas componentes:

$$\begin{bmatrix} x + 3y \\ 2x - y \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}.$$

Além disso, desta forma surge o **produto de uma matriz por um vetor**:

$$\begin{bmatrix} 1 & 3 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix},$$

definido de modo que as duas últimas equações acima signifiquem a mesma coisa. Esta última notação é conhecida como a **forma matricial** do sistema linear.

Exemplo 9. O sistema, que já apareceu nas notas da semana anterior

$$\begin{cases} x_1 + 2x_2 + x_3 = 12 \\ x_1 - 3x_2 + 5x_3 = 1 \\ 2x_1 - x_2 + 3x_3 = 10 \end{cases}$$

pode ser representado matricialmente por

$$\begin{bmatrix} 1 & 2 & 1 \\ 1 & -3 & 5 \\ 2 & -1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 12 \\ 1 \\ 10 \end{bmatrix}.$$

De forma mais sucinta,

$$\boxed{A\vec{x} = \vec{b}}$$

onde

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & -3 & 5 \\ 2 & -1 & 3 \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \text{e} \quad \vec{b} = \begin{bmatrix} 12 \\ 1 \\ 10 \end{bmatrix}.$$

Mais geralmente, uma matriz do tipo $m \times n$ (leia-se "m por n") é uma matriz com m linhas e n colunas:

$$A = (a_{ij}) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

e pode estar associada a um sistema com m equações e n variáveis, já que o produto

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix},$$

quando visto componente a componente, representa:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

2.2 Combinações lineares de vetores

Uma **combinação linear** dos vetores $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k$ é um vetor da forma:

$$\vec{v} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k = \sum_{i=1}^{k} x_i\vec{v}_i,$$

onde $x_1, x_2, \ldots, x_k$ são números reais. Em outras palavras, uma combinação linear é uma soma de múltiplos dos vetores $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k$.

Exemplo 10. Para os vetores

$$\vec{v}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \text{e} \quad \vec{v}_2 = \begin{bmatrix} 1 \\ 3 \end{bmatrix},$$

algumas combinações lineares são:

- $\vec{v}_1 + \vec{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$;
- $4\vec{v}_1 = 4\vec{v}_1 + 0\vec{v}_2 = 4\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 4 \\ -4 \end{bmatrix}$;
- $\vec{v}_2 = 0\vec{v}_1 + 1\vec{v}_2 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$;
- $0\vec{v}_1 + 0\vec{v}_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \vec{0}$.

Geometricamente, qualquer vetor do plano pode ser representado como combinação linear de vetores que não são colineares.

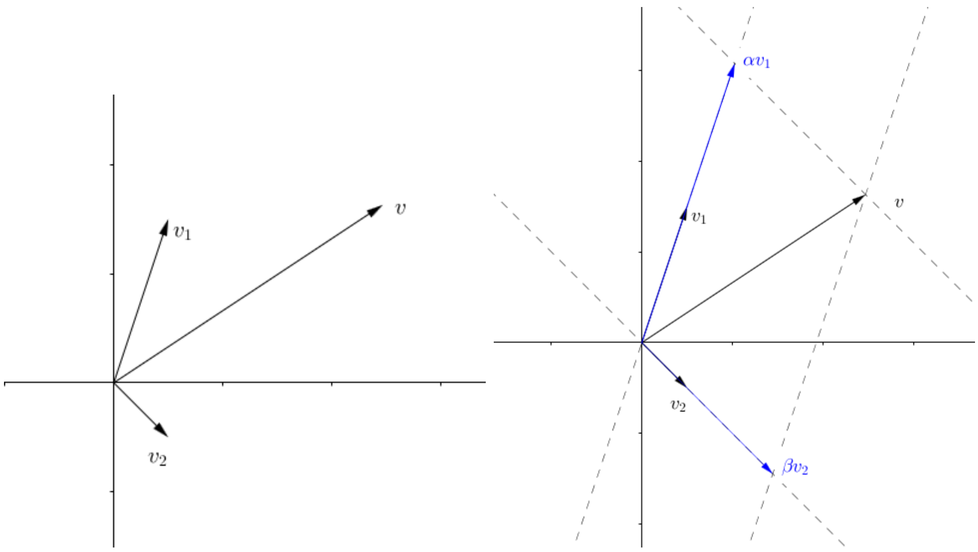
Vamos ilustrar este fato na figura abaixo. Por exemplo, para representar o vetor $\vec{v}$ como combinação linear de $\vec{v}_1$ e $\vec{v}_2$, podemos traçar retas paralelas aos vetores, passando pela origem e pela extremidade do vetor $\vec{v}$, como na figura da direita. Assim, pela interpretação geométrica da soma de vetores (ver início das notas), sabemos que $\vec{v}$ é a soma dos vetores em azul. Mas estes, por construção, são colineares aos vetores iniciais e logo múltiplos destes. Concluímos que

$$\vec{v} = \alpha\vec{v}_1 + \beta\vec{v}_2,$$

isto é, $\vec{v}$ é uma combinação linear de $\vec{v}_1$ e $\vec{v}_2$. $\triangleleft$

De forma mais geral, nós dizemos que um vetor $\vec{v} \in \mathbb{R}^m$ é combinação linear dos k vetores $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k \in \mathbb{R}^m$ quando conseguirmos encontrar números reais $x_1, x_2, \ldots, x_k$ tais que

$$x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k = \vec{v}.$$



Nós vamos nos referir a este tipo de equação por **equação vetorial**. Para decidir se um vetor é combinação linear de outros, devemos decidir se existem estes números reais $x_1, x_2, \ldots, x_k$. Escrevendo em coordenadas:

$$\vec{v}_1 = \begin{bmatrix} v_{11} \\ v_{21} \\ v_{31} \\ \vdots \\ v_{m1} \end{bmatrix}, \ \vec{v}_2 = \begin{bmatrix} v_{12} \\ v_{22} \\ v_{32} \\ \vdots \\ v_{m2} \end{bmatrix}, \ \vec{v}_3 = \begin{bmatrix} v_{13} \\ v_{23} \\ v_{33} \\ \vdots \\ v_{m3} \end{bmatrix}, \ \cdots, \ \vec{v}_k = \begin{bmatrix} v_{1k} \\ v_{2k} \\ v_{3k} \\ \vdots \\ v_{mk} \end{bmatrix}, \ \vec{v} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_m \end{bmatrix},$$

vemos que encontrar os coeficientes da combinação linear, caso estes existam, equivale a resolver a equação vetorial

$$x_1 \begin{bmatrix} v_{11} \\ v_{21} \\ v_{31} \\ \vdots \\ v_{m1} \end{bmatrix} + x_2 \begin{bmatrix} v_{12} \\ v_{22} \\ v_{32} \\ \vdots \\ v_{m2} \end{bmatrix} + x_3 \begin{bmatrix} v_{13} \\ v_{23} \\ v_{33} \\ \vdots \\ v_{m3} \end{bmatrix} + \cdots + x_k \begin{bmatrix} v_{1k} \\ v_{2k} \\ v_{3k} \\ \vdots \\ v_{mk} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_m \end{bmatrix},$$

que, por sua vez, é equivalente a

$$\begin{bmatrix} v_{11}x_1 + v_{12}x_2 + v_{13}x_3 + \cdots + v_{1k}x_k \\ v_{21}x_1 + v_{22}x_2 + v_{23}x_3 + \cdots + v_{2k}x_k \\ v_{31}x_1 + v_{32}x_2 + v_{33}x_3 + \cdots + v_{3k}x_k \\ \vdots \\ v_{m1}x_1 + v_{m2}x_2 + v_{m3}x_3 + \cdots + v_{mk}x_k \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_m \end{bmatrix}.$$

Mas isto é resolver um sistema linear! Portanto novo conceito está intimamente relacionado com tudo o que já tínhamos visto antes.

O **espaço gerado** por todas as combinações lineares dos vetores $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k$ é denotado por $\text{Span}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$.

- Vimos que o conjunto gerado pelos dois vetores do exemplo anterior é o plano inteiro, isto é, $\text{Span}\{\vec{v}_1, \vec{v}_2\} = \mathbb{R}^2$.
- Nota que o vetor nulo sempre pertence ao espaço gerado, assim como todos os vetores $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k$.
- O conjunto gerado por apenas um vetor não nulo representa uma reta.

2.3 Sistemas lineares e combinações lineares das colunas

A associação feita acima pode ser resumida como:

> Resolver o sistema linear $A\vec{x} = \vec{b}$ equivale a decidir se o vetor $\vec{b}$ é uma combinação linear das colunas de A.

ou ainda

> Resolver o sistema linear $A\vec{x} = \vec{b}$ equivale a decidir se o vetor $\vec{b}$ pertence ao espaço gerado pelas colunas de A.

Desta forma, tudo o que já vimos sobre existência de soluções para sistemas lineares pode ser "traduzido" para o contexto de combinações lineares e espaços gerados (faça as traduções!).

3

Semana 3

3.1 Dependência e independência linear

Nós vimos, por definição, que um vetor $\vec{v} \in \mathbb{R}^m$ é combinação linear dos k vetores $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k \in \mathbb{R}^m$ quando conseguirmos encontrar números reais $x_1, x_2, \dots, x_k$ tais que

$$x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k = \vec{v}.$$

Vimos também que para decidir se um vetor é combinação linear de outros, devemos verificar se existem estes números reais $x_1, x_2, \dots, x_k$. Em coordenadas, isto é equivalente a resolver o sistema linear:

$$\begin{bmatrix} v_{11} & v_{12} & v_{13} & \cdots & v_{1k} \\ v_{21} & v_{22} & v_{23} & \cdots & v_{2k} \\ v_{31} & v_{32} & v_{33} & \cdots & v_{3k} \\ \vdots & \vdots & \vdots & & \vdots \\ v_{m1} & v_{m2} & v_{m3} & \cdots & v_{mk} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_m \end{bmatrix}.$$

Agora, dizemos que os vetores $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k \in \mathbb{R}^m$ são **linearmente independentes (LI)** se nenhum dos vetores puder ser escrito combinação linear dos demais. Uma forma de escrever isto matematicamente é a seguinte:

$$\boxed{\text{Se } x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k = \vec{0}, \text{ então } x_1 = x_2 = \cdots = x_k = 0.}$$

De fato, imaginem que pelo menos um dos coeficientes acima fosse diferente de zero, digamos $x_2 \neq 0$. Daí podemos dividir por x_2 e conseguiríamos escrever o vetor $\vec{v}_2$ como combinação linear dos demais:

$$\vec{v}_2 = -\frac{x_1}{x_2}\,\vec{v}_1 - \frac{x_3}{x_2}\,\vec{v}_3 - \cdots - \frac{x_k}{x_2}\,\vec{v}_k.$$

Se os vetores $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k \in \mathbb{R}^m$ não forem linearmente independentes, então nós dizemos que eles são **linearmente dependentes**.

Exemplo 11. Vamos verificar se os vetores

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

são LI ou LD.

Para verificar se são LI, nós devemos considerar a equação vetorial

$$x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \vec{0} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Como a equação é homogênea, temos pelo menos a solução trivial: $x_1 = 0$, $x_2 = 0$ e $x_3 = 0$. Se esta for a única solução, então os vetores são LI. Se existir alguma outra solução que não seja a trivial, então os vetores são LD.

Resolver a equação vetorial equivale a resolver o sistema linear

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

É fácil de ver que a única solução deste sistema é a trivial e, portanto, os vetores são LI. Se este sistema não fosse fácil de resolver, devemos começar a resolvê-lo por escalonamento. Se existissem variáveis livres, então existiriam infinitas soluções. $\lhd$

Exemplo 12. Analisamos agora os vetores

$$\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}.$$

Como no exemplo anterior, para verificar se são LI, nós devemos considerar a equação vetorial

$$x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Como a equação é homogênea, temos pelo menos a solução trivial: $x_1 = 0$, $x_2 = 0$ e $x_3 = 0$. Se esta for a única solução, então os vetores são LI. Se existir alguma outra solução que não seja a trivial, então os vetores são LD.

Resolver a equação vetorial equivale a resolver o sistema linear

$$\begin{bmatrix} 1 & 1 & 1 & -1 \\ 0 & 1 & 1 & 2 \\ 1 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Em um sistema com mais variáveis do que equações, podemos não ter soluções (no caso de alguma linha inconsistente) ou podemos ter infinitas soluções (no caso de haver variáveis livres). Não é possível ter unicidade. Como já sabemos que em um sistema homogêneo, sempre temos pelo menos a solução trivial $x_1 = 0$, $x_2 = 0$, $x_3 = 0$ e $x_4 = 0$, podemos concluir que existem infinitas soluções. Logo, o conjunto de vetores é LD.

O argumento que utilizamos no Exemplo 12 acima é geral e se aplica em várias situações. Temos

$$\boxed{\text{Se } k > m, \text{ então os vetores } \vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k \in \mathbb{R}^m \text{ são necessariamente LD.}}$$

Desta forma, não existem quatro vetores LI em $\mathbb{R}^3$. Não conseguiríamos encontrar sete vetores LI em $\mathbb{R}^5$ e assim por diante.

Façamos agora algumas observações gerais sobre independência linear:

- O vetor nulo, mesmo que sozinho, é LD. Se $\vec{v} = \vec{0}$ estiver em um conjunto de vetores, este conjunto será automaticamente LD.

- Ao considerarmos apenas dois vetores $\vec{u}$ e $\vec{v}$, dizer que estes vetores são LD significa que eles são múltiplos um dou outro e, portanto, colineares.

- Veremos mais adiante no curso, que a dimensão do espaço gerado por vetores linearmente independentes é exatamente o número de vetores. Desta maneira, concluiremos, por exemplo, que o conjunto gerado por dois vetores do espaço tridimensional é um espaço linear de dimensão dois: um plano.

3.2 Independência linear e sistemas lineares

Na última semana, vimos que resolver o sistema linear $A\vec{x} = \vec{b}$ equivale a decidir se o vetor $\vec{b}$ é uma combinação linear das colunas de A.

A situação da seção anterior é a de decidir se os vetores $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k$ são LI ou LD. Isto, consequentemente, equivale a decidir se existe solução não trivial do sistema linear homogêneo

$$A\vec{x} = \vec{0},$$

onde a matriz A é formada com os vetores $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k$ como colunas.

Exemplo 13. Naturalmente, podemos traduzir os exemplos da seção anterior desta forma. Ou seja no Exemplo 11, temos que a matriz

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

tem as colunas linearmente independentes (LI). Por outro lado, as colunas da matriz

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

do Exemplo 12 são linearmente dependentes (LD). $\triangleleft$

Podemos pensar ainda de outra forma. Considere uma matriz A de ordem $m \times n$. Para decidir se as colunas de A são LI, devemos procurar por soluções não triviais do sistema linear cuja matriz associada é A. Ao resolver um **sistema homogêneo** por escalonamento, sabemos que

- Se existirem mais colunas do que linhas (*i.e.* $m < n$), então as colunas de A são necessariamente LD.

- Se existirem mais linhas do que colunas (*i.e.* $m > n$), procedemos por escalonamento. Se todas as colunas da matriz A possuirem posição de pivô, então as colunas são LI (pois daí a única solução do sistema homogêneo é a trivial). No caso de alguma coluna não possuir posição de pivô, o sistema homogêneo possui pelo menos uma variável livre; logo, as colunas de A são LD.

Exemplo 14. Decidir se as colunas de A são LI:

$$A = \begin{bmatrix} 2 & 1 & -1 & 8 \\ -3 & -1 & 2 & -11 \\ 6 & 2 & -4 & 22 \\ -2 & 1 & 2 & -3 \end{bmatrix}.$$

Como as colunas de A são quatro vetores de $\mathbb{R}^4$, pode ser que sejam LI (caso tivesse uma coluna a mais, saberíamos automaticamente serem LD). Aplicando o algoritmo de escalonamento, sabemos que

$$A \sim \begin{bmatrix} 2 & 1 & -1 & 8 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Notando que a última coluna não possui posição de pivô, concluimos que as colunas de A são LD. Isto pois, neste caso, o sistema linear homogêneo associado, $A\vec{x} = \vec{0}$, cuja matriz aumentada associada é

$$\left[\begin{array}{cccc|c} 2 & 1 & -1 & 8 & 0 \\ 0 & 1 & 1 & 2 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right],$$

que possui uma variável livre (e portanto outras soluções que não a trivial). $\triangleleft$

3.3 Transformações lineares

Uma **transformação linear** é um tipo especial de função que associa um vetor a outro vetor

$$T : \{\text{vetores}\} \to \{\text{vetores}\}$$

e com a propriedade adicional

$$T\big(a\vec{u} + b\vec{v}\big) = aT(\vec{u}) + bT(\vec{v}), \quad \text{para todos } a, b \in \mathbb{R} \quad \text{e todos } \vec{u}, \vec{v}.$$

Em geral, pensamos em uma transformação linear como "transformando um vetor em outro".

Exemplo 15. A aplicação $T : \mathbb{R}^3 \to \mathbb{R}^3$ dada por $T(\vec{u}) = 3\vec{u}$ é a transformação linear que transforma um vetor de $\mathbb{R}^3$ no vetor que tem o triplo do comprimento.

Exemplo 16. Também é linear a transformação que faz uma rotação de um ângulo $\pi/2$. Voltaremos a este exmeplo com mais detalhes na seção seguinte.

Exemplo 17. Consideramos a transformação $T : \mathbb{R}^2 \to \mathbb{R}^3$, que transforma vetores de $\mathbb{R}^2$ em vetores de $\mathbb{R}^3$, dada pela fórmula

$$T(x_1, x_2) = (x_1 - 3x_2, 3x_1 + 5x_2, -x_1 + x_2).$$

Nota que, excepcionalmente, estamos representando os vetores como linhas. É comum quando tratamos de transformações lineares. Esta fórmula nos diz, por exemplo, que os vetores

$$\vec{u} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \text{e} \vec{v} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

são transformados nos vetores

$$T(\vec{u}) = \begin{bmatrix} 1 - 3 \cdot (-1) \\ 3 \cdot 1 + 5 \cdot (-1) \\ -1 + (-1) \end{bmatrix} = \begin{bmatrix} 4 \\ -2 \\ -2 \end{bmatrix} \quad \text{e} \quad T(\vec{v}) = \begin{bmatrix} 0 - 3 \cdot 3 \\ 3 \cdot 0 + 5 \cdot 3 \\ -0 + 3 \end{bmatrix} = \begin{bmatrix} -9 \\ 15 \\ 3 \end{bmatrix}.$$

3.4 Matriz de uma transformação linear

Voltemos à transformação linear

$$T(x_1, x_2) = (x_1 - 3x_2, 3x_1 + 5x_2, -x_1 + x_2)$$

do Exemplo 17. Lembramos que um vetor qualquer de $\mathbb{R}^2$ pode ser escrito como combinação linear dos vetores

$$\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{e} \quad \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

De fato,

$$\vec{u} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = x_1\vec{e}_1 + x_2\vec{e}_2.$$

Logo, utilizando a propriedade de a transformação ser linear, temos que

$$T(\vec{u}) = T\big(x_1\vec{e}_1 + x_2\vec{e}_2\big) = x_1 T(\vec{e}_1) + x_2 T(\vec{e}_2).$$

Calculamos $T(\vec{e}_1)$ e $T(\vec{e}_2)$ pela fórmula dada da transformação linear:

$$T(\vec{e}_1) = \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix} \quad \text{e} \quad T(\vec{e}_2) = \begin{bmatrix} -3 \\ 5 \\ 1 \end{bmatrix}.$$

Concluimos que

$$T(\vec{u}) = x_1 \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix} + x_2 \begin{bmatrix} -3 \\ 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Desta forma, associamos uma matriz de ordem 3×2 à transformação linear $T : \mathbb{R}^2 \to \mathbb{R}^3$.

O procedimento que aplicamos acima não é particular do exemplo que analisamos, de modo que é sempre possível associar a uma transformação linear $T : \mathbb{R}^n \to \mathbb{R}^m$ uma matriz A de ordem $m \times n$, chamada **matriz canônica associada à transformação linear** T ou apenas **matriz associada** a T, cujas colunas são os vetores $T(\vec{e}_1), T(\vec{e}_2), T(\vec{e}_3), \ldots, T(\vec{e}_n) \in \mathbb{R}^m$ (e portanto n colunas com m componentes cada).

Exemplo 18. A transformação linear $T : \mathbb{R}^3 \to \mathbb{R}^3$, $T(\vec{u}) = 5\vec{u}$, satisfaz

$$T(\vec{e}_1) = \begin{bmatrix} 5 \\ 0 \\ 0 \end{bmatrix}, \quad T(\vec{e}_2) = \begin{bmatrix} 0 \\ 5 \\ 0 \end{bmatrix} \quad \text{e} \quad T(\vec{e}_3) = \begin{bmatrix} 0 \\ 0 \\ 5 \end{bmatrix}.$$

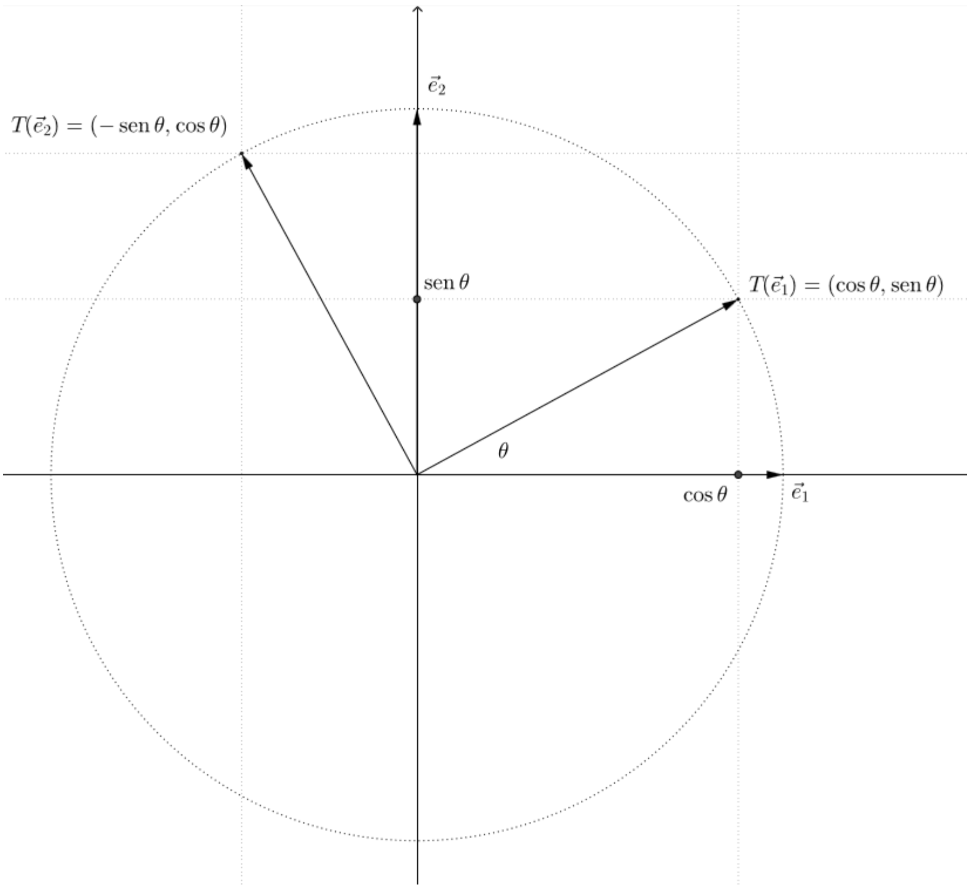
Assim, podemos escrever

$$T(\vec{x}) = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Exemplo 19. O método acima pode também ser aplicado para conseguirmos uma fórmula para transformações lineares como as do Exemplo 16. Vamos considerar a aplicação $T : \mathbb{R}^2 \to \mathbb{R}^2$ definida por

$$T(\vec{x}) = \text{rotação no sentido anti-horário de } \vec{x} \text{ por um ângulo } \theta \in (0, 2\pi).$$

A matriz da transformação linear tem por colunas a imagem por T dos vetores $\vec{e}_1$ e $\vec{e}_2$. Observamos que (ver figura)



$$T(\vec{e}_1) = \begin{bmatrix} \cos\theta \\ \operatorname{sen}\theta \end{bmatrix}, \quad T(\vec{e}_2) = \begin{bmatrix} -\operatorname{sen}\theta \\ \cos\theta \end{bmatrix}.$$

Logo, concluimos que

$$T(\vec{x}) = \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

3.5 Transformações injetoras, sobrejetoras e invertíveis

Como de costume, dada uma função $f : A \to B$, diz-se que A é o domínio de f enquanto B é o contradomínio. A imagem de f é um subconjunto de B: todos os elementos $y \in B$ tais que $f(x) = y$ ou, intuitivamente, "todos os elementos de B que são atingidos pela função f".

Na hora de decidir se uma função é invertível ou não, duas propriedades são essenciais:

1) cada elemento de B ser a imagem de no máximo um elemento de A, caso em que f é dita **injetora**;

2) a imagem de f ser igual ao contradomínio, caso em que f diz-se **sobrejetora**.

Quando estas duas propriedades são satisfeitas, isto é, quando a função f é injetora e sobrejetora, vale que f é **invertível**: podemos encontrar uma função $f^{-1} : B \to A$ que satisfaz

$$f^{-1}\big(f(x)\big) = x, \text{ para todo } x \in A \quad \text{e} \quad f\big(f^{-1}(y)\big) = y, \text{ para todo } y \in B.$$

Ao tentarmos encontrar uma **função inversa** f^{-1}, as propriedades de ser injetora e sobrejetora, aparecem naturalmente.

Vamos analisar quando que transformações lineares são injetoras, sobrejetoras e/ou invertíveis.

3.5.1 Transformações lineares injetoras

A transformação linear $T : \mathbb{R}^n \to \mathbb{R}^m$ é injetora quando, para $\vec{b} \in \mathbb{R}^m$, a equação

$$T(\vec{x}) = \vec{b}$$

possuir uma única solução ou nenhuma (no máximo uma – comparar com a definição acima). Como vimos, existe uma matriz A de ordem $m \times n$ associada à transformação linear T, de modo que temos que analisar as soluções de

$$A\vec{x} = \vec{b}.$$

Recaimos novamente em um sistema linear!

No caso particular em que $\vec{b} = \vec{0}$, o sistema homogêneo $A\vec{x} = \vec{0}$ sempre possui a solução trivial $\vec{x} = \vec{0}$. Neste caso, para que a transformação linear seja injetora devemos verificar que esta é a única solução de $A\vec{x} = \vec{0}$. Na verdade, é possível verificar (ver livro David C. Lay):

$$\boxed{T \text{ é injetora se, e somente se, } A\vec{x} = \vec{0} \text{ possui apenas a solução trivial.}}$$

Em particular, este caso somente é possível se $n \leq m$ (reler observações que seguem o Exemplo 3).

Notem que a afirmação acima é também equivalente a

$$\boxed{T \text{ é injetora se, e somente se, as colunas de } A \text{ são linearmente independentes.}}$$

3.5.2 Transformações lineares sobrejetoras

A transformação linear $T : \mathbb{R}^n \to \mathbb{R}^m$ é sobrejetora quando, para todo $\vec{b} \in \mathbb{R}^n$, a equação

$$T(\vec{x}) = \vec{b}$$

possui alguma solução (comparar com a definição anterior). Seja A a matriz de ordem $m \times n$ associada a T. Assim, para verificar que T é sobrejetora, devemos verificar que, para qualquer $\vec{b} \in \mathbb{R}^m$, o sistema linear

$$A\vec{x} = \vec{b}$$

possui ao menos uma solução (uma ou infinitas). Isto é equivalente a:

$$\boxed{T \text{ é sobrejetora se, e somente se, as colunas de } A \text{ geram todo o espaço } \mathbb{R}^m.}$$

Em particular, este caso somente é possível se $n \geq m$ (veremos mais adiante no curso – a imagem da transformação $T : \mathbb{R}^n \to \mathbb{R}^m$ será de dimensão no máximo igual a n).

Exemplo 20. Considere a transformação linear T cuja matriz associada é

$$A = \begin{bmatrix} 5 & 3 & 1 & 1 \\ 0 & -1 & 1 & -1 \\ 0 & 0 & 0 & 3 \end{bmatrix}.$$

Como são quatro colunas de $\mathbb{R}^3$, vemos que as colunas são LD e, portanto, T não é injetora.

Por outro lado, a matriz já está em forma escalonada. Vamos analisar se o sistema

$$A\vec{x} = \vec{b} \iff \left[\begin{array}{cccc|c} 5 & 3 & 1 & 1 & b_1 \\ 0 & -1 & 1 & -1 & b_2 \\ 0 & 0 & 0 & 3 & b_3 \end{array}\right]$$

possui solução para todo $\vec{b} \in \mathbb{R}^3$. De fato, o sistema possui solução (já que nenhuma linha da sua forma escalonada é inconsistente). Em verdade, o sistema possui uma variável livre. Logo, T é sobrejetora.

Exemplo 21. Considere a transformação linear T cuja matriz associada é

$$A = \begin{bmatrix} 3 & 1 \\ 5 & 7 \\ 0 & -4 \end{bmatrix}.$$

Como são somente duas colunas, é fácil ver que uma não é múltipla da outra: por causa das primeiras componentes, podemos pensar que a primeira coluna é três vezes a primeira, mas verificando a segunda linha já não dá certo $3 \cdot 7 \neq 5$. Logo, as colunas são LI e a transformação T é injetora.

Por outro lado, as colunas de A geram um espaço de dimensão no máximo 2; logo, não geram $\mathbb{R}^3$. Portanto, T não é sobrejetora.

3.5.3 Transformações lineares invertíveis

Pelas observações das últimas subseções, só é possível da transformação linear $T : \mathbb{R}^n \to \mathbb{R}^m$ ser invertível quando $m = n$, pois T deverá ser ao mesmo tempo injetora e sobrejetora. Veremos na semana seguinte um método para calcular a transformação inversa T^{-1}, que será também uma transformação linear.

4

Semana 4

Em construção ... Caso deseja colaborar com a escrita deste livro, veja como em:
https://github.com/livroscolaborativos/AlgebraLinear

5

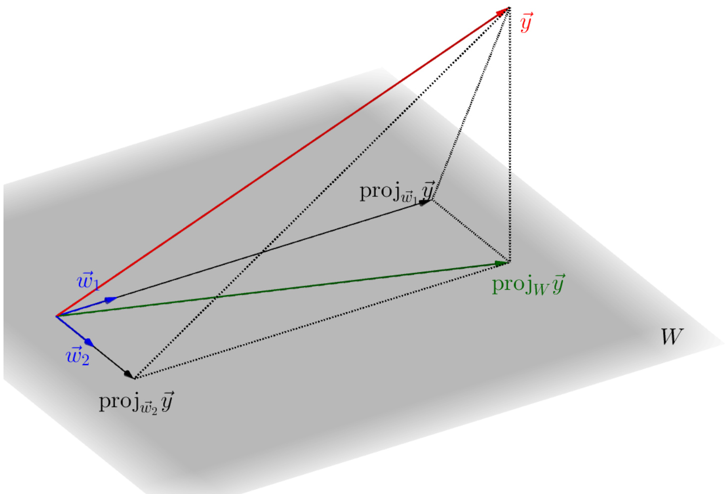
Semana 12

5.1 Projeção ortogonais sobre subespaços

Na última semana, estudamos projeções ortogonais e vimos que a projeção ortogonal de $\vec{y}$ na direção de um vetor $\vec{w}$ pode ser calculada por

$$\operatorname{proj}_{\vec{w}} \vec{y} = \frac{\vec{y} \cdot \vec{w}}{\vec{w} \cdot \vec{w}} \, \vec{w}$$

Gostaríamos de endereçar agora a seguinte questão: como obter a projeção de um vetor $\vec{y} \in \mathbb{R}^3$ sobre um plano W?



Vamos trabalhar geométricamente: suponhamos que o plano

$$W = \operatorname{Span}\{\vec{w}_1, \vec{w}_2\}$$

e que os vetores $\vec{w}_1$ e $\vec{w}_2$ são ortogonais. Em outras palavras, $\{\vec{w}_1, \vec{w}_2\}$ é uma base ortogonal de W. Como veremos abaixo, o Processo de Gram-Schmidt fornece um algoritmo para obter uma base ortogonal de um espaço vetorial a partir de uma outra base qualquer e, logo, sempre é possível considerar uma base ortogonal.

A **projeção ortogonal sobre o subespaço** W, denotada por $\operatorname{proj}_W \vec{y}$, é definida como

$$\operatorname{proj}_W \vec{y} = \operatorname{proj}_{\vec{w}_1} \vec{y} + \operatorname{proj}_{\vec{w}_2} \vec{y} = \frac{\vec{y} \cdot \vec{w}_1}{\vec{w}_1 \cdot \vec{w}_1} \, \vec{w}_1 + \frac{\vec{y} \cdot \vec{w}_2}{\vec{w}_2 \cdot \vec{w}_2} \, \vec{w}_2.$$

Notamos que este vetor pertence a W, já que é uma combinação linear dos elementos de uma base de W. Mas mais fundamental é o fato de a projeção ser feita em uma direção ortogonal ao plano: devemos verificar que

$$\vec{y} - \operatorname{proj}_W \vec{y} \ \text{ é ortogonal a } W.$$

Qualquer vetor de W pode ser escrito como

$$\vec{w} = \frac{\vec{w}\cdot\vec{w}_1}{\vec{w}_1\cdot\vec{w}_1}\,\vec{w}_1 + \frac{\vec{w}\cdot\vec{w}_2}{\vec{w}_2\cdot\vec{w}_2}\,\vec{w}_2.$$

Exercício 1. Utilize as propriedades do produto escalar e a ortogonalidade da base $\{\vec{w}_1, \vec{w}_2\}$ para verificar que

$$\vec{w}\cdot\big(\vec{y} - \operatorname{proj}_W \vec{y}\big) = 0.$$

Exemplo 22. Vamos calcular a projeção do vetor

$$\vec{y} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

sobre o plano gerado pelos vetores

$$\vec{w}_1 = \begin{bmatrix} 0 \\ 3 \\ -7 \end{bmatrix} \quad \text{e} \quad \vec{w}_2 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}.$$

Comlculamos

$$\begin{aligned}
\operatorname{proj}_W \vec{y} &= \frac{\vec{y}\cdot\vec{w}_1}{\vec{w}_1\cdot\vec{w}_1}\,\vec{w}_1 + \frac{\vec{y}\cdot\vec{w}_2}{\vec{w}_2\cdot\vec{w}_2}\,\vec{w}_2 = \frac{3-7}{9+49}\,\vec{w}_1 + \frac{1+2}{1+4}\,\vec{w}_2 \\
&= -\frac{2}{29}\begin{bmatrix} 0 \\ 3 \\ -7 \end{bmatrix} + \frac{3}{5}\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 3/5 \\ -6/29 \\ 244/145 \end{bmatrix}.
\end{aligned}$$

Desenvolvemos acima o conceito de projeção ortogonal sobre um subespaço do espaço $\mathbb{R}^3$ de dimensão três apenas pela conveniência da visualização geométrica. Na verdade, o conceito pode ser definido em qualquer dimensão.

Seja $W \subset \mathbb{R}^n$ um subespaço vetorial de dimensão k e $\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}$ uma base ortogonal de W. Dado um vetor $\vec{y} \in \mathbb{R}^n$ qualquer, definimos a **projeção ortogonal de $\vec{y}$ sobre W** como

$$\operatorname{proj}_W \vec{y} = \frac{\vec{y}\cdot\vec{w}_1}{\vec{w}_1\cdot\vec{w}_1}\,\vec{w}_1 + \frac{\vec{y}\cdot\vec{w}_2}{\vec{w}_2\cdot\vec{w}_2}\,\vec{w}_2 + \cdots + \frac{\vec{y}\cdot\vec{w}_k}{\vec{w}_k\cdot\vec{w}_k}\,\vec{w}_k$$

Pode-se verificar, como no Exercício acima, que

$$\vec{w}\cdot\big(\vec{y} - \operatorname{proj}_W \vec{y}\big) = 0, \text{ para qualquer } \vec{w} \in W,$$

isto é, a projeção $\operatorname{proj}_W \vec{y}$ pertence a W e $\vec{y} - \operatorname{proj}_W \vec{y}$ é ortogonal a todo elemento de W, assim como nossa intuição esperaria.

Exemplo 23. Calcular a projeção do vetor

$$\vec{y} = \begin{bmatrix} 1 \\ 0 \\ -1 \\ -3 \\ 2 \end{bmatrix}$$

sobre o espaço tridimensional gerado pelos vetores

$$\vec{w}_1 = \begin{bmatrix} 0 \\ 4 \\ 1 \\ 6 \\ 0 \end{bmatrix}, \quad \vec{w}_2 = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \\ -1 \end{bmatrix} \quad \text{e} \quad \vec{w}_3 = \begin{bmatrix} 0 \\ 0 \\ 2 \\ 4 \\ 1 \end{bmatrix}.$$

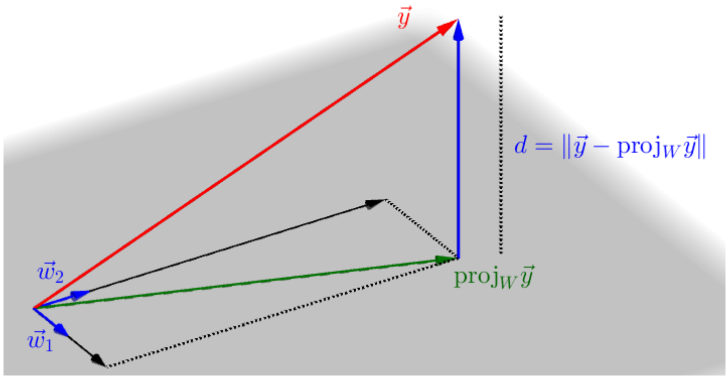
Devemos calcular

$$\begin{aligned}\operatorname{proj}_W \vec{y} &= \frac{\vec{y}\cdot\vec{w}_1}{\vec{w}_1\cdot\vec{w}_1}\,\vec{w}_1 + \frac{\vec{y}\cdot\vec{w}_2}{\vec{w}_2\cdot\vec{w}_2}\,\vec{w}_2 + \frac{\vec{y}\cdot\vec{w}_3}{\vec{w}_3\cdot\vec{w}_3}\,\vec{w}_3\\ &= \frac{-1-18}{16+1+36}\,\vec{w}_1 + \frac{1-2-2}{1+4+1}\,\vec{w}_2 + \frac{-2-12+2}{4+16+1}\,\vec{w}_3\\ &= -\frac{19}{53}\begin{bmatrix}0\\4\\1\\6\\0\end{bmatrix} - \frac{1}{2}\begin{bmatrix}1\\0\\2\\0\\-1\end{bmatrix} - \frac{12}{21}\begin{bmatrix}0\\0\\2\\4\\1\end{bmatrix} = \begin{bmatrix}-1/2\\-76/53\\-2784/1113\\-4938/1113\\-1/12\end{bmatrix}.\end{aligned}$$

Confira estas contas com uma calculadora (os coeficientes dos vetores nem sempre são bonitinhos!)

5.2 Distância a subespaços e a melhor aproximação

A projeção ortogonal pode ser utilizada para calcular distâncias entre o pontos e subespaços.



Dado um ponto $P = (x_1, x_2, \ldots, x_n)$, formamos o vetor

$$\vec{y} = \begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}$$

cujas componentes são as coordenadas de P. Como vimos na seção anterior, o vetor $\vec{y} - \operatorname{proj}_W \vec{y}$ é ortogonal ao subespaço W. Sendo a distância de P até W denotada por $\operatorname{dist}(P, W)$ e definida como a menor distância entre P e elementos de W, temos que

$$\operatorname{dist}(P, W) = \|\vec{y} - \operatorname{proj}_W \vec{y}\|.$$

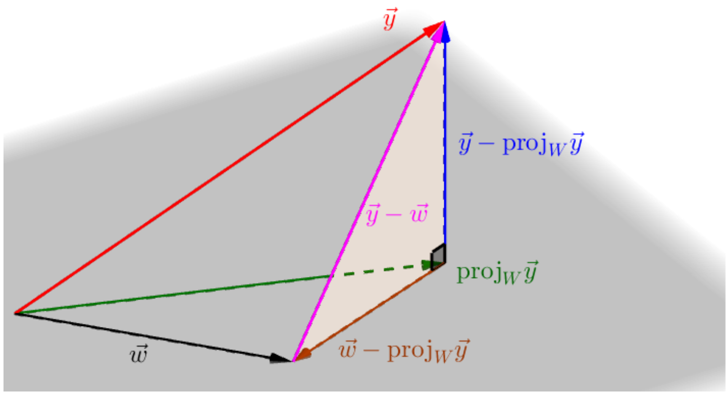
Justificativa desta afirmação. Dado qualquer outro vetor $\vec{w} \in W$, o Teorema de Pitágoras implica que (ver figura)

$$\|\vec{y} - \vec{w}\|^2 = \|\vec{w} - \operatorname{proj}_W \vec{y}\|^2 + \|\vec{y} - \operatorname{proj}_W \vec{y}\|^2 > \|\vec{y} - \operatorname{proj}_W \vec{y}\|^2.$$

Em outras palavras, $\|\vec{y} - \operatorname{proj}_W \vec{y}\|$ é de fato a menor distância entre P e elementos de W. $\square$

Exemplo 24. Calcular a distância do ponto $P = (1, 2, -1, 0, 7)$ até o plano $W = \operatorname{Span}\{\vec{u}, \vec{v}\}$ (contido em $\mathbb{R}^5$) gerado pelos vetores

$$\vec{u} = \begin{bmatrix}1\\0\\2\\0\\5\end{bmatrix} \quad \text{e} \quad \vec{v} = \begin{bmatrix}1\\-1\\2\\0\\3\end{bmatrix}.$$



Para isto, consideramos o vetor

$$\vec{y} = \begin{bmatrix} 1 \\ 2 \\ -1 \\ 0 \\ 7 \end{bmatrix}$$

e vamos calcular $\|\vec{y} - \mathrm{proj}_W \vec{y}\|$:

$$\vec{y} - \mathrm{proj}_W \vec{y} = \vec{y} - \frac{\vec{y}\cdot\vec{u}}{\vec{u}\cdot\vec{u}}\,\vec{u} - \frac{\vec{y}\cdot\vec{v}}{\vec{v}\cdot\vec{v}}\,\vec{v} = \begin{bmatrix} 1 \\ 2 \\ -1 \\ 0 \\ 7 \end{bmatrix} - \frac{34}{30}\begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \\ 5 \end{bmatrix} - \frac{18}{15}\begin{bmatrix} 1 \\ -1 \\ 2 \\ 0 \\ 3 \end{bmatrix}$$
$$= \begin{bmatrix} 1 \\ 2 \\ -1 \\ 0 \\ 7 \end{bmatrix} - \frac{17}{15}\begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \\ 5 \end{bmatrix} - \frac{18}{15}\begin{bmatrix} 1 \\ -1 \\ 2 \\ 0 \\ 3 \end{bmatrix} = \begin{bmatrix} -4/3 \\ 16/5 \\ -17/3 \\ 0 \\ -34/15 \end{bmatrix}$$

Desta forma, a distância de P a W é

$$\mathrm{dist}(P, W) = \|\vec{y} - \mathrm{proj}_W \vec{y}\| = \sqrt{\frac{16}{9} + \frac{256}{25} + \frac{289}{9} + \frac{1156}{225}} = \sqrt{\frac{16\cdot 25 + 256\cdot 9 + 289\cdot 25 + 1156}{225}}$$
$$= \frac{\sqrt{11085}}{15} \simeq 7{,}02.$$

5.3 O Processo de ortogonalização de Gram–Schmidt

O **Processo de Gram–Schmidt** é um algoritmo para obter uma base ortogonal (ou ortonormal) a partir de uma base qualquer. De maneira mais geral, o método permite transformar um conjunto de vetores linearmente independentes em um conjunto ortogonal que gera o mesmo espaço vetorial.

Vamos começar com um exemplo.

Exemplo 25. Consideramos os vetores

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad \text{e} \quad \vec{v}_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

O espaço gerado por $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ é todo o espaço $\mathbb{R}^3$, já que temos três vetores linearmente independentes em um espaço de dimensão três.

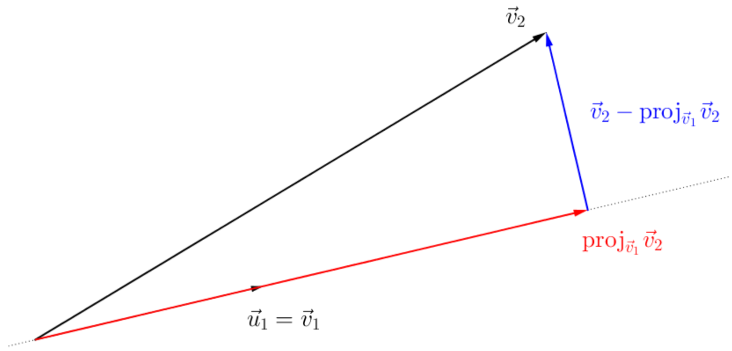
O processo consiste em fixar qualquer um dos vetores como o primeiro dos vetores do conjunto que virá a ser ortogonal. Por exemplo, fixamos

$$\vec{u}_1 = \vec{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

Em seguida, já vimos que, ao fazer a projeção de $\vec{v}_2$ sobre $\vec{v}_1$, o vetor

$$\vec{v}_2 - \mathrm{proj}_{\vec{v}_1} \vec{v}_2$$

é ortogonal a $\vec{v}_1$, como ilustra a figura.



Assim sendo, definimos o segundo vetor do nosso conjunto que será ortogonal como o vetor

$$\vec{u}_2 = \vec{v}_2 - \mathrm{proj}_{\vec{u}_1} \vec{v}_2 = \vec{v}_2 - \frac{\vec{v}_2 \cdot \vec{u}_1}{\vec{u}_1 \cdot \vec{u}_1} \vec{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1 \\ -1/2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}.$$

Temos que $\vec{u}_1$ e $\vec{u}_2$ são ortogonais e que estão no mesmo plano, de modo que também temos

$$\mathrm{Span}\{\vec{u}_1, \vec{u}_2\} = \mathrm{Span}\{\vec{v}_1, \vec{v}_2\}.$$

Vamos escrever momentariamente $W = \mathrm{Span}\{\vec{u}_1, \vec{u}_2\}$.

No próximo passo do processo, o terceiro vetor pode ser obtido como

$$\vec{u}_3 = \vec{v}_3 - \mathrm{proj}_W \vec{v}_3,$$

pois, desta forma, $\vec{u}_3$ é ortogonal a todos os vetores de W; em particular, é ortogonal a ambos $\vec{u}_1$ e $\vec{u}_2$. Além disso, como $\vec{u}_1$ e $\vec{u}_2$ já são vetores ortogonais, podemos calcular a projeção sobre W como de costume:

$$\vec{u}_3 = \vec{v}_3 - \mathrm{proj}_W \vec{v}_3 = \vec{v}_3 - \frac{\vec{v}_3 \cdot \vec{u}_1}{\vec{u}_1 \cdot \vec{u}_1} \vec{u}_1 - \frac{\vec{v}_3 \cdot \vec{u}_2}{\vec{u}_2 \cdot \vec{u}_2} \vec{u}_2.$$

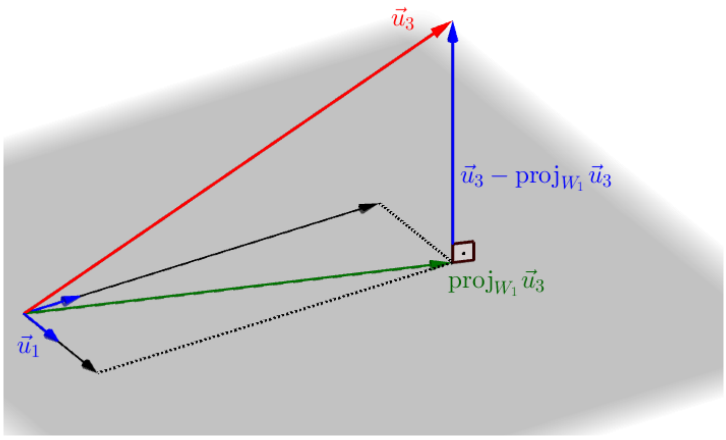
Calculando, temos (observe como dois fatores $1/2$ cancelam no último termo)

$$\vec{u}_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \frac{2-1}{1+4+1} \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{6} \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} -2/3 \\ 2/3 \\ 2/3 \end{bmatrix} = \frac{2}{3} \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

Concluimos assim que o conjunto

$$\{\vec{u}_1, \vec{u}_2, \vec{u}_3\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1/2 \\ 1 \\ -1/2 \end{bmatrix}, \begin{bmatrix} -2/3 \\ 2/3 \\ 2/3 \end{bmatrix} \right\}$$

é ortogonal e gera o mesmo espaço que $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$.



Na realidade, se tivéssemos considerado múltiplos dos vetores acima, não perderíamos a propriedade de ortogonalidade, e temos

$$\mathrm{Span}\{\vec{u}_1, \vec{u}_2, \vec{u}_3\} = \mathrm{Span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

"Colocar em evidência" os fatores comuns desta forma e "desconsiderá-los" pode facilitar as contas em alguns casos.

Observamos também que, fazendo a normalização dos vetores, podemos obter uma base ortonormal:

$$\begin{cases} \|\vec{u}\| = \sqrt{1+0+1} = \sqrt{2} \\ \|\vec{v}\| = \sqrt{1+4+1} = \sqrt{6} \\ \|\vec{w}\| = \sqrt{1+1+1} = \sqrt{3} \end{cases}$$

e assim,

$$\frac{\vec{u}_1}{\|\vec{u}_1\|} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \quad \frac{\vec{u}_2}{\|\vec{u}_2\|} = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{6} \\ 2/\sqrt{6} \\ -1/\sqrt{6} \end{bmatrix} \text{ e } \frac{\vec{u}_3}{\|\vec{u}_3\|} = \frac{1}{\sqrt{3}} \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$$

forma um conjunto ortonormal que gera o mesmo espaço vetorial que $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$. $\triangleleft$

Em seguida, mostramos como funciona o processo de maneira geral (um bom entendimento do exemplo acima deve deixar claro o porquê deste processo funcionar): sejam k vetores linearmente independentes $\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_k$ pertencentes ao espaço n-dimensional $\mathbb{R}^n$. Vamos denotar

$$W = \mathrm{Span}\{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_k\}$$

e construir uma base ortogonal (e em seguida ortonormal) para W. Começamos com

$$\boxed{\vec{u}_1 = \vec{v}_1}$$

e definimos

$$\boxed{\vec{u}_2 = \vec{v}_2 - \mathrm{proj}_{\vec{u}_1} \vec{v}_2 = \vec{v}_2 - \frac{\vec{v}_2 \cdot \vec{u}_1}{\vec{u}_1 \cdot \vec{u}_1} \vec{u}_1.}$$

Temos que $\vec{u}_1$ e $\vec{u}_2$ são ortogonais e que estão no mesmo plano, de modo que também temos

$$\mathrm{Span}\{\vec{u}_1, \vec{u}_2\} = \mathrm{Span}\{\vec{v}_1, \vec{v}_2\} = W_2 \quad \rightsquigarrow \quad \text{notação.}$$

Em seguida, definimos

$$\boxed{\vec{u}_3 = \vec{v}_3 - \mathrm{proj}_{W_2} \vec{v}_3 = \vec{v}_3 - \frac{\vec{v}_3 \cdot \vec{u}_1}{\vec{u}_1 \cdot \vec{u}_1} \vec{u}_1 - \frac{\vec{v}_3 \cdot \vec{u}_2}{\vec{u}_2 \cdot \vec{u}_2} \vec{u}_2.}$$

Temos que $\vec{u}_1, \vec{u}_2, \vec{u}_3$ ortogonais e

$$\mathrm{Span}\{\vec{u}_1, \vec{u}_2, \vec{u}_3\} = \mathrm{Span}\{\vec{v}_1, \vec{v}_2, \vec{v}_3\} = W_3 \quad \rightsquigarrow \text{ notação.}$$

Em seguida

$$\boxed{\vec{u}_4 = \vec{v}_4 - \mathrm{proj}_{W_3} \vec{v}_4 = \vec{v}_4 - \frac{\vec{v}_4 \cdot \vec{u}_1}{\vec{u}_1 \cdot \vec{u}_1} \vec{u}_1 - \frac{\vec{v}_4 \cdot \vec{u}_2}{\vec{u}_2 \cdot \vec{u}_2} \vec{u}_2 - \frac{\vec{v}_4 \cdot \vec{u}_3}{\vec{u}_3 \cdot \vec{u}_3} \vec{u}_3.}$$

Temos que $\vec{u}_1, \vec{u}_2, \vec{u}_3, \vec{u}_4$ ortogonais e

$$\mathrm{Span}\{\vec{u}_1, \vec{u}_2, \vec{u}_3, \vec{u}_4\} = \mathrm{Span}\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4\} = W_4 \quad \rightsquigarrow \text{ notação.}$$

E assim seguimos fazendo as contas até chegar ao último vetor: sendo

$$\mathrm{Span}\{\vec{u}_1, \vec{u}_2, \cdots, \vec{u}_{k-1}\} = \mathrm{Span}\{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_{k-1}\} = W_{k-1} \quad \rightsquigarrow \text{ notação,}$$

escrevemos

$$\boxed{\vec{u}_k = \vec{v}_k - \mathrm{proj}_{W_{k-1}} \vec{v}_k = \vec{v}_k - \frac{\vec{v}_k \cdot \vec{u}_1}{\vec{u}_1 \cdot \vec{u}_1} \vec{u}_1 - \frac{\vec{v}_k \cdot \vec{u}_2}{\vec{u}_2 \cdot \vec{u}_2} \vec{u}_2 - \cdots - \frac{\vec{v}_k \cdot \vec{u}_{k-1}}{\vec{u}_{k-1} \cdot \vec{u}_{k-1}} \vec{u}_{k-1}.}$$

Temos que $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ é um conjunto ortogonal e

$$\mathrm{Span}\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\} = \mathrm{Span}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{u}_k\} = W \quad \rightsquigarrow \text{ que é o espaço inicial,}$$

como queríamos.

Este é a parte mais "pesada" do método. Por exemplo, se agora quiséssemos uma base ortonormal para W, basta considerar o conjunto

$$\left\{ \frac{\vec{u}_1}{\|\vec{u}_1\|}, \frac{\vec{u}_2}{\|\vec{u}_2\|}, \ldots, \frac{\vec{u}_k}{\|\vec{u}_k\|} \right\},$$

que continua sendo ortogonal, mas agora formado por vetores unitários.

Exemplo 26 (Retirado de Wikipedia.org)**.** Vamos obter uma base ortonormal para o espaço coluna $\mathrm{Col}\, A$ da matriz

$$A = \begin{bmatrix} 12 & -51 & 4 \\ 6 & 167 & -68 \\ -4 & 24 & -41 \end{bmatrix}.$$

Lembrando que $\mathrm{Col}\, A$ é o espaço gerado pelas colunas de A, vamos aplicar o processo de Gram–Schmidt para os vetores

$$\vec{v}_1 = \begin{bmatrix} 12 \\ 6 \\ -4 \end{bmatrix}, \vec{v}_2 = \begin{bmatrix} -51 \\ 167 \\ 24 \end{bmatrix}, \vec{v}_3 = \begin{bmatrix} 4 \\ -68 \\ -41 \end{bmatrix}.$$

O Processo começa como:

$$\vec{u}_1 = \vec{v}_1 = \begin{bmatrix} 12 \\ 6 \\ -4 \end{bmatrix},$$

$$\vec{u}_2 = \vec{v}_2 - \frac{\vec{v}_2 \cdot \vec{u}_1}{\vec{u}_1 \cdot \vec{u}_1} \vec{u}_1 = \begin{bmatrix} -51 \\ 167 \\ 24 \end{bmatrix} - \frac{-612 + 1002 - 96}{144 + 36 + 16} \begin{bmatrix} 12 \\ 6 \\ -4 \end{bmatrix} = \begin{bmatrix} -51 \\ 167 \\ 24 \end{bmatrix} - \frac{296}{196} \begin{bmatrix} 12 \\ 6 \\ -4 \end{bmatrix} = \begin{bmatrix} -69 \\ 158 \\ 30 \end{bmatrix}$$

Em seguida (conferir estas contas com o auxílio de uma calculadora)

$$\vec{u}_3 = \vec{v}_3 - \frac{\vec{v}_3 \cdot \vec{u}_1}{\vec{u}_1 \cdot \vec{u}_1} \vec{u}_1 - \frac{\vec{v}_3 \cdot \vec{u}_2}{\vec{u}_2 \cdot \vec{u}_2} \vec{u}_2 = \begin{bmatrix} 4 \\ -68 \\ -41 \end{bmatrix} - \frac{48 - 408 + 164}{196} \begin{bmatrix} 12 \\ 6 \\ -4 \end{bmatrix} - \frac{-276 - 10744 - 1230}{4761 + 24694 + 900} \begin{bmatrix} -69 \\ 158 \\ 30 \end{bmatrix}$$
$$= \begin{bmatrix} 4 \\ -68 \\ -41 \end{bmatrix} + \frac{19\cancel{6}}{\cancel{1}96} \begin{bmatrix} 12 \\ 6 \\ -4 \end{bmatrix} + \frac{12250}{30355} \begin{bmatrix} -69 \\ 158 \\ 30 \end{bmatrix} = \begin{bmatrix} -58/5 \\ 6/5 \\ -33 \end{bmatrix}.$$

Para obter uma base ortonormal para o espaço coluna, vamos ainda considerar

$$\frac{\vec{u}_1}{\|\vec{u}_1\|}, \frac{\vec{u}_2}{\|\vec{u}_2\|}, \frac{\vec{u}_3}{\|\vec{u}_3\|}.$$

Calculamos

$$\begin{cases} \|\vec{u}_1\| = \sqrt{196} = 14 \\ \|\vec{u}_2\| = \sqrt{30625} = 175 \\ \|\vec{u}_3\| = \dfrac{\sqrt{30625}}{5} = 35 \end{cases}$$

Portanto, uma base ortonormal para $\operatorname{Col} A$ é

$$\vec{n}_1 = \frac{\vec{u}_1}{\|\vec{u}_1\|} = \begin{bmatrix} 6/7 \\ 3/7 \\ -2/7 \end{bmatrix}, \quad \vec{n}_2 = \frac{\vec{u}_2}{\|\vec{u}_2\|} = \begin{bmatrix} -69/175 \\ 158/175 \\ 6/35 \end{bmatrix} \quad \text{e} \quad \vec{n}_3 = \frac{\vec{u}_3}{\|\vec{u}_3\|} = \begin{bmatrix} -58/175 \\ 6/175 \\ -33/35 \end{bmatrix}. \triangleleft$$

5.4 Fatoração QR e matrizes ortogonais

O método de Gram-Schmidt nos permite fatorar uma matriz A cujas colunas são LI em um produto

$$A = QR,$$

onde Q é uma matriz cujas colunas formam um conjunto ortonormal e R é uma matriz triangular superior. Tal fatoração é muito importante do ponto de vista computacional, como veremos em breve, quando usarmos a fatoração QR como uma alternativa para resolução de sistemas lineareas por mínimos quadrados. Também é a base de um eficiente método numérico para encontrar autovalores de uma matriz quadrada.

Matrizes cujas colunas formam um conjunto ortonormal são muito úteis e por isso recebem um nome especial:são chamadas de matrizes ortogonais. Aqui cabe uma observação relevante: este é um exemplo de uma situação onde a terminologia clássica está longe de ser satisfatória: matrizes cujas colunas formam um conjunto *ortonormal* são chamadas de matrizes *ortogonais*.

Faremos agora uma exposição da fatoração QR usando uma matriz inicial A que tem 3 colunas, observando que a construção abaixo é facilmente generalizada para uma matriz de qualquer dimensão (desde que tenha colunas LI).

Dada uma matriz $A = [\vec{v}_1 \vec{v}_2 \vec{v}_3]$ cujas colunas são LI e são denotadas por $\vec{v}_1$, $\vec{v}_2$, e $\vec{v}_3$, o processo de Gram-Schmidt começa com a construção de uma base ortogonal formada pelos vetores $\vec{u}_1$, $\vec{u}_2$, e $\vec{u}_3$, para o espaço-coluna da matriz A. Para isso, definimos

$$\begin{cases} \vec{u}_1 = \vec{v}_1 \\ \vec{u}_2 = \vec{v}_2 - \operatorname{proj}_{\vec{u}_1} \vec{v}_2 \\ \vec{u}_3 = \vec{v}_3 - \operatorname{proj}_{\vec{u}_1, \vec{u}_2} \vec{v}_3 \end{cases}$$

Portanto existem coeficientes b_1, c_1 e c_2 tais que

$$\begin{cases} \vec{u}_1 = \vec{v}_1 \\ \vec{u}_2 = \vec{v}_2 - b_1 \vec{u}_1 \\ \vec{u}_3 = \vec{v}_3 - c_1 \vec{u}_1 - c_2 \vec{u}_2 \end{cases}$$

Em um segundo momento, cada um dos três vetores $\vec{u}_1$, $\vec{u}_2$, e $\vec{u}_3$ é normalizado, ou seja, dividimos cada um deles pelo seu comprimento. Assim, $\vec{u}_1$, $\vec{u}_2$, e $\vec{u}_3$ formam uma base ortonormal para o espaço coluna de A, e existem coeficientes α_1, β_1, β_2, γ_1, γ_2 e γ_3 tais que

$$\begin{cases} \vec{u}_1 = \alpha_1 \vec{v}_1 \\ \vec{u}_2 = \beta_2 \vec{v}_2 - \beta_1 \vec{u}_1 \\ \vec{u}_3 = \gamma_3 \vec{v}_3 - \gamma_1 \vec{u}_1 - \gamma_2 \vec{u}_2 \end{cases}$$

Se colocarmos as colunas da matriz A em função dos vetores da base ortonormal que acaba de ser obtida, temos

$$\begin{cases} \vec{v}_1 = \alpha_1^{-1}\vec{u}_1 \\ \vec{v}_2 = \beta_2^{-1}\vec{u}_2 + \beta_2^{-1}\beta_1\vec{u}_1 \\ \vec{v}_3 = \gamma_3^{-1}\vec{u}_3 + \gamma_3^{-1}\gamma_1\vec{u}_1 + \gamma_3^{-1}\gamma_2\vec{u}_2 \end{cases}$$

ou trocando a ordem das parcelas no lado direito,

$$\begin{cases} \vec{v}_1 = \alpha_1^{-1}\vec{u}_1 \\ \vec{v}_2 = \beta_2^{-1}\beta_1\vec{u}_1 + \beta_2^{-1}\vec{u}_2 \\ \vec{v}_3 = \gamma_3^{-1}\gamma_1\vec{u}_1 + \gamma_3^{-1}\gamma_2\vec{u}_2 + \gamma_3^{-1}\vec{u}_3 \end{cases}$$

Assim, ao definirmos a matriz $Q = [\vec{u}_1\vec{u}_2\vec{u}_3]$, obtemos

$$A = QR$$

onde

$$R = \begin{bmatrix} \alpha_1^{-1} & \beta_2^{-1}\beta_1 & \gamma_3^{-1}\gamma_1 \\ 0 & \beta_2^{-1} & \gamma_3^{-1}\gamma_2 \\ 0 & 0 & \gamma_3^{-1} \end{bmatrix}$$

visto que cada coluna do produto QR, é definido como a combinação linear das colunas da matriz Q com pesos dados pela respectiva coluna da matriz R.

Note ainda que R é invertível, visto que sua diagonal principal contém apenas termos não-nulos.

A construção acima é facilmente generalizada para matrizes quaisquer e nos permite enunciar o seguinte fato (lembre que uma matriz ortogonal é uma matriz cujas colunas formam um conjunto ortonormal):

Teorema 27. *Qualquer matriz A cujas colunas são linearmente independentes pode ser fatorada na forma $A = QR$, onde Q é uma matriz ortogonal e R é uma matriz triangular superior invertível. Ainda, o espaço coluna das matrizes A e Q é o mesmo, ou seja, as colunas de Q formam uma base ortonormal para o subespaço gerado pelas colunas de A.*

Note que a fatoração QR está para o processo de Gram-Schmidt assim como a fatoração LU está para o método de linha-redução a forma escada!

Bem, vimos que a matriz Q tem como colunas a base ortonormal gerada pelo processo de Gram-Schmidt. E como podemos calcular a matriz R no caso geral ?

Usando o seguinte fato surpreendente:

Teorema 28. *uma matriz ortogonal é sempre invertível e sua inversa é dada pela sua transposta, ou seja $Q^{-1} = Q^T$.*

A prova deste teorema é simples:

$$Q^TQ = \begin{bmatrix} — & \vec{u}_1 & — \\ — & \vec{u}_2 & — \\ — & \vec{u}_3 & — \end{bmatrix} \begin{bmatrix} | & | & | \\ \vec{u}_1 & \vec{u}_2 & \vec{u}_3 \\ | & | & | \end{bmatrix} = \begin{bmatrix} \vec{u}_1\cdot\vec{u}_1 & \vec{u}_1\cdot\vec{u}_2 & \vec{u}_1\cdot\vec{u}_2 \\ \vec{u}_2\cdot\vec{u}_1 & \vec{u}_2\cdot\vec{u}_2 & \vec{u}_2\cdot\vec{u}_3 \\ \vec{u}_3\cdot\vec{u}_1 & \vec{u}_3\cdot\vec{u}_2 & \vec{u}_3\cdot\vec{u}_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I.$$

Assim, se $A = QR$, basta multiplicar ambos os lados desta equação por Q^T pela esquerda para obtermos:

$$Q^TA = Q^TQR \Longrightarrow R = Q^TA.$$

No último exemplo da seção anterior, obtemos através do método de Gram-Schmidt uma base ortonormal para $\operatorname{Col} A$ onde

$$A = \begin{bmatrix} 12 & -51 & 4 \\ 6 & 167 & -68 \\ -4 & 24 & -41 \end{bmatrix}.$$

Vamos definir a matriz Q como a matriz formada pelos vetores desta base ortonormal:

$$Q = \begin{bmatrix} 6/7 & -69/175 & -58/175 \\ 3/7 & 158/175 & 6/175 \\ -2/7 & 6/35 & -33/35 \end{bmatrix}$$

Vimos que $R = Q^T A$ e portanto

$$R = Q^T A = \begin{bmatrix} 6/7 & 3/7 & -2/7 \\ -69/175 & 158/175 & 6/35 \\ -58/175 & 6/175 & -33/35 \end{bmatrix} \begin{bmatrix} 12 & -51 & 4 \\ 6 & 167 & -68 \\ -4 & 24 & -41 \end{bmatrix} = \begin{bmatrix} 14 & 21 & -14 \\ 0 & 175 & -70 \\ 0 & 0 & 35 \end{bmatrix}.$$

Finalmente podemos escrever

$$A = \begin{bmatrix} 12 & -51 & 4 \\ 6 & 167 & -68 \\ -4 & 24 & -41 \end{bmatrix} = \begin{bmatrix} 6/7 & -69/175 & -58/175 \\ 3/7 & 158/175 & 6/175 \\ -2/7 & 6/35 & -33/35 \end{bmatrix} \begin{bmatrix} 14 & 21 & -14 \\ 0 & 175 & -70 \\ 0 & 0 & 35 \end{bmatrix} = QR. \triangleleft$$

5.5 Matrizes Ortogonais preservam comprimento e ângulo

Vamos aproveitar para falar um pouco mais sobre matrizes ortogonais. Um bom exemplo de matriz ortogonal 2×2 é dada pela matriz de rotação, ou seja, pela matriz

$$Q = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix},$$

onde $0 \leq \theta \leq 2\pi$. Podemos ver que esta matriz define uma transformação linear no plano que rotaciona vetores em um ângulo θ no sentido anti-horário. Assim, tal transformação preserva comprimento e ângulo entre vetores. Estas propriedades não são particulares da matriz de rotação: todas as matrizes ortogonais se comportam da mesma forma. O teorema a seguir traz estas informações e mais algumas.

Teorema 29. *Seja $Q = [\vec{u}_1, ..., \vec{u}_n]$ uma matriz ortogonal $n \times n$. Então, para quaisquer vetores $\vec{v}$ e $\vec{w}$ em $\mathbb{R}^n$ temos*

i. $Q\vec{v} \cdot Q\vec{w} = \vec{v} \cdot \vec{w}$

ii. $\|Q\vec{v}\| = \|\vec{v}\|$

iii. o ângulo entre $Q\vec{v}$ e $Q\vec{w}$ é igual ao ângulo entre $\vec{v}$ e $\vec{w}$

A prova do primeiro item é consequência direta da linearidade do produto interno e da hipótese das colunas formarem uma base ortonormal: se $\vec{v} = (\alpha_1, ..., \alpha_n)$ e $\vec{w} = (\beta_1, ..., \beta_n)$, então

$$Q\vec{v} \cdot Q\vec{w} = \left(\sum_{i=1}^{n} \alpha_i \vec{u}_i \right) \cdot \left(\sum_{j=1}^{n} \beta_j \vec{u}_j \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \beta_j \; \vec{u}_i \cdot \vec{u}_j$$

devido a linearidade do produto interno (ou seja a propriedade distributiva aplicada tanto na primeira quanto na segunda componente do produto interno). Agora, lembrando que $\vec{u}_i \cdot \vec{u}_j = 0$ se $i \neq j$, a ultima expressão iguala-se a

$$\sum_{i=1}^{n} \alpha_i \beta_i \; \vec{u}_i \cdot \vec{u}_i.$$

Finalmente, lembrando que cada coluna $\vec{u}_i$ de Q tem comprimento unitário, concluimos que

$$Q\vec{v} \cdot Q\vec{w} = \sum_{i=1}^{n} \alpha_i \beta_i = \vec{v} \cdot \vec{w}.$$

Os próximos dois itens seguem diretamente do primeiro item, visto que

$$\|Q\vec{v}\| = \sqrt{Q\vec{v} \cdot Q\vec{v}} = \sqrt{\vec{v} \cdot \vec{v}} = \|\vec{v}\|$$

e o ângulo entre Qv e Qw é definido como

$$\arccos\left(\frac{Qv \cdot Qw}{\|Qv\|\|Qw\|}\right) = \arccos\left(\frac{v \cdot w}{\|v\|\|w\|}\right),$$

que por definição é o ângulo entre v e w. Assim finalizamos a prova do teorema.

Um outro exemplo de matriz ortogonal é dado pelas matrizes de reflexão. Um caso particular é dado pela reflexão em torno do eixo horizontal:

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

Finalizamos esta seção enunciando um teorema cujo primeiro item já foi provado e cujo segundo item é consequencia direta do primeiro e do fato de que $\det(AB) = \det(A)\det(B)$: PS NAO VALE A PENA PRECISA DE ALGO MAIS...

Teorema 30. *Seja* $Q = [\vec{u}_1, ..., \vec{u}_n]$ *uma matriz ortogonal. Temos*

i. $Q^{-1} = Q^T$

ii. $\det(Q) = 1$ *ou* $\det(Q) = -1$.

6

Semana 13

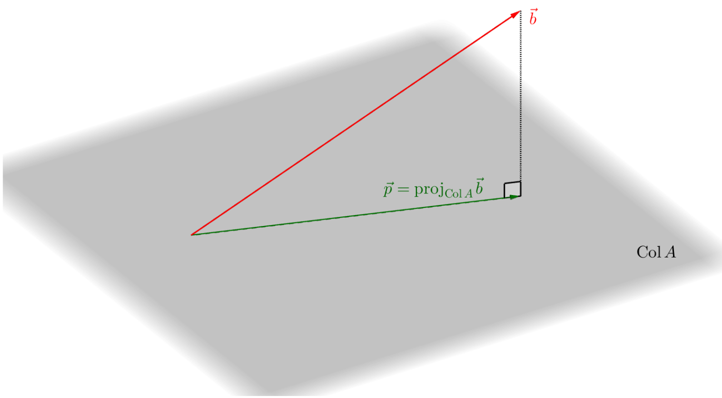
6.1 Método dos Mínimos Quadrados

Sistemas lineares aparecem como modelos matemáticos de vários fenômenos e em várias situações. Acontece que alguns sistemas simplesmente não possuem soluções e ficamos sem saber como proceder. O **Método dos Mínimos Quadrados** é uma técnica que nos permite, de forma aproximada, retirar alguma informação desses sistemas impossíveis. A terminologia se deve ao fato de que, como veremos, este método minimiza a soma dos quadrados dos erros obtidos na aproximação.

Como sabemos, resolver o sistema linear

$$A\vec{x} = \vec{b}$$

consiste em encontrar um vetor $\vec{x}$ que satisfaça a esta equação. Na terminologia que construimos ao longo do curso, isto significa que $\vec{b}$ pertence ao espaço coluna da matriz A, isto é, $\vec{b} \in \operatorname{Col} A$. Desta forma, não é possível resolver o sistema quando $\vec{b} \notin \operatorname{Col} A$.



O Método dos Mínimos Quadrados consiste em olhar para o vetor $\vec{p} = \operatorname{proj}_{\operatorname{Col} A} \vec{b}$ e resolver o sistema linear associado

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

Esta solução de $A\vec{x} = \vec{p}$ é chamada de **solução de mínimos quadrados**. A ideia é (ver figura) considerar o vetor em $\operatorname{Col} A$ que é “mais próximo” de $\vec{b}$ e cujo sistema linear associado possua solução. Neste contexto, estamos pensando em mais próximo no sentido que

$$\|\vec{b} - \operatorname{proj}_{\operatorname{Col} A} \vec{b}\| \leq \|\vec{b} - \vec{c}\|, \text{ para qualquer } \vec{c} \in \operatorname{Col} A,$$

isto é, no sentido de ser a projeção a melhor aproximação de $\vec{b}$ no espaço coluna de A. Escrevendo os vetores em coordenadas:

$$\vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \quad \text{e} \quad \operatorname{proj}_{\operatorname{Col} A} \vec{b} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix},$$

temos que o valor

$$\|\vec{b} - \operatorname{proj}_{\operatorname{Col} A} \vec{b}\| = \sum_{i=1}^{n} (b_i - p_i)^2$$

é chamado de **erro da aproximação**. Logo, como anunciado, a soma dos quadrados dos erros obtidos cada componente é o mínimo possível.

Exemplo 31. Considere o sistema linear

$$\begin{cases} x_1 + x_2 - 2x_3 = 3 \\ x_1 - 2x_3 = 2 \\ x_2 + 2x_3 = 0 \\ x_1 + x_2 - 2x_3 = 0 \end{cases} \rightsquigarrow A\vec{x} = \begin{bmatrix} 1 & 1 & -2 \\ 1 & 0 & -2 \\ 0 & 1 & 2 \\ -1 & -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 0 \\ 0 \end{bmatrix} = \vec{b}.$$

Este sistema é claramente impossível, pois a primeira equação é inconsistente com a última. De forma mais automática, um passo no escalonamento da matriz aumentada associada revelaria a mesma conclusão:

$$[\, A \mid \vec{b}\,] = \begin{bmatrix} 1 & 1 & -2 & 3 \\ 1 & 0 & -2 & 2 \\ 0 & 1 & 2 & 0 \\ -1 & -1 & 2 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & -2 & 3 \\ 1 & 0 & -2 & 2 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}.$$

Vamos tentar encontrar, conforme descrito acima, uma solução de mínimos quadrados. Passos a serem realizados:

- Para calcular a projeção de $\vec{b}$ sobre $\operatorname{Col} A$, precisamos, em primeiro lugar, obter uma base ortogonal do espaço $\operatorname{Col} A$; isto pode ser feito pelo Processo de Gram–Schmidt;
- Calcular a projeção $\vec{p} = \operatorname{proj}_{\operatorname{Col} A} \vec{b}$;
- Resolver o sistema linear $A\vec{x} = \vec{p}$.

Embora seja possível realizar estes três passos, temos de convir que seria muito trabalhoso. Por isto, vamos analisar um pouco mais da teoria para encontrar um método mais eficaz. ◁

Suponhamos que:

- A é uma matriz $m \times n$,
- $\vec{x} \in \mathbb{R}^n$,
- $\vec{b} \in \mathbb{R}^m$ e
- o sistema $A\vec{x} = \vec{b}$ não possui solução.

A solução de mínimos quadrados é a solução de $A\vec{x} = \vec{p}$, onde $\vec{p} = \operatorname{proj}_{\operatorname{Col} A} \vec{b}$. Observamos, por outro lado, que, sendo $\vec{p}$ a projeção sobre o espaço coluna de A, o vetor $\vec{b}-\vec{p}$ deve ser ortogonal a todos os elementos de $\operatorname{Col} A$. Em particular, se escrevermos A em termos de suas colunas

$$A = \begin{bmatrix} | & | & & | \\ \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \\ | & | & & | \end{bmatrix}, \qquad \left(\text{que, em particular, implica } A^T = \begin{bmatrix} — & \vec{a}_1 & — \\ — & \vec{a}_2 & — \\ & \vdots & \\ — & \vec{a}_n & — \end{bmatrix} \right)$$

devemos ter

$$\vec{a}_j \cdot (\vec{b} - \vec{p}) = 0, \text{ para todo } j \iff A^T(\vec{b} - \vec{p}) = \vec{0} \iff A^T\vec{b} = A^T\vec{p} = A^T A\vec{x}.$$

Esta última linha é válida porque o produto escalar $\vec{a}_j \cdot (\vec{b} - \vec{p})$ é justamente a entrada j do vetor obtido ao multiplicar $A^T(\vec{b} - \vec{p})$.

Concluimos, desta forma, que, se $\vec{x}$ é uma solução de mínimos quadrados, ou seja, se $A\vec{x} = \vec{p}$, então necessariamente

$$\boxed{A^T A\vec{x} = A^T\vec{b}.} \tag{6.1}$$

Exercício 2 (Teórico)**.** Justifique que o conjunto de soluções do sistema (6.1) coincide com o conjunto das soluções de mínimos quadrados do sistema $A\vec{x} = \vec{b}$.

Tudo isto implica que podemos utilizar a equação (6.1) para encontrar as soluções de mínimos quadrados de $A\vec{x} = \vec{b}$.

Exemplo 32 (De volta ao Exemplo 31)**.** Calculamos

$$A^T A = \begin{bmatrix} 1 & 1 & 0 & -1 \\ 1 & 0 & 1 & -1 \\ -2 & -2 & 2 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & -2 \\ 1 & 0 & -2 \\ 0 & 1 & 2 \\ -1 & -1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 2 & -6 \\ 2 & 3 & -2 \\ -6 & -2 & 16 \end{bmatrix}$$

e

$$A^T\vec{b} = \begin{bmatrix} 1 & 1 & 0 & -1 \\ 1 & 0 & 1 & -1 \\ -2 & -2 & 2 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \\ -10 \end{bmatrix}$$

Para resolvermos o sistema $A^T A\vec{x} = A^T\vec{b}$, vamos escalonar a matriz aumentada associada:

$$\begin{bmatrix} 3 & 2 & -6 & 5 \\ 2 & 3 & -2 & 3 \\ -6 & -2 & 16 & -10 \end{bmatrix} \sim \begin{bmatrix} 3 & 2 & -6 & 5 \\ 0 & 5 & 6 & -1 \\ 0 & 2 & 4 & 0 \end{bmatrix} \sim \begin{bmatrix} 3 & 2 & -6 & 5 \\ 0 & 5 & 6 & -1 \\ 0 & 0 & 4 & 1 \end{bmatrix}$$

Logo,

$$\begin{cases} 3x_1 + 2x_2 - 6x_3 = 5 \\ 5x_2 + 6x_3 = -1 \\ 4x_3 = 1 \end{cases}.$$

Assim,

$$x_3 = \frac{1}{4} \implies 5x_2 + 6 \cdot \frac{1}{4} = -1 \implies x_2 = -\frac{1}{2} \implies 3x_1 - 1 - \frac{3}{2} = 5 \implies x_1 = \frac{3}{2}.$$

Portanto, uma solução de mínimos quadrados é

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3/2 \\ -1/2 \\ 1/4 \end{bmatrix}.$$

Observe como isto é muito menos trabalhoso do que o método que havíamos esquematizado quando discutimos este sistema no Exemplo 31. $\triangleleft$

Exercício 3. Colocar em prática o método discutido no Exemplo 31 e comparar o resultado com o que obtivemos no Exemplo 32.

Observação 33. Caso a matriz A já seja uma matriz ortogonal, podemos calcular projeções diretamente, pois uma base ortogonal para $\operatorname{Col} A$ já esté disponível desde o começo. Neste caso, ambos os métodos exigiriam um trabalho parecido.

Nossa aproximação linear foi encontrada de modo a minimizar o erro quadrático. Para encontrar o erro, deveríamos calcular

$$\|\vec{b} - \operatorname{proj}_{\operatorname{Col} A} \vec{b}\|.$$

No entanto, resolvemos de outra maneira e não calculamos a projeção. Observando que a solução de mínimos quadrados satisfaz $A\vec{x} = \operatorname{proj}_{\operatorname{Col} A} \vec{b}$, podemos calcular o erro por

$$\text{erro } = \|\vec{b} - A\vec{x}\|,$$

onde $\vec{x}$ é a solução de mínimos quadrados.

No Exemplo 32 acima, podemos calcular o erro desta forma. De fato,

$$A\vec{x} = \begin{bmatrix} 1 & 1 & -2 \\ 1 & 0 & -2 \\ 0 & 1 & 2 \\ -1 & -1 & 2 \end{bmatrix} \begin{bmatrix} 3/2 \\ -1/2 \\ 1/4 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1 \\ 0 \\ -1/2 \end{bmatrix} \implies \vec{b} - A\vec{x} = \begin{bmatrix} 5/2 \\ 1 \\ 0 \\ 1/2 \end{bmatrix}$$

$$\implies \text{erro } = \|\vec{b} - A\vec{x}\| = \sqrt{\frac{25}{4} + 1 + \frac{1}{4}} = \frac{\sqrt{30}}{2}.$$

6.2 Fatoração QR e Mínimos Quadrados

Vimos que o processo de Gram-Schmidt nos permite escrever uma matriz A cujas linhas são linearmente independentes na forma $A = QR$, onde Q é uma matriz ortogonal, R é triangular superior invertível, e as colunas de Q formam uma base ortonormal para o espaço-coluna de A. Chamamos esta fatoração de A de fatoração QR.

A fatoração QR tem uma importante aplicação na solução de problemas em mínimos quadrados. Ocorre que, em muitos problemas, a matriz $A^T A$ é uma matriz *mal-condicionada*, ou seja uma matriz que é muito sensível aos erros de aproximação cometidos em cálculos numéricos usualmente executados no tratamento de problemas aplicados. Os exemplos que virão a seguir nos mostram situações em que há uma grande variação de magnitude entre as entradas da matriz $A^T A$, e isto costuma ser fonte de instabilidade numérica.

Para isso, a fatoração QR oferece uma solução alternativa em mínimos quadrados para um sistema linear $Ax = b$: Se $A = QR$, então

$$A^T = R^T Q^T$$

e

$$A^T A = R^T Q^T Q R = R^T R,$$

onde usamos o fato de que $Q^{-1} = Q^T$. Assim as equações normais $A^T A\vec{x} = A^T\vec{b}$ se reduzem a

$$R^T R\vec{x} = R^T Q^T \vec{b}.$$

Ainda, como R é invertível podemos eliminar (cancelar) R^T na equação acima, obtendo o seguinte:

Teorema 34. *Se $A = QR$ com Q ortogonal e R triangular superior invertível, então a solução em mínimos quadrados do sistema $Ax = b$ é única e dada pela solução do sistema*

$$R\vec{x} = Q^T\vec{b}.$$

Note que o fato de R ser triangular superior torna a resolução do sistema acima pouco custosa do ponto de vista computacional.

Alternativamente temos uma expressão explícita para solução em mínimos quadrados:

$$\vec{x} = R^{-1} Q^T \vec{b}.$$

É importante notar que existem algoritmos para fatoração QR que são *numéricamente estáveis, ou seja pouco suscetiveis a erros de aproximação ou arredondamento. Tais algorimos são diferentes do processo de Gram Schmidt usual.*

6.3 Regressão Linear Simples

*Vamos apresentar uma aplicação de algumas das técnicas da seção anterior à Estatística ou Econometria. Uma **regressão linear simples** é uma equação, tipicamente da forma*

$$y = a + bx,$$

para estimar os valores (x, y) apenas conhecendo alguns valores específicos $(x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k)$. A ideia é tentar capturar como que mudanças na variável independente x afetam a variável dependente y.

Exemplo 35. [1] Com o intuito de analisar se é razoável supor que há um relação linear entre a idade de um motorista e quão longe ele consegue ver, uma empresa (Last Resource, Inc., Bellefonte, PA) coletou dados de 30 motoristas. Para simplificar as nossas contas, vamos listar abaixo *apenas alguns* destes dados.

Idade	Distância (em m)
20	590
32	410
41	460
49	380
66	350

Podemos pensar em y como a distância e em x como a idade. Gostaríamos de achar uma relacção linear da forma

$$y = a + bx.$$

Desta forma, os dados obtidos implicam que

$$\begin{cases} b + 20a = 590 \\ b + 32a = 410 \\ b + 41a = 460 \\ b + 49a = 380 \\ b + 66a = 350 \end{cases} \quad \rightsquigarrow \quad \begin{bmatrix} 1 & 20 \\ 1 & 32 \\ 1 & 41 \\ 1 & 49 \\ 1 & 66 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 590 \\ 410 \\ 460 \\ 380 \\ 350 \end{bmatrix}$$

Ora, dificilmente um sistema linear com duas incógnitas e cinco equações terá solução (só terá solução se todos os pontos do conjunto de dados estiverem alinhados em uma reta perfeita!). Vamos procurar por uma solução de mínimos quadrados. Isto é **regressão linear simpes**.

Denotando

$$A = \begin{bmatrix} 1 & 20 \\ 1 & 32 \\ 1 & 41 \\ 1 & 49 \\ 1 & 66 \end{bmatrix} \quad \text{e} \quad \vec{b} = \begin{bmatrix} 590 \\ 410 \\ 460 \\ 380 \\ 350 \end{bmatrix}$$

precisamos calcular

$$A^T A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 20 & 32 & 41 & 49 & 66 \end{bmatrix} \begin{bmatrix} 1 & 20 \\ 1 & 32 \\ 1 & 41 \\ 1 & 49 \\ 1 & 66 \end{bmatrix} = \begin{bmatrix} 5 & 208 \\ 208 & 9832 \end{bmatrix}$$

e

$$A^T \vec{b} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 20 & 32 & 41 & 49 & 66 \end{bmatrix} \begin{bmatrix} 590 \\ 410 \\ 460 \\ 380 \\ 350 \end{bmatrix} = \begin{bmatrix} 2190 \\ 85500 \end{bmatrix}.$$

Em seguida, a matriz associada aumentada pode ser reduzida por escalonamento:

$$[\, A^T A \mid \vec{b} \,] = \begin{bmatrix} 5 & 208 & 2190 \\ 208 & 9832 & 85500 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 468510/737 \\ 0 & 1 & -7005/1474 \end{bmatrix} \implies \begin{cases} a \simeq 635.7 \\ b \simeq -4.75 \end{cases}.$$

Os números são feios, mas as contas feitas foram as que sempre fizemos ao escalonar uma matriz até sua forma escalonada reduzida.

A conclusão é que a reta de mínimos quadrados que melhor aproxima os nossos dados é a reta

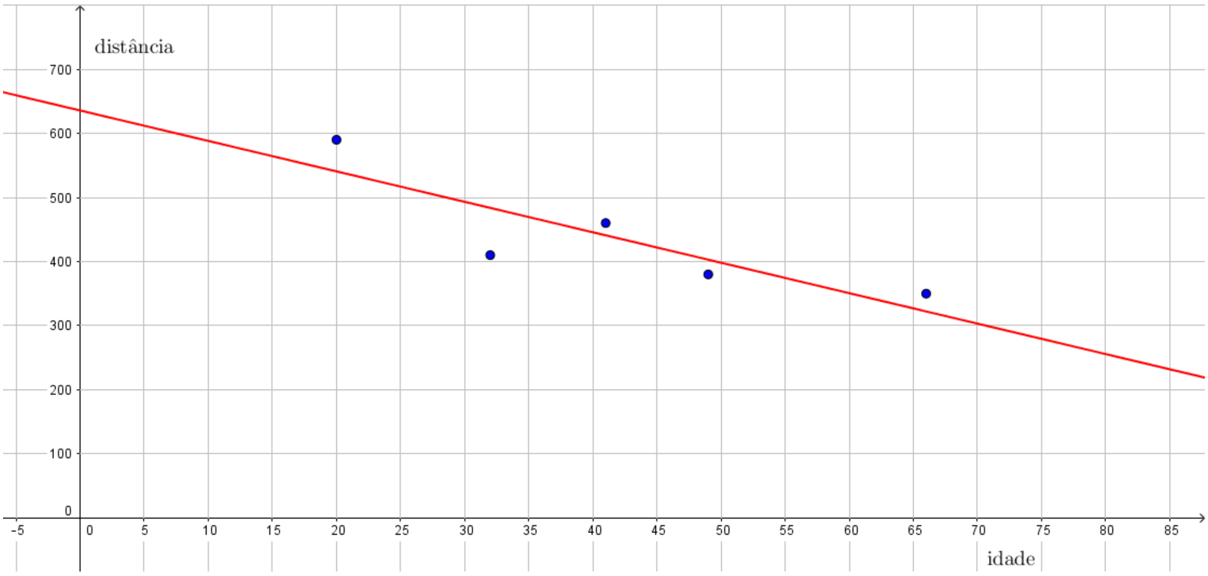
$$y = a + bx = 635.7 - 4.75x.$$

[1] Exemplo adaptado de https://onlinecourses.science.psu.edu/stat501/node/257

O erro de mínimos quadrados nesta aproximação (ou norma do resíduo) pode ser calculado como

$$A\vec{x} = \begin{bmatrix} 1 & 20 \\ 1 & 32 \\ 1 & 41 \\ 1 & 49 \\ 1 & 66 \end{bmatrix} \begin{bmatrix} 635.7 \\ 4.75 \end{bmatrix} \simeq \begin{bmatrix} 730.7 \\ 787.7 \\ 830.45 \\ 868.45 \\ 949.2 \end{bmatrix} \simeq \text{proj}_{\text{Col}\,A}\,\vec{b} \implies \text{ erro } = \|\vec{b} - A\vec{x}\| \simeq 947.3$$

Na figura abaixo, mostramos um esboço da reta que melhor aproxima os dados deste exemplo. ◁



De uma maneira geral, para encontrar uma ***reta de melhor ajuste*** *a uma quantidade de pontos (dados coletados de algum problema)*

$$(x_1, y_1),\ (x_2, y_2),\ \dots, (x_k, y_k),$$

devemos procurar pela solução (a, b) *de mínimos quadrados do sistema linear*

$$\begin{cases} a + bx_1 = y_1 \\ a + bx_2 = y_2 \\ \quad\vdots \\ a + bx_k = y_k \end{cases} \quad \leftrightsquigarrow \quad \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_k \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix}.$$

Vejamos um outro exemplo:

Exemplo 36. Encontrar a reta que melhor se ajusta aos pontos do plano

$$(1,2),\quad (2,3),\quad (2,4),\quad (3,2),\quad (4,3),\quad (4,4),\quad (5,3),\quad (5,5),\quad (6,4).$$

Como vimos acima, podemos considerar

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 4 \\ 1 & 5 \\ 1 & 5 \\ 1 & 6 \end{bmatrix} \quad \text{e} \quad \vec{b} = \begin{bmatrix} 2 \\ 3 \\ 4 \\ 2 \\ 3 \\ 4 \\ 3 \\ 5 \\ 4 \end{bmatrix}$$

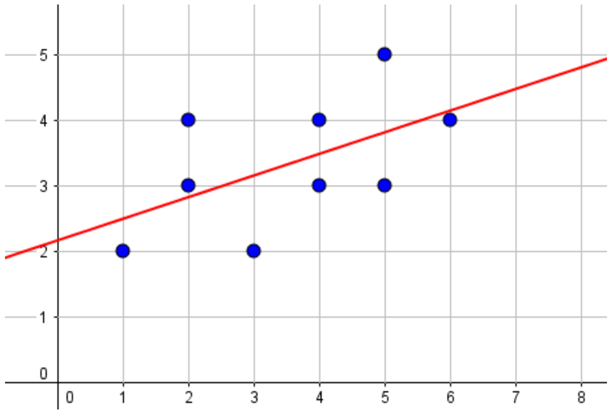
Assim,

$$A^TA = \begin{bmatrix} 9 & 32 \\ 32 & 136 \end{bmatrix} \quad \text{e} \quad A^T\vec{b} = \begin{bmatrix} 30 \\ 114 \end{bmatrix}.$$

A solução de mínimos quadrados, que é a solução de $A^TA\vec{x} = A^T\vec{b}$, é dada por

$$b = 54/25 = 2.16, a = 33/100 = 0.33.$$

A reta procurada é portanto $y = 0.33x + 2.16$.



6.4 Regressão não linear

Podemos também procurar (em contraste com a seção anterior) por funções não lineares que se ajustem a um conjunto de pontos. Por exemplo, dado um certo conjunto de pontos

$$(x_1, y_1),\ (x_2, y_2),\ \ldots, (x_k, y_k),$$

vamos procurar por uma parábola

$$y = a + bx + cx^2$$

que melhor se ajuste a este conjunto de pontos, no sentido de que a soma dos quadrados dos erros seja a menor possível. Isto corrensponde a procurar pela solução (a, b, c) *de mínimos quadrados do seguinte sistema linear*

$$\begin{cases} a + bx_1 + cx_1^2 = y_1 \\ a + bx_2 + cx_2^2 = y_2 \\ \vdots \\ a + bx_k + cx_k^2 = y_k \end{cases} \quad \rightsquigarrow \quad \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \\ 1 & x_k & x_k^2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix}.$$

Vejamos como ficaria a parábola que se ajusta aos dados do Exemplo 35:

Exemplo 37. Nossos pontos no Exemplo 35 são

$$(20, 590),\ (32, 410),\ (41, 460),\ (49, 380),\ (66, 350).$$

Para encontrar a parábola de melhor ajuste como acima, devemos procurar pela solução de mínimos quadrados do sistema

$$\begin{bmatrix} 1 & 20 & 400 \\ 1 & 32 & 1024 \\ 1 & 41 & 1681 \\ 1 & 49 & 2401 \\ 1 & 66 & 4356 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 590 \\ 410 \\ 460 \\ 380 \\ 350 \end{bmatrix} \quad \rightsquigarrow \quad A\vec{x} = \vec{b}.$$

Calculamos (com o auxílio de uma calculadora)

$$A^TA = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 20 & 32 & 41 & 49 & 66 \\ 400 & 1024 & 1681 & 2401 & 4356 \end{bmatrix} \begin{bmatrix} 1 & 20 & 400 \\ 1 & 32 & 1024 \\ 1 & 41 & 1681 \\ 1 & 49 & 2401 \\ 1 & 66 & 4356 \end{bmatrix} = \begin{bmatrix} 5 & 208 & 9862 \\ 208 & 9862 & 514834 \\ 9862 & 514834 & 28773874 \end{bmatrix},$$

$$A^TA = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 20 & 32 & 41 & 49 & 66 \\ 400 & 1024 & 1681 & 2401 & 4356 \end{bmatrix} \begin{bmatrix} 590 \\ 410 \\ 460 \\ 380 \\ 350 \end{bmatrix} = \begin{bmatrix} 2190 \\ 85500 \\ 3866080 \end{bmatrix}$$

e resolvemos (escalonando, por exemplo, com o auxílio de um computador)

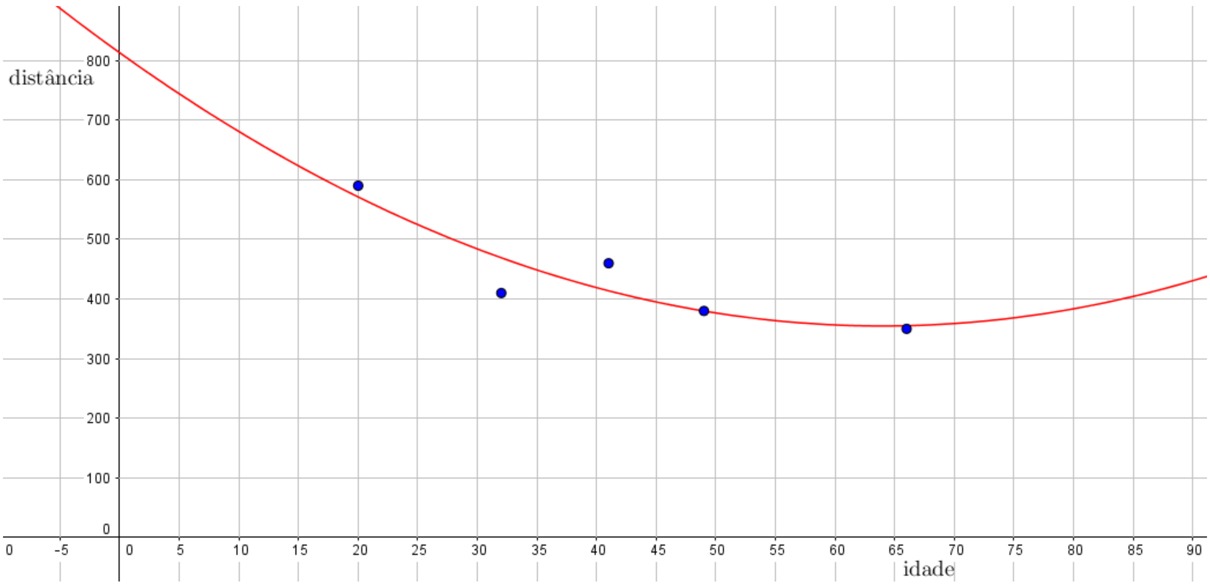
$$\begin{bmatrix} 5 & 208 & 9862 & 2190 \\ 208 & 9862 & 514834 & 85500 \\ 9862 & 514834 & 28773874 & 3866080 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 28482529840/35036713 \\ 0 & 1 & 0 & -1505841055/105110139 \\ 0 & 0 & 1 & 11779375/105110139 \end{bmatrix}.$$

Logo, aproximando estas frações, obtemos

$$\begin{cases} a \simeq 812.934 \\ b \simeq -14.326 \\ c \simeq 0.112 \end{cases}$$

e a parábola desejada é, aproximadamente

$$y = 812.934 - 14.326x + 0.112x^2.$$



Observamos que, apesar de o erro quadrático ser provavelmente menor, esta parábola se torna crescente a partir de um certo momento, fazendo com que o modelo não seja tão razoável para idades maiores. Por exemplo, é pouco provável que (na média), pessoas com 95 anos vejam a uma distância maior do que pessoas de 65 anos, como sugere a parábola acima.

6.5 Regressão linear múltipla

*Uma **regressão linear múltipla** é o problema análogo à regressão linear simples no caso em que a variável dependente pode depender de mais fatores independentes. Tipicamente, queremos encontrar uma equação afim que melhor se ajusta a alguns dados conhecidos.*

No caso especial em que y depende de apenas outros dois fatores, escreveremos

$$z = a + bx + cy,$$

que, geometricamente, representa um plano no espaço tridimensional $\mathbb{R}^3$. Seja agora uma conjunto de dados:

$$(x_1, y_1, z_1),\ (x_1, y_1, z_1), \ldots, (x_k, y_k, z_k).$$

Queremos encontrar coeficientes (a, b, c) que satisfaçam:

$$\begin{cases} a + bx_1 + cy_1 = z_1 \\ a + bx_2 + cy_2 = z_2 \\ \quad \vdots \\ a + bx_k + cy_k = z_k \end{cases} \quad \leftrightsquigarrow \quad \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \vdots & \vdots & \vdots \\ 1 & x_k & y_k \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_k \end{bmatrix}.$$

Quanto maior o número de dados, mais provável de o sistema ser impossível (podemos fazer a analogia geométrica de que por três pontos não colineares no espaço passa um único plano; se aumentarmos o número de pontos, mais difícil que haja um plano contendo todos eles). Por isto, procuramos por uma solução de mínimos quadrados.

Exemplo 38. Uma pesquisa com 214 mulheres na em uma universidade americana[2] coletou informações sobre a altura das participantes, assim como a altura de seus pais. Abaixo, listamos *apenas alguns destes dados*, para que nossas contas não fiquem tão extensas. Fizemos também uma mudança de unidades nas alturas (de polegadas) para centímetros,

Altura	Altura Mãe	Altura Pai
152	155	165
162	155	160
165	170	173
170	163	183
173	168	183
183	165	183

Queremos encontrar uma solução de mínimos quadrados para o sistema linear

$$\begin{bmatrix} 1 & 155 & 165 \\ 1 & 155 & 160 \\ 1 & 170 & 173 \\ 1 & 163 & 183 \\ 1 & 168 & 183 \\ 1 & 165 & 183 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 152 \\ 162 \\ 165 \\ 170 \\ 173 \\ 183 \end{bmatrix} \quad \leftrightsquigarrow \quad A\vec{x} = \vec{b}.$$

Calculamos

$$A^TA = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 155 & 155 & 170 & 163 & 168 & 165 \\ 165 & 160 & 173 & 183 & 183 & 183 \end{bmatrix} \begin{bmatrix} 1 & 155 & 165 \\ 1 & 155 & 160 \\ 1 & 170 & 173 \\ 1 & 163 & 183 \\ 1 & 168 & 183 \\ 1 & 165 & 183 \end{bmatrix} = \begin{bmatrix} 6 & 976 & 1047 \\ 976 & 158968 & 170553 \\ 1047 & 170553 & 183221 \end{bmatrix}$$

e

$$A^T\vec{b} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 155 & 155 & 170 & 163 & 168 & 165 \\ 165 & 160 & 173 & 183 & 183 & 183 \end{bmatrix} \begin{bmatrix} 152 \\ 162 \\ 165 \\ 170 \\ 173 \\ 183 \end{bmatrix} = \begin{bmatrix} 1005 \\ 163689 \\ 175803 \end{bmatrix}.$$

[2] Referência: `https://onlinecourses.science.psu.edu/stat501/node/292`

Por escalonamento,

$$\begin{bmatrix} 6 & 976 & 1047 & 1005 \\ 976 & 158968 & 170553 & 163689 \\ 1047 & 170553 & 183221 & 175803 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 2154573/145769 \\ 0 & 1 & 0 & 14475/145769 \\ 0 & 0 & 1 & 114081/145769 \end{bmatrix} \implies \begin{cases} a \simeq 14.781 \\ b \simeq 0.099 \\ c \simeq 0.783 \end{cases}$$

A equação de melhor ajuste procurada é, portanto, aproximadamente,

$$z \simeq 14.781 + 0.099x + 0.783y.$$

Tente calcular sua altura z a partir da altura de sua mãe x e de seu pai y. O teste deve funcionar melhor para mulheres!