LUCAS

NOTAS DE ÁLGEBRA LINEAR

IMPA TECH - 2024

Sumário

Parte I

Teoria

Introdução

Nesse capítulo introduzimos as noções de vetor e de matriz, discutimos um pouco as relações entre esses objetos. Introduzimos sistemas lineares, espaço linha e espaço coluna. Finalizamos com um problema interessante. Como esse é um capítulo introdutório e um pouco motivacional nem tudo precisa ser plenamente entendido. Algumas coisas são ditas aqui apenas para motivar sua curiosidade.

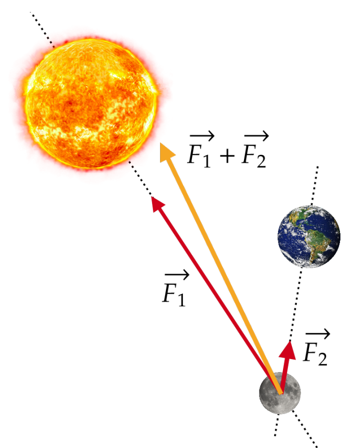

Figura 1: Forças agindo sobre a Lua num sistema com Terra, Sol e Lua.

Vetores

É provável que um físico, uma cientista da computação e um matemático discordem sobre o que é um *vetor*.

Na física, em especial na mecânica clássica, vetores são usados para caracterizar grandezas como força, velocidade, aceleração e distância. Essas grandezas dependem não apenas de um valor numérico associado, mas também de sentido e direção. Como exemplo, considere o sistema com Terra, Sol e Lua representado (em duas dimensões e sem escala) na Figura 1. A força $\overrightarrow{F_1}$ entre a Lua e o Sol é maior que a força $\overrightarrow{F_2}$ entre Lua e Terra (e por isso a flecha tem maior comprimento), além disso, a orientação também importa. Também é comum operar somando as várias forças que agem sobre um mesmo corpo.

Na computação vetores são simples listas ordenadas de números. Por exemplo,

$$v = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 5 \end{bmatrix} \text{ e } w = \begin{bmatrix} 3 \\ 0 \\ -1 \\ 8 \end{bmatrix}.$$

É comum querer somar vetores ou multiplicar vetores por escalares, por exemplo

$$5v - w = \begin{bmatrix} 2 \\ 5 \\ 11 \\ 17 \end{bmatrix}.$$

Além disso, ao contrário dos vetores que encontramos na física, que normalmente tem 3 ou 4 coordenadas, é comum encontrar, na computação, vetores com milhões e até bilhões de entradas[1].

[1] O que existe em comum entre as noções de "coordenadas"e de "entrada"?

Já na matemática, um vetor é um elemento de um espaço vetorial V. Eu não espero que essa frase faça sentido nesse momento e também não pretendo elaborar sobre a definição formal de espaço vetorial por enquanto. Pelo contrário, vou ser evasivo e contar uma história.

No livro *Os elementos*, Euclides começa com as seguintes definições (a tradução é livre):

(i) Ponto é o que não tem partes, ou o que não tem grandeza alguma.

(ii) Linha é o que tem comprimento sem largura. As extremidades da linha são pontos.

(iii) Linha reta é aquela que está posta egualmente entre as suas extremidades.

(iv) Superfície é o que tem comprimento e largura.

E continua definindo extremidade, superfície plana, ângulo, etc. Algumas páginas depois Euclides enuncia seus famosos postulados:

(i) Pode-se traçar uma única reta ligando quaisquer dois pontos.

(ii) Pode-se continuar (de uma maneira única) qualquer reta finita continuamente em uma reta.

(iii) Pode-se traçar um círculo com qualquer centro e com qualquer raio.

(iv) Todos os ângulos retos são iguais.

(v) Se uma reta, ao cortar outras duas, forma ângulos internos, no mesmo lado, cuja soma é menor do que dois ângulos retos, então estas duas retas encontrar-se-ão no lado onde estão os ângulos cuja soma é menor do que dois ângulos retos.

Figura 2: Circle Limit III, Escher, 1959. Uma representação artística do modelo do disco de Poincaré.

O quinto postulado de Euclides, conhecido como o *postulado das paralelas*, foi motivo de polêmica durante toda história da geometria, atravessando dois mil anos: alguns achavam que seria possível provar o postulado partindo dos outros quatro, outros que ele era independente. Essa longa discussão motivou a descoberta de geometrias alternativas onde o quinto postulado é substituído por outros. Dessa forma, até mesmo na matemática moderna, os postulados de Euclides (e suas alternativas) ainda sobrevivem.

Por outro lado, as definições feitas por Euclides não fazem parte da formulação moderna da geometria: são recursivas, confusas e até vagas. O que seria uma parte de um ponto? O que seria ter grandeza? O que é uma extremidade? O que é ter comprimento e largura? E comprimento, mas não largura?

Para evitar definições recursivas ou vagas, e também para construir uma teoria tão generalista e abstrata quanto possível, *no lugar de explicar o que é um vetor vamos dizer o que esperamos de um vetor*. Levando em consideração as noções de vetor da física e da computação parece natural dizer que vetores são objetos que podemos somar e multiplicar por escalares. Ou seja, se v, w são vetores e α, β são números reais, gostaríamos que

$$\alpha v + \beta w$$

também seja um vetor. De fato, vamos exigir mais algumas propriedades, como comutatividade da soma, existência de um elemento neutro para a soma, etc. Veremos isso mais detalhadamente nas próximas páginas. Veremos também que para uma matemática, não só os vetores da física e da computação são exemplos de vetores, mas também é possível enxergar funções ou mesmo sequências de Fibonacci como vetores.

Matrizes e sistemas lineares

Claramente, matrizes são arranjos retangulares de números, por exemplo

$$A = \begin{bmatrix} 1 & 7 & 11 & 40 \\ 6 & -1 & 8 & \frac{4}{5} \\ 0 & 0 & 1 & 7 \end{bmatrix}.$$

Mais do que isso, matrizes estão normalmente associadas à sistemas lineares. Por exemplo, podemos representar o sistema de equações lineares

$$\begin{cases} x + 7y + 11z + 40w = 1 \\ 6x - y + 8z + \frac{4}{5}w = 8 \\ z + 7w = 0 \end{cases}$$

como

$$\begin{bmatrix} 1 & 7 & 11 & 40 \\ 6 & -1 & 8 & \frac{4}{5} \\ 0 & 0 & 1 & 7 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} 40 \\ 8 \\ 0 \end{bmatrix}.$$

Muitas coisas aconteceram nessa última passagem. Primeiro recordamos que para multiplicar uma matriz de l linhas e c colunas por um vetor é necessário que o vetor tenha c entradas. Depois que a multiplicação é feita pegando cada linha da matriz, pareando com o vetor, multiplicando coordenada a coordenada e somando. Por fim, usamos também que dois vetores são iguais quando suas entradas são iguais.

É muito comum representar um sistema linear de maneira compacta simplesmente escrevendo

$$A\overrightarrow{x} = \overrightarrow{b} \tag{1}$$

onde A é como antes,

Tome cuidado para não confundir $\overrightarrow{x}$ com x.

$$\overrightarrow{x} = \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \text{ e } \overrightarrow{b} = \begin{bmatrix} 40 \\ 8 \\ 0 \end{bmatrix}.$$

Olhando com atenção para a equação (1), em especial para o lado esquerdo da igualdade, percebemos que matrizes são, na verdade, *maneiras de mapear vetores em outros vetores*. Basta substituir valores diferentes de $\overrightarrow{x}$ e perceber que acharemos diferentes valores para $A\overrightarrow{x}$. Como veremos a seguir, matrizes não apenas mapeiam vetores em outros vetores, mas fazem isso de maneira linear. Ou seja, dados dois vetores $\overrightarrow{x}, \overrightarrow{y}$ e um escalar α, vale que

$$A\left(\alpha\overrightarrow{x} + \overrightarrow{y}\right) = \alpha A\overrightarrow{x} + A\overrightarrow{y}.$$

Num segundo momento, veremos que matrizes são formas de representar *transformações lineares*, que são simplesmente funções capazes de mapear vetores em outros vetores preservando as propriedades de soma e multiplicação por escalar. Veremos, por exemplo, que a derivada é um exemplo de transformação linear.

Interpretação geométrica de sistemas lineares

Tome o seguinte sistema linear

$$\begin{cases} x + y = 4 \\ 2x - y = -1 \end{cases},$$

cuja solução é $x = 1$ e $y = 3$.

Nesse caso, é fácil desenhar a reta definida por cada uma das equações no plano, como feito na Figura 3. A solução do sistema é dada pelo ponto de encontro das retas. Em especial, se as retas fossem paralelas não haveria solução. Por outro lado, se as retas estivessem sobrepostas haveria infinitas soluções. Um sistema com três variáveis poderia ser representado num espaço tridimensional[2] e

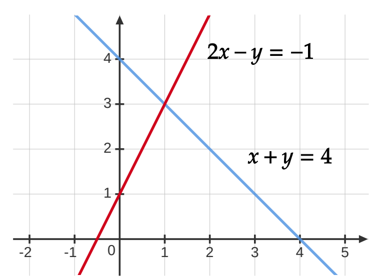

Figura 3: Representação geométrica das linhas de um sistema linear.

[2] O que é dimensão?

cada equação definiria um plano[3].

Uma figura mais interessante aparece quando olhamos pras colunas da matriz que representa esse sistema linear. Começamos escrevendo

$$\begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$$

e observamos que

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

Ou seja, ao escolher valores para x e y estamos combinando as colunas da matriz $\begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}$. Resolver o sistema é achar os valores de x e y cuja combinação é $\begin{bmatrix} 4 \\ -1 \end{bmatrix}$, como ilustrado na Figura 4.

Nos próximos capítulos veremos que a ideia "combinar vetores" ou mais formalmente *combinações lineares* de vetores tem um papel crucial na teoria que vamos desenvolver.

[3] E um sistema com quatro ou mais variáveis?

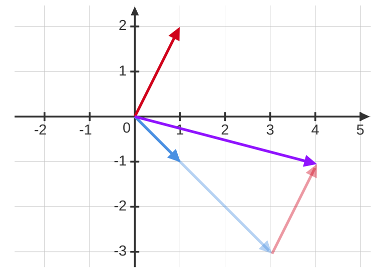

Figura 4: Representação geométrica das colunas de um sistema linear.

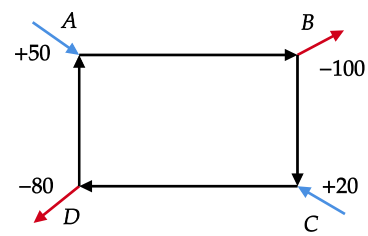

Figura 5: Possível configuração para o tráfego.

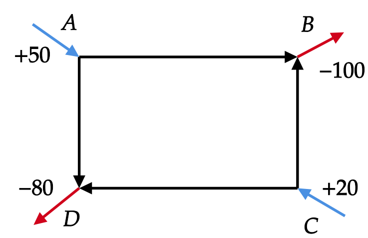

Figura 6: Outra possível configuração para o tráfego.

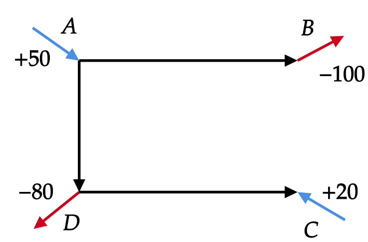

Figura 7: E mais uma possível configuração para o tráfego.

Dois problemas

Embora o leitor ainda não tenha as ferramentas necessárias para resolver os problemas que seguem, vale a pena pensar um pouco sobre eles e entender as dificuldades envolvidas.

Problema um: tráfego urbano

Sejam A, B, C, D junções de ruas como na Figura 5. O fluxo total de carros passando por A em uma hora é de 50 carros em sentido positivo (ou seja, entram mais carros para dentro do sistema do que saem). O fluxo total por B é de 100 carros em sentido negativo, por C de 20 carros no sentido positivo e por D de 80 carros no sentido negativo. Quantos carros passam pela estrada entre A e B por hora? E se consideramos o arranjo da Figura 6? E o da Figura 7?

Problema dois: vírus

Um vírus surgiu na cidade do Rio de Janeiro e se espalha pelo ar. A cada hora existe uma chance de 1 em 5 de uma pessoa sadia se infectar, também existe uma chance de 1 em 10 de uma pessoa infectada conseguir se curar (mas a cura, infelizmente, não é definitiva). Se a infecção começou com 1% da população carioca infectada, o percentual de infectados na cidade vai se estabilizar eventualmente? E se a infecção começar com 5% da população infectada?

Próximos passos

Mesmo que superficialmente, nossa discussão já passou por sistemas planetários, formalismo matemático, funções, sequências de Fibonacci, sistemas lineares, derivadas, tráfego urbano e vírus. Álgebra linear é uma disciplina onipresente na matemática e em qualquer área quantitativa.

O curso tem dois objetivos principais. Na primeira parte do curso discutiremos como operar com vetores e matrizes, desenvolvendo tecnicas e algoritmos para resolver sistemas lineares e outros problemas. Na segunda parte vamos introduzir um formalismo maior e construir uma teoria sobre vetores e transformações lineares.

Fazendo contas com matrizes e vetores

Esse é um capítulo sobre contas e algumas definições básicas que serão essenciais durante o curso. Embora pareça maçante (e muitas vezes seja mesmo), precisamos nos certificar de que entendemos as definições e operações básicas antes de prosseguirmos. Não subestime esse capítulo: saber fazer uma conta de várias maneiras diferentes ou mesmo escrever de várias maneiras diferentes muitas vezes é o segredo para conseguir demonstrar um teorema ou desenhar um algoritmo mais eficiente.

Neste capítulo, consideraremos $\overrightarrow{v}$ e $\overrightarrow{w}$ vetores com entradas

$$\overrightarrow{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{R}^n \text{ e } \overrightarrow{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \in \mathbb{R}^n,$$

além de A e B matrizes com $m \times n$ e $n \times p$ entradas[4], ou seja,

[4] Ou seja, m linhas e n colunas

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

e

$$B = (b_{ij})_{n \times p} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \ddots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{bmatrix}.$$

Somando e multiplicando

Podemos somar vetores e também multiplicar por escalares realizando as operações coordenada a coordenada. Se $\alpha \in \mathbb{R}$, então

$$\alpha \overrightarrow{v} + \overrightarrow{w} = \begin{bmatrix} \alpha v_1 + w_1 \\ \alpha v_2 + w_2 \\ \vdots \\ \alpha v_n + w_n \end{bmatrix}.$$

Também podemos multiplicar uma matriz por um escalar $\alpha \in \mathbb{R}$ realizando a multiplicação coordenada a coordenada, ou seja

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

Como A tem n colunas e $\overrightarrow{v}$ tem n coordenadas podemos multiplicar $A\overrightarrow{v}$ e obter um vetor de m coordenadas da forma

$$A\overrightarrow{v} = \begin{bmatrix} a_{11}v_1 + a_{12}v_2 + \cdots + a_{1n}v_n \\ a_{21}v_1 + a_{22}v_2 + \cdots + a_{2n}v_n \\ \vdots \\ a_{m1}v_1 + a_{m2}v_2 + \cdots + a_{mn}v_n \end{bmatrix} \in \mathbb{R}^m.$$

A multiplicação de duas matrizes A e B, denotada por AB, é uma operação definida quando o número de colunas de A é igual ao número de linhas de B. Se A é uma matriz $m \times n$ e B é uma matriz $n \times p$, então o produto AB resultará em uma matriz $m \times p$. A entrada (i, j) da matriz resultante AB é obtida multiplicando cada elemento da i-ésima linha de A pelos elementos correspondentes da j-ésima coluna de B e somando esses produtos. Matematicamente, a entrada (i, j) de $C = AB$ é dada por [5]:

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj}.$$

[5] Talvez seja preciso gastar um tempinho entendendo o somatório abaixo (somatório é o nome desse Σ grande com índice k indo de $k = 1$ até $k = n$).

É interessante notar que o produto de $A\overrightarrow{v}$ pode ser visto como um produto de matrizes. Basta interpretar $\overrightarrow{v}$ como uma matriz de n linhas e 1 coluna.

Por fim, note que **o produto de matrizes não é comutativo!**[6] Ou seja, pode ocorrer que $AB \neq BA$ mesmo quando $n = m = p$ e o produto está bem definido. Por outro lado, **o produto de matriz é associativo**, ou seja, se A, B e C são matrizes e a multiplicação ABC faz sentido, então $ABC = (AB)C = A(BC)$.

[6] Tome, por exemplo, $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ e $B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$

A transposta e a inversa de uma matriz

A transposta de uma matriz A, denotada por A^T, é obtida trocando suas linhas por colunas e vice-versa. Se A é uma matriz $m \times n$, então a matriz transposta A^T terá dimensões $n \times m$. Logo:

$$A^T = (a_{ji})_{n\times m} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}.$$

Note que enquanto o produto por A mapeia vetores de $\mathbb{R}^n$ para $\mathbb{R}^m$, o produto por A^T faz o contrário: mapeia vetores de $\mathbb{R}^m$ para $\mathbb{R}^n$.

Como o produto entre uma matriz A e um vetor $\overrightarrow{v} \in \mathbb{R}^n$ mapeia $\overrightarrow{v}$ em um vetor $A\overrightarrow{v} \in \mathbb{R}$, é natural pensar na matriz que realiza a operação inversa. Dizemos que uma matriz C ($n \times m$):

- é inversa pela direita de A se $AC = I_m$;
- é inversa pela esquerda de A se $CA = I_n$;
- é inversa de A se é inversa pela esquerda e pela direita.

Acima usamos que I_n é a matriz identidade

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}_{n\times n}.$$

Nem todas as matrizes tem inversa, nem inversa pela direita ou pela esquerda. Quando A tem inversa denotamos a inversa de A por

$$A^{-1}.$$

De fato, saber se uma matriz tem ou não inversa é um problema muito próximo do problema de resolver sistemas lineares[7]. Portanto, é o propósito de boa parte do nosso curso.

[7] Por que?

Algumas propriedades

Seja B inversa pela esquerda de A e C inversa pela direita de A, então

$$B = BI_m = B(AC) = (BA)C = I_nC = C.$$

Logo, se A tem inversa pela direita e pela esquerda as inversas são iguais. Ou seja, a inversa A^{-1} de uma matriz invertível é única.

Além disso, como veremos no futuro, apenas matrizes quadradas (com $n = m$) podem possuir inversa. Por enquanto, vamos fingir que não sabemos disso.

As operações que discutimos satisfazem as seguintes propriedades:

- $(AB)^T = B^T A^T$;
- Quando A e B possuem inversa, então AB é invertível e $(AB)^{-1} = B^{-1}A^{-1}$;
- Quando a inversa de A existe, então a inversa da inversa de A é A, em outras palavras, $(A^{-1})^{-1} = A$;
- Se $\alpha \neq 0$ e A tem inversa, então $(\alpha A)^{-1} = \frac{1}{\alpha}A^{-1}$;
- Se A é invertível, inversa e transposta comutam, ou seja, $(A^T)^{-1} = (A^{-1})^T$;

Vá em frente, tente provar essas propriedades. Aqui vamos provar apenas a última propriedade. Basta mostrar que

$$A^T(A^{-1})^T = I_n \text{ e } (A^{-1})^T A^T = I_m.$$

Podemos usar a primeira propriedade da lista, mostrando que

$$A^T(A^{-1})^T = (A^{-1}A)^T = I_n^T = I_n$$

e

$$(A^{-1})^T A^T = (AA^{-1})^T = I_m^T = I_m.$$

A norma Euclideana e o produto escalar

Uma propriedade muito importante na definição física de vetor é o seu módulo ou tamanho, que aqui conheceremos como norma Euclideana. A norma Euclideana de um vetor $\overrightarrow{v}$ é denotada por $\|\overrightarrow{v}\|_2$ e é definida como[8]

$$\|\overrightarrow{v}\|_2 = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

[8] Note que a definição se parece muito com o Teorema de Pitágoras. Isso não é uma coincidência.

Definimos também o produto interno entre dois vetores $\overrightarrow{v}, \overrightarrow{w} \in \mathbb{R}^n$ como

$$\langle \overrightarrow{v}, \overrightarrow{w} \rangle = v_1w_1 + v_2w_2 + \cdots + v_nw_n = \sum_{i=1}^{n} v_iw_i.$$

O produto interno tem uma interpretação natural em termos da lei dos cossenos. Quaisquer dois vetores $\overrightarrow{v} \in \mathbb{R}^n$ e $\overrightarrow{v} \in \mathbb{R}^n$ quando posicionados numa mesma origem definem um plano (e isso independe de n!). Nesse plano esses vetores definem um ângulo θ.

Podemos formar um triângulo com lados de tamanhos $\|\vec{v}\|_2$, $\|\vec{w}\|_2$ e $\|\vec{w} - \vec{v}\|_2$. Além disso, o ângulo θ entre $\vec{v}$ e $\vec{w}$ aparece no triângulo.

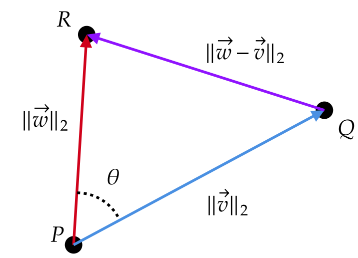

Figura 8: Dados vetores $\vec{v}$ e $\vec{v}$ podemos formar um triângulo relacionando as normas Euclideanas e o ângulo θ entre os vetores.

A lei dos cossenos diz que um triângulo de vértices PQR satisfaz

$$|QR|^2 = |PQ|^2 + |PR|^2 - 2|PQ||PR|\cos\left(\widehat{RPQ}\right).$$

Aplicando a lei dos cossenos no triangulo da Figura 8 temos

$$\|\vec{w} - \vec{v}\|_2^2 = \|\vec{v}\|_2^2 + \|\vec{w}\|_2^2 - 2\|\vec{v}\|_2\|\vec{w}\|_2\cos\theta,$$

além disso, vale que (faça as contas)

$$\|\vec{w} - \vec{v}\|_2^2 = \|\vec{v}\|_2^2 + \|\vec{w}\|_2^2 - 2\langle\vec{v}, \vec{w}\rangle.$$

Logo,

$$\langle\vec{v}, \vec{w}\rangle = \|\vec{v}\|_2\|\vec{w}\|_2\cos\theta. \qquad (2)$$

A identidade (2) fornece uma interpretação geométrica muito útil para aplicações. Assuma que $\|\vec{v}\|_2$ e $\|\vec{v}\|_2$ são fixas, mas que podemos escolher $\vec{v}$ e $\vec{w}$ de maneira a variar o ângulo θ: quando o produto escalar é alto então $\vec{v}$ e $\vec{w}$ estão próximos; quando o produto escalar é próximo de zero então $\vec{v}$ e $\vec{w}$ são quase perpendiculares; quando o produto escalar é negativo e com valor absoluto alto então $\vec{v}$ e $\vec{w}$ apontam em sentidos opostos. Ou seja, **o produto escalar mede a similaridade entre dois vetores**.[9]

[9] Uma curiosidade: a regressão linear, um dos métodos clássicos de aprendizado de máquina, usa o produto escalar como medida de similaridade.

O produto escalar e a norma Euclideana estão relacionados por

$$\|\vec{v}\|_2 = \sqrt{\langle\vec{v}, \vec{v}\rangle}$$

e

$$\langle\vec{v}, \vec{w}\rangle = \frac{\|\vec{v} + \vec{w}\|_2^2 - \|\vec{v}\|_2^2 - \|\vec{w}\|_2^2}{2}.$$

Além disso, valem as seguintes propriedades (prove):

- O produto escalar é simétrico, ou seja, $\langle\vec{v}, \vec{w}\rangle = \langle\vec{w}, \vec{v}\rangle$;
- Além disso, para $\vec{v}, \vec{w}, \vec{u} \in \mathbb{R}^n$ e $\alpha \in \mathbb{R}$,

$$\langle\alpha\vec{v} + \vec{w}, \vec{u}\rangle = \alpha\langle\vec{v}, \vec{u}\rangle + \langle\vec{w}, \vec{u}\rangle.$$

Vale observar também que as propriedades vistas aqui valem em maior generalidade com as definições certas. Veremos no futuro que a norma Euclideana é apenas um exemplo de norma e que o produto escalar entre vetores é apenas um exemplo de produto interno. Além disso, as propriedades vistas vão continuar valendo em maior generalidade.

O traço de uma matriz

Definimos também mais uma operação relevante, o traço, que será definido apenas para matrizes quadradas. Definimos

$$\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii}.$$

Além disso, o traço satisfaz as seguintes propriedades:

- O traço é linear. Seja α um escalar e A e B matrizes $n \times n$, então:

$$\mathrm{tr}(\alpha A + B) = \alpha\,\mathrm{tr}(A) + \mathrm{tr}(B).$$

- O traço do produto não depende da ordem:

$$\mathrm{tr}(AB) = \mathrm{tr}(BA).$$

Exercícios

Exercício 1. *Mostre que se existe um vetor $\overrightarrow{v} \neq \overrightarrow{0}$ tal que $A\overrightarrow{v} = \overrightarrow{0}$ então A não possui inversa. Nesse caso, é possível que A tenha inversa pela direita? E pela esquerda?*

Exercício 2. *Sejam $\overrightarrow{v}, \overrightarrow{w} \in \mathbb{R}^n$ vetores e V, W as matrizes $nx1$ correspondentes aos vetores. Mostre que*

(i) $\langle \overrightarrow{v}, \overrightarrow{w} \rangle = V^T W = \mathrm{tr}(VW^T)$.

(ii) Se $A = (a_{ij})_{n\times n}$, então

$$\langle \overrightarrow{v}, A\overrightarrow{w} \rangle = V^T A W = \sum_{i,j\in[n]} a_{ij} v_i w_j = W^T A^T V = \langle A^T\overrightarrow{v}, \overrightarrow{w} \rangle.$$

Exercício 3. *Uma das mais conhecidas arquiteturas para redes neurais é o multilayer perceptron, cuja arquitetura baseia-se em compor multiplicações por matrizes com ativações não-lineares. Tome A uma matriz $m \times n$ e B uma matriz $p \times m$. Defina $\sigma : \mathbb{R}^m \mapsto \mathbb{R}^m$ como sendo a aplicação $x \mapsto \max(0, x)$ coordenada a coordenada, ou seja, σ zera as entradas negativas e preserva as entradas positivas. Chamamos σ de função de ativação[10]. Uma rede $\varphi : \mathbb{R}^n \mapsto \mathbb{R}^p$ com duas camadas lineares é dada pelo mapa*

$$\varphi\left(\overrightarrow{v}\right) = B\sigma\left(A\overrightarrow{v}\right),$$

onde as matrizes A e B normalmente são aprendidas treinando a rede neural. Escolha m, p, A e B de maneira que φ seja a função identidade em $\mathbb{R}^n$. Justifique sua escolha.

[10] Essa escolha de σ em específico é conhecida como ReLU.

Eliminação Gaussiana

Nosso objetivo nesse capítulo é achar uma maneira sistemática de resolver sistemas lineares.

O que podemos esperar de sistemas lineares?

Antes de tentar resolver sistemas lineares vamos olhar alguns exemplos com diferentes comportamentos:

1. O sistema
$$\begin{cases} 2x + y = 1 \\ 2x = 2 \end{cases}$$
tem solução $x = 1$, $y = -1$. Essa solução é claramente única.

2. O sistema
$$\begin{cases} -2x + y = 1 \\ -4x + 2y = 2 \end{cases}$$
tem infinitas soluções, basta fixar um valor de x e tomar $y = 1 + 2x$.

3. Por fim, o sistema
$$\begin{cases} x + 3y = 1 \\ 2x + 6y = 3 \end{cases}$$
não tem nenhuma solução.

O Algoritmo

Existem várias estratégias para solucionar sistemas lineares, algumas se aplicam mais facilmente a sistemas com poucas equações e poucas variáveis, outras são mais gerais. Estudaremos nesse capítulo o processo de eliminação Gaussiana.

O objetivo da eliminação Gaussiana é transformar um sistema linear em um novo sistema equivalente (com as mesmas soluções) que pode ser resolvido trivialmente com substituição retroativa. Tome, por exemplo,

$$\begin{bmatrix} 1 & 2 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \\ 1 \end{bmatrix}.$$

Resolvendo primeiro z, depois y e depois x é fácil ver que $z = 1$, $y = 3$ e $x = -4$. Note cada variável que foi resolvida requeria apenas o valor de variáveis já conhecidas! Isso se dá pelo formato triangular da matriz. Dizemos que uma matriz é triangular superior quando ela tem apenas zeros abaixo da diagonal e inferior quando ela tem apenas zeros acima da diagonal. **O objetivo da eliminação Gaussiana é transformar uma matriz qualquer em uma matriz equivalente que seja triangular superior**[11], é comum também exigir que essa triangular superior tenha apenas valores zero ou um na diagonal.

[11] De maneira geral, dizemos que duas matrizes A e B são equivalentes se existem matrizes inversíveis P, Q tais que $QB = AP$.

Para realizar a eliminação Gaussiana podemos realizar três operações ditas elementares [12]:

[12] Observe que a primeira operação pode ser obtida a partir da aplicação das outras duas, mas por motivos que se tornarão claros posteriormente vamos manter a primeira operação.

1. Trocar duas linhas de lugar (i.e., permutar duas linhas). Vamos denotar a troca da i-ésima linha com a j-ésima linha por $L_i \leftrightarrow L_j$.

2. Multiplicar uma linha por um número $\alpha \neq 0$. Denotaremos $L_i \leftarrow \alpha L_i$.

3. Substituir uma linha pelo resultado da soma dessa linha com um múltiplo de qualquer outra linha. Denotaremos $L_i \leftarrow L_i + \alpha L_j$, onde $\alpha \in \mathbb{R}$.

Vamos entender o processo com o seguinte exemplo:

$$\begin{bmatrix} 2 & 1 & -1 \\ -3 & -1 & 2 \\ -3 & 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 8 \\ -11 \\ -3 \end{bmatrix}.$$

Para simplificar, é usual representar o sistema na forma de uma *matriz aumentada* como abaixo:

$$\left[\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ -3 & -1 & 2 & -11 \\ -2 & 1 & 2 & -3 \end{array}\right]$$

Vamos realizar as operações elementares para transformar essa matriz em uma forma triangular superior:

Passo 1: Dividir a primeira linha por 2 para fazer o elemento a_{11} (primeiro elemento da primeira linha) igual a 1, ou seja, aplicar $L_1 \leftarrow \frac{1}{2}L_1$:

Note que a operação $L_1 \leftarrow \frac{1}{2}L_1$ não passa de multiplicar pela esquerda por $E_{L_1 \leftarrow \frac{1}{2}L_1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

$$\left[\begin{array}{ccc|c} 1 & \frac{1}{2} & -\frac{1}{2} & 4 \\ -3 & -1 & 2 & -11 \\ -2 & 1 & 2 & -3 \end{array}\right]$$

Passo 2: Multiplicar a primeira linha por 3 e adicionar à segunda linha para zerar o elemento a_{21} (primeiro elemento da segunda linha), ou seja, aplicar $L_2 \leftarrow 3L_1 + L_2$:

Note que a operação $L_2 \leftarrow 3L_1 + L_2$ não passa de multiplicar pela esquerda por $E_{L_2 \leftarrow 3L_1 + L_2} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

$$\left[\begin{array}{ccc|c} 1 & \frac{1}{2} & -\frac{1}{2} & 4 \\ 0 & \frac{1}{2} & \frac{1}{2} & 1 \\ -2 & 1 & 2 & -3 \end{array}\right]$$

Passo 3: Multiplicar a primeira linha por 2 e adicionar à terceira linha para zerar o elemento a_{31} (primeiro elemento da terceira linha), ou seja, aplicar $L_3 \leftarrow 2L_1 + L_3$:

Note que a operação $L_3 \leftarrow 2L_1 + L_3$ não passa de multiplicar pela esquerda por $E_{L_3 \leftarrow 2L_1 + L_3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}$.

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

Passo 4: Multplicar a segunda linha por 2 para fazer o elemento a_{22} (segundo elemento da segunda linha) igual a 1, ou seja, aplicar $L_2 \leftarrow 2L_2$:

Quem é $E_{L_2 \leftarrow 2L_2}$?

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

Passo 5: Multiplicar a segunda linha por 2 e subtrair da terceira linha para zerar o elemento a_{32} (segundo elemento da terceira linha), ou seja, aplicar $L_3 \leftarrow 2L_2 + L_3$:

Quem é $E_{L_3 \leftarrow 2L_2 + L_3}$?

$$\left[\begin{array}{ccc|c} 1 & \frac{1}{2} & -\frac{1}{2} & 4 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 1 \end{array}\right]$$

Como o terceiro elemento da diagonal já é 1 podemos parar. Note que já é possível resolver o sistema linear por substituição retroativa [13]. Note que, nesse caso, não precisamos fazer uso da operação de trocar duas linhas de lugar. Tente aplicar o processo de eliminação Gaussiana nos sistemas

[13] Qual é a solução?

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 5 \\ 4 & 6 & 8 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

e

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 5 \\ 4 & 4 & 8 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

Chamamos as entradas diagonais unitárias de **pivôs**. Uma matriz $n \times n$ que pode ser transformada em uma matriz triangular superior com n pivôs não nulos é dita matriz **não singular**, caso contrário dizemos que a matriz é singular. Além disso, vamos chamar o número de pivôs não nulos da matriz triangular superior obtida ao final do processo de eliminação Gaussiana de uma matriz A de **posto** (ou *rank*, em inglês) de A.

Matrizes elementares, permutações e a decomposição LU

As matrizes correspondentes as operações elementares são ditas matrizes elementares. A forma geral dessas matrizes é:

$$E_{L_i \leftrightarrow L_j} = \begin{array}{cc} \begin{array}{ccccccc} & & \text{coluna } i & & \text{coluna } j & & \end{array} & \\ \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & \dots & 1 & & \\ & & \vdots & \ddots & \vdots & & \\ & & 1 & \dots & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix} & \begin{array}{l} \\ \\ \text{linha } i \\ \\ \text{linha } j \\ \\ \\ \end{array} \end{array}$$

$$E_{L_i \leftarrow \alpha L_i} = \begin{array}{cc} \begin{array}{ccccc} & & \text{coluna } i & & \end{array} & \\ \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & \alpha & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix} & \begin{array}{l} \\ \\ \text{linha } j \\ \\ \\ \end{array} \end{array}$$

$$E_{L_i \leftarrow \alpha L_j + L_i} = \begin{array}{cc} \begin{array}{ccccccc} & & \text{coluna } i & & \text{coluna } j & & \end{array} & \\ \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & \dots & \alpha & & \\ & & \vdots & \ddots & \vdots & & \\ & & 0 & \dots & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix} & \begin{array}{l} \\ \\ \text{linha } i \\ \\ \text{linha } j \\ \\ \\ \end{array} \end{array}$$

Todas as matrizes acima possuem inversa[14] e, portanto, o produto delas também. Note também que o escalonamento pode ser feito de

[14] Prove!

maneira que as matrizes elementares das regras 2 e 3 são escolhidas para serem sempre triangulares inferiores (basta seguir metodicamente escalonando em ordem a primeira linha, a segunda linha, a terceira linha...).

Logo, escalonar uma matriz A é encontrar uma matriz E (produto das matrizes elementares usadas em cada etapa) tal que

$$EA$$

é uma matriz triangular superior. Note que se E for obtida usando apenas as regras 2 e 3 então E é triangular inferior[15].

[15] Prove que o produto de matrizes triangulares inferiores é triangular inferior

De fato, a menos de uma permutação inicial P (obtida repetindo a regra 1 conforme necessário) podemos sempre assumir que E foi obtida usando apenas as regras 2 e 3. Nesse caso,

$$E(PA)$$

é uma matriz triangular superior, que vamos denotar por U, logo:

$$E(PA) = U.$$

Além disso, vale que E tem inversa e essa inversa é triangular inferior[16], vamos denotar $L = E^{-1}$. Logo:

[16] Prove que a inversa das matrizes elementares obtidas pelas regras 2 e 3 são matrizes triangulares inferiores.

$$PA = LU.$$

A equação acima é tão importante, que vou repetir ela com palavras: **toda matriz A pode ser escrita, a menos de uma permutação P, como o produto de uma matriz L inversível triangular inferior e uma matriz U triangular superior**. Esse fato é conhecido como **decomposição LU**. Além disso, a decomposição LU pode ser obtida como um subproduto da eliminação Gaussiana.

O seguinte teorema é consequência da discussão que acabamos de fazer:

Teorema 1 (Decomposição LU)**.** *Seja A uma matriz $n \times m$. Existem matrizes P, L e U satisfazendo*

- *P é $n \times n$ e é uma matriz de permutação, ou seja, existe $k \in \mathbb{N}$ e tuplas de índices $\{(i_l, j_l)\}_{l=1}^{k}$ tais que*

$$P = \prod_{l=1}^{k} E_{L_{i_l} \leftrightarrow L_{j_l}}.$$

- *L é $n \times n$, possui inversa e é triangular inferior;*

- *U é $n \times m$ e triangular superior e sua diagonal tem apenas entradas zero ou um.*

E, além disso,

$$PA = LU.$$

Resolvendo sistemas lineares

Vamos agora fazer uso da decomposição $PA = LU$ para resolver

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

Multiplicando por P obtemos

$$(PA)\overrightarrow{x} = P\overrightarrow{b}.$$

E então

$$LU\overrightarrow{x} = P\overrightarrow{b}.$$

Vamos chamar de $\overrightarrow{y}$ o valor (ainda desconhecido) de $U\overrightarrow{x}$. Então vale

$$L\overrightarrow{y} = P\overrightarrow{b} \text{ e } U\overrightarrow{x} = \overrightarrow{y}.$$

Em outras palavras, uma vez que se conhece uma decomposição $PA = LU$ para a matriz A, podemos resolver $A\overrightarrow{x} = \overrightarrow{b}$ resolvendo primeiro $L\overrightarrow{y} = P\overrightarrow{b}$ por substituição e depois $U\overrightarrow{x} = \overrightarrow{y}$ também por substituição!

Note que caso $A\overrightarrow{x} = \overrightarrow{b}$ não tenha solução, então o sistema dado por $L\overrightarrow{y} = P\overrightarrow{b}$ e $U\overrightarrow{x} = \overrightarrow{y}$ também não terá solução. Mas como L possui inversa (por construção, já que é inversa de produto de matrizes elementares), então vale que

$$A\overrightarrow{x} = \overrightarrow{b} \text{ tem solução sse } U\overrightarrow{x} = \overrightarrow{y} \text{ tem solução.}$$

Dois exemplos de decomposição $PA = LU$

Vamos retomar o exemplo sugerido anteriormente, tome:

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 5 \\ 4 & 6 & 8 \end{bmatrix}.$$

Aplicando as operações $L_2 \leftarrow L_2 - 2L_1$, $L_3 \leftarrow L_3 - 4L_1$, $L_2 \leftrightarrow L_3$, $L_2 \leftarrow \frac{1}{2}L_2$ e $L_3 \leftarrow \frac{1}{3}L_3$, o processo de eliminação Gaussiana resulta em:

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 5 \\ 4 & 6 & 8 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 3 \\ 4 & 6 & 8 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 3 \\ 0 & 2 & 4 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 4 \\ 0 & 0 & 3 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 3 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$

Tomando $P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ podemos realizar a eliminação Gaussiana sobre $PA = \begin{bmatrix} 1 & 1 & 1 \\ 4 & 6 & 8 \\ 2 & 2 & 5 \end{bmatrix}$ sem fazer uso da operação de permutação. De fato, aplicando $L_2 \leftarrow L_2 - 4L_1$, $L_3 \leftarrow L_3 - 2L_1$, $L_2 \leftarrow \frac{1}{2}L_2$ e

$L_3 \leftarrow \frac{1}{3}L_3$, obtemos:

$$\begin{bmatrix}1 & 1 & 1\\4 & 6 & 8\\2 & 2 & 5\end{bmatrix} \longrightarrow \begin{bmatrix}1 & 1 & 1\\0 & 2 & 4\\2 & 2 & 5\end{bmatrix} \longrightarrow \begin{bmatrix}1 & 1 & 1\\0 & 2 & 4\\0 & 0 & 3\end{bmatrix} \longrightarrow \begin{bmatrix}1 & 1 & 1\\0 & 1 & 2\\0 & 0 & 3\end{bmatrix} \longrightarrow \begin{bmatrix}1 & 1 & 1\\0 & 1 & 2\\0 & 0 & 1\end{bmatrix}$$

Em outras palavras,

$$PA = \underbrace{\begin{bmatrix}1 & 0 & 0\\4 & 1 & 0\\0 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\2 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\0 & 2 & 0\\0 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 3\end{bmatrix}}_{\mathrm{L}}\underbrace{\begin{bmatrix}1 & 1 & 1\\0 & 1 & 2\\0 & 0 & 1\end{bmatrix}}_{\mathrm{U}}$$

Note que, nesse caso, conseguimos obter uma decomposição $PA = LU$ em que U tem pivôs em todas as linhas. Logo, $A\overrightarrow{x} = \overrightarrow{b}$ tem solução para qualquer escolha de $\overrightarrow{b}$. Isso não será verdade no próximo exemplo.

Agora tome

$$A = \begin{bmatrix}1 & 1 & 1\\2 & 2 & 5\\4 & 4 & 8\end{bmatrix}.$$

Realizando o processo de eliminação Gaussiana obtemos

$$A = \underbrace{\begin{bmatrix}1 & 0 & 0\\2 & 1 & 0\\0 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\4 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\0 & 3 & 0\\0 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 4\end{bmatrix}}_{\mathrm{L}}\underbrace{\begin{bmatrix}1 & 1 & 1\\0 & 0 & 1\\0 & 0 & 1\end{bmatrix}}_{\mathrm{U}}$$

Note que a matriz U obtida tem apenas 2 pivôs, um na primeira e outro na terceira linha. Além disso,

$$U\overrightarrow{x} = \begin{bmatrix}x_1 + x_2 + x_3\\x_3\\x_3\end{bmatrix}.$$

Ou seja, $U\overrightarrow{x} = \overrightarrow{y}$ só pode ter solução se $y_2 = y_3$. Logo, $A\overrightarrow{x} = \overrightarrow{b}$ não tem solução para qualquer escolha de $\overrightarrow{b}$ [17].

[17] Justifique com um exemplo.

Não é coincidência que um posto menor que o número de linhas tenha levado à um sistema $A\overrightarrow{x} = \overrightarrow{b}$ que não tem solução para certas escolhas de $\overrightarrow{b}$, veremos isso com mais detalhes no que segue.

Quando inversas existem? E como calcular?

Podemos usar a eliminação Gaussiana (ou mesmo a decomposição LU) de uma matriz A para descobrir se ela tem inversa e obter essa inversa. Começamos caracterizando um caso especial, o das matrizes quadradas U triangulares superiores cuja diagonal possui apenas entradas unitárias. Ou seja, matrizes da forma

$$U = \begin{bmatrix} 1 & u_{12} & \cdots & u_{1n} \\ 0 & 1 & \cdots & u_{2n} \\ 0 & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Teorema 2. *Toda matriz quadrada U triangular superior cuja diagonal tenha apenas entradas unitárias é inversível.*

Demonstração. Se $B = \begin{bmatrix} \overrightarrow{b_1} & \overrightarrow{b_2} & \cdots & \overrightarrow{b_n} \end{bmatrix}$ é matriz $n \times n$ com entradas coluna $\overrightarrow{b_i}$, $i \in [n]$, então

$$UB = \begin{bmatrix} U\overrightarrow{b_1} & U\overrightarrow{b_2} & \cdots & U\overrightarrow{b_n} \end{bmatrix}.$$

Seja $\overrightarrow{e_i}$ o vetor cuja única entrada não nula é a i-ésima. Logo, B é inversa pela direita de U se, e somente se,

$$U\overrightarrow{b_i} = \overrightarrow{e_i} \;\; \forall i \in [n].$$

Todos os n sistemas acima possuem solução por substituição retroativa, logo U possui inversa pela direita. Analogamente, é possível mostrar que U possui inversa pela esquerda. $\square$

Note que se uma matriz qualquer A possui inversa pela esquerda B e inversa pela direita C, então $B = C$ e a inversa é única. Basta observar que

$$B = B(AC) = (BA)C = C.$$

Agora usamos a decomposição LU para generalizar o teorema anterior para o caso geral:

Teorema 3. *Seja A uma matriz $m \times n$ e seja $PA = LU$ uma decomposição LU de A. As seguintes afirmações valem:*

(i) A tem inversa pela esquerda se, e somente se, a matriz triangular superior U tem inversa pela esquerda.

(ii) A tem inversa pela direita se, e somente se, a matriz triangular superior U tem inversa pela direita.

(iii) Se $n > m$, então U não tem inversa pela esquerda.

(iv) Se $n < m$, então U não tem inversa pela direita.

(v) Se $n \leq m$ e a diagonal de U não tem entradas nulas, então U tem inversa pela esquerda.

(vi) Se $n \geq m$ e a diagonal de U não tem entradas nulas, então U tem inversa pela direita.

Demonstração. Mostraremos apenas os itens ímpares, já que os pares seguem analogamente.

Começamos mostrando (i). Se A tem inversa pela esquerda, então existe B tal que $BA = I_n$. Usando que $A = P^{-1}LU$ temos

$$(BP^{-1}L)U = BP^{-1}LU = I_n.$$

Logo, $BP^{-1}L$ é inversa pela esquerda de U. Agora assuma que B é inversa pela esquerda de U, ou seja, $BU = I_n$ então

$$(BL^{-1}P)A = BU = I_d.$$

Portanto, $BL^{-1}P$ é inversa pela esquerda de A.

Agora mostramos (iii). Uma inversa pela esquerda para U seria uma matriz B de dimensões $n \times m$ tal que $BU = I_n$. Como U é triangular superior então pode ser vista em blocos da forma

$$U = \begin{bmatrix} U' & 0_{m\times(n-m)} \end{bmatrix}_{m\times n}$$

para uma matriz U' triangular superior quadrada de dimensões $m \times m$. Além disso, dado qualquer B temos:

$$BU = \begin{bmatrix} BU' & 0_{n\times(n-m)} \end{bmatrix}_{n\times n}.$$

Logo, se $n > m$, $BU \neq I_n$ para qualquer escolha de B.

Por fim, mostraremos (v). Podemos escrever

$$U = \begin{bmatrix} U' \\ 0_{(m-n)\times n} \end{bmatrix}_{m\times n}$$

para uma matriz U' triangular superior quadrada de dimensões $n \times n$. Além disso, U' não tem entradas nulas na diagonal. Pelo Teorema 2, existe U'^{-1}. Agora observe que

$$\begin{bmatrix} U'^{-1} & 0_{n\times(m-n)} \end{bmatrix}_{n\times m} U = I_n.$$

Logo, U possui inversa pela esquerda. □

Entre muitas coisas, o resultado acima mostra que apenas matrizes quadradas possuem inversa por ambos os lados.

Exercícios

Exercício 4. *Mostre que um sistema linear da forma $A\vec{x} = \vec{b}$ satisfaz exclusivamente um de três casos: tem uma única solução; não tem solução alguma; tem infinitas soluções.*

Exercício 5. *Mostraremos que ao somar duas equações e substituir uma delas pela soma as soluções não mudam. Sejam* $\overrightarrow{v}, \overrightarrow{w} \in \mathbb{R}^n$ *vetores,* $b_1, b_2 \in \mathbb{R}$ *e* $\alpha \in \mathbb{R}$*. Mostre que*

$$\left\{ \overrightarrow{x} \in \mathbb{R}^n : \langle \overrightarrow{x}, \overrightarrow{v} \rangle = b_1 \text{ e } \langle \overrightarrow{x}, \overrightarrow{w} \rangle = b_2 \right\}$$
$$=$$
$$\left\{ \overrightarrow{x} \in \mathbb{R}^n : \langle \overrightarrow{x}, \overrightarrow{v} \rangle = b_1 \text{ e } \langle \overrightarrow{x}, \overrightarrow{w} + \alpha \overrightarrow{v} \rangle = b_2 + \alpha b_1 \right\}.$$

Exercício 6. *Mostre que não existe escolha de valores para* $a, b, c, d, e, f, g, h \in \mathbb{R}$ *tal que*

$$\begin{bmatrix} 0 & a & 0 & 0 & 0 \\ b & 0 & c & 0 & 0 \\ 0 & d & 0 & e & 0 \\ 0 & 0 & f & 0 & g \\ 0 & 0 & 0 & h & 0 \end{bmatrix} \text{ tenha inversa.}$$

Espaços vetoriais

No capítulo anterior entendemos sistemas lineares a partir da visão do espaço de linhas e desenvolvemos ferramentas para obter soluções de sistemas lineares, além de discutir inversão de matrizes. O objetivo desse capítulo será criar ferramentas para que seja possível entender sistemas lineares pelo espaço das colunas. Além disso, vamos finalmente responder o que é um vetor (numa visão mais ampla) e formalizar a noção de dimensão.

Espaço e subespaço vetorial

Até agora falamos sobre vetores como coisas que podemos somar e multiplicar por escalares. Essa ideia é formalizada na definição abaixo:

Definição 1 (Espaço vetorial)**.** *Seja V um conjunto, $\cdot : \mathbb{R} \times V \to V$ e $+ : V \times V \to V$ operações. Dizemos que $(V, \cdot, +)$ (ou simplesmente V) é um espaço vetorial se valem*

1. *(comutatividade) $v + w = w + v$ para quaisquer $v, w \in V$;*
2. *(associatividade da soma) $(v + w) + z = w + (v + z)$ para quaisquer $v, w, z \in V$;*
3. *(elemento neutro da soma) existe $0 \in V$ tal que $v + 0 = v$ para todo $v \in V$;*
4. *(existência da inversa da soma) para todo $v \in V$ existe um elemento $-v \in V$ tal que $v + (-v) = 0$;*
5. *(associatividade do produto por escalar) $\alpha \cdot (\beta v) = (\alpha\beta) \cdot v$ para quaisquer $\alpha, \beta \in \mathbb{R}$ e $v \in V$;*
6. *(distributividade do produto escalar) $(\alpha + \beta) \cdot v = \alpha \cdot v + \beta \cdot v$ para quaisquer $\alpha, \beta \in \mathbb{R}$ e $v \in V$;*
7. *(distributividade da soma) $\alpha \cdot (v + w) = \alpha \cdot v + \alpha \cdot w$ para quaisquer $\alpha \in \mathbb{R}$ e $v, w \in V$;*

8. *(elemento neutro do produto) vale que* $1 \cdot v = v$ *para todo* $v \in V$*.*

Definição 2 (Subespaço vetorial)**.** *Seja* $(V, \cdot, +)$ *um espaço vetorial e* $W \subset V$*, dizemos que* W *é um subespaço vetorial de* V *se* $(W, \cdot, +)$ *é um espaço vetorial.*

Alguns exemplos usuais de espaço e subespaço vetoriais são:

- O conjunto $\mathbb{R}^n$ é um espaço vetorial com a soma e produto definidos usualmente. Além disso, se $v, w \in \mathbb{R}^n$, então

$$\{\alpha v + \beta w : \alpha, \beta \in \mathbb{R}\}$$

  é um subespaço vetorial de $\mathbb{R}^n$[18].

[18] Prove!

- Se S é um conjunto qualquer e V é o conjunto das funções $f : S \to \mathbb{R}$ com soma e produto definidos usualmente, então V é um espaço vetorial. Se $R \subset S$ e W é o conjunto das funções que se anulam em R então W é um espaço vetorial[19].

[19] Prove!

- Se $V = \mathbb{R}[x]$ denota os polinômios em x, então V é espaço vetorial. O conjunto $W = \mathbb{R}_n[x]$ de polinômios com grau menor que n é um subespaço vetorial[20].

[20] Prove! Além disso, argumente se o conjunto dos polinômios de grau exatamente $n - 1$ forma ou não um subespaço de V.

Combinações lineares, bases e dimensão

Seja V um espaço vetorial e $v_1, \dots, v_n \in V$. Se $\alpha_1, \dots, \alpha_n$ são reais quaisquer, dizemos que

$$\alpha_1 v_1 + \dots + \alpha_n v_n$$

é uma combinação linear dos vetores $v_1, \dots, v_n$.

De fato, o conjunto de todas as combinações lineares desses n vetores gera um subespaço de V conhecido como *span* de $v_1, \dots, v_n$[21] e denotado por

[21] Prove!

$$< v_1, \dots, v_n >= \{\alpha_1 v_1 + \dots + \alpha_n v_n : \alpha_i \in \mathbb{R}\, \forall i \in [n]\}.$$

Uma pergunta natural é se algum dos vetores usados é desnecessário para a construção do *span*. Por exemplo, tome

$$v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, v_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \text{ e } v_3 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

É fácil ver que $< v_1, v_2, v_3 >= \mathbb{R}^2$, mas também vale que

$$< v_1, v_2 >=< v_1, v_3 >=< v_2, v_3 >= \mathbb{R},$$

ou seja, qualquer um dos 3 vetores pode ser removido da lista e ainda seremos capazes de gerar $\mathbb{R}^2$!

Note que para que um vetor v_i seja desnecessário é preciso que seja possível escrevê-lo usando os restantes, ou seja, v_i é desnecessário se, e somente se,

$$v_i = \sum_{j \neq i} \alpha_j v_j$$

para alguma escolha de pesos α_i. A seguinte definição captura essa ideia:

Definição 3 (Dependência e independência linear)**.** *Dizemos que $v_1, \ldots, v_n \in V$ são vetores linearmente independentes quando*

$$\sum_{i=1}^{n} \alpha_i v_i = 0$$

só tem solução tomando $\alpha_1 = \cdots = \alpha_n = 0$. Caso contrário dizemos que são linearmente dependentes.

Outro conceito adjacente é o conceito de base, que captura a ideia de quantidade mínima de vetores para gerar um espaço.

Definição 4. *Dizemos que $v_1, \ldots, v_n$ é base de V se é um conjunto linearmente independente e gerar V, ou seja,*

$$< v_1, \ldots, v_n >= V.$$

No exemplo do começo dessa seção vimos que $\mathbb{R}^2$ podia ser gerado sempre com 2 vetores, olhando cuidadosamente percebemos que não é possível gerar com 1 vetor. Esse fato não é uma coincidência, de fato vale que:

Lema 1. *Seja V um espaço vetorial gerado por um número finito de vetores. Então V possui uma base e, além disso, todas as bases de V tem o mesmo número de elementos.*

Demonstração. Como V é gerado por um número finito de vetores então $V =< v_1, \ldots, v_r >$ para algum r. Se $v_1, \ldots, v_r$ for linearmente independente então é base, caso contrário existe um vetor v_i que pode ser escrito como combinação linear dos outros. Removemos esse vetor e iteramos o argumento até que restem apenas vetores linearmente independentes e que sigam gerando V[22].

[22] Formalize isso com uma indução finita.

Agora assuma duas bases com tamanhos distintos, sejam elas $v_1, \ldots, v_n$ e $w_1, \ldots, w_m$. Sem perda de generalidade assumimos $m > n$. Por ser base, existem α_{ij} com $i \in [n]$ e $j \in [m]$ tais que

$$w_j = \alpha_{1j} v_1 + \cdots + \alpha_{nj} v_n.$$

Para quaisquer β_j, $j \in [m]$, podemos reescrever

$$\sum_{j=1}^{m} \beta_j w_j = \sum_{j=1}^{m} \beta_j \left(\sum_{i=1}^{n} \alpha_{ij} v_i \right) = \sum_{i=1}^{n} \left(\sum_{j=1}^{m} \beta_j \alpha_{ij} \right) v_i. \tag{3}$$

Como $v_1, \dots, v_n$ é um conjunto linearmente independente sabemos que (3) se anula se, e somente se,

$$\sum_{j=1}^{m} \beta_j \alpha_{ij} = 0 \, \forall i \in [n].$$

Olhando com cuidado, podemos reescrever as equações acima como um sistema linear:

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{n1} & \alpha_{n2} & \cdots & \alpha_{nm} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix} = 0$$

A matriz $(\alpha_{ij})_{n \times m}$ tem mais colunas do que linhas pois assumimos $m > n$. Logo, o sistema acima possui infinitas soluções[23]. Por outro lado, (3) se anula se, e somente se, $\beta_j = 0$ para todo $j \in [m]$. Um absurdo pois acabamos de observar que existem infinitas soluções.

[23] Justifique. Note também que isso só é verdade pois do lado direito temos o vetor nulo.

□

Se V é gerado por um conjunto finito de vetores então sempre existe uma base e toda base tem o mesmo número de elementos, então esse número independente da base escolhida. **Chamaremos a quantidade de elementos numa base de V de dimensão de V.**

Agora conseguimos entender por que o espaço $\mathbb{R}^d$ tem dimensão d. Seja e_i o vetor em que apenas a i-ésima entrada vale 1 e as outras valem zero. Então $e_1, \dots, e_d$ é uma base[24] de $\mathbb{R}^d$. Para ver isso note que

[24] Essa base é especial e é chamada de base canônica.

- *é conjunto gerador* pois

$$v = \begin{bmatrix} v_1 \\ \vdots \\ v_d \end{bmatrix} = \sum_{i=1}^{d} v_i e_i$$

- e é *conjunto linearmente independente* pois

$$\sum_{i=1}^{d} \alpha_i e_i = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_d \end{bmatrix} = 0$$

ocorre se, e somente se, $\alpha_1 = \cdots = \alpha_d = 0$.

No lema anterior, começamos com um conjunto gerador e dai construímos uma base. O próximo lema mostra que é possível completar um conjunto linearmente independente para obter uma base:

Lema 2. *Seja V um espaço vetorial de dimensão $n < \infty$ e $v_1, \ldots, v_r \in V$ um conjunto linearmente independente. Então existem $v_{r+1}, \ldots, v_n \in V$ tais que $v_1, \ldots, v_n$ é base de V.*

Demonstração. Basta observar que se $v_1, \ldots, v_r$ não gera V então existe $v_{r+1} \in V$ tal que

$$v_{r+1} = \sum_{i=1}^{r} \alpha_i v_i$$

não tem solução α_i, $i \in [r]$. Logo, $v_1, \ldots, v_{r+1}$ é linearmente independente[25]. Agora basta iterar o argumento para obter $v_{r+2}, \ldots, v_n$. Ao chegarmos em n vetores linearmente independentes teremos uma base[26] e o processo termina. □

[25] Prove que se v_r não pode ser escrito como combinação linear de vetores linearmente independentes $v_1, \ldots, v_r$, então $v_1, \ldots, v_{r+1}$ é linearmente independente

[26] Mostre que n vetores linearmente independentes em um espaço vetorial de dimensão n sempre formam uma base.

Coordenadas

O conceito de base também permite formalizar a noção de coordenadas. Note que se $v_1, \ldots, v_n$ é base de V então todo vetor $v \in V$ pode ser escrito como

$$v = \sum_{i=1}^{n} \alpha_i v_i$$

para alguma escolha de constantes $\alpha_i \in \mathbb{R}$. Curiosamente, acontece que a escolha dessas constantes é única, ou seja, para cada vetor $v \in V$ só existe uma maneira de escrevê-lo como soma dos vetores de uma base. Para ver isso assuma que existem duas maneiras, sejam α_i e α_i', $i \in [n]$, tais que

$$v = \sum_{i=1}^{n} \alpha_i v_i = \sum_{i=1}^{n} \alpha_i' v_i,$$

logo

$$0 = v - v = \sum_{i=1}^{n} (\alpha_i - \alpha_i') v_i$$

e usando que bases são linearmente independentes temos $\alpha_i = \alpha_i'$ para todo $i \in [n]$. Como essa representação é única então podemos usar ela pra representar v. Ou seja, no lugar de escrever

$$v = \alpha_1 v_1 + \cdots + \alpha_n v_n$$

podemos escrever

$$v = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}_{v_1, \ldots, v_n} .$$

Note que as coordenadas de um vetor v mudam de acordo com a base escolhida, mesmo que o vetor não mude. Também vale observar que no espaço $\mathbb{R}^n$ as entradas do vetor correspondem com as suas coordenadas na base canônica. Por exemplo,

$$\begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix}_{e_1,e_2,e_3} = \begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix}_{e_1+e_2,e_2+e_3,e_3}.$$

Um espaço vetorial nem tão diferente assim

Seja $\mathbb{R}_+$ o conjunto dos reais positivos. Vale que $(\mathbb{R}_+^n, +, \cdot)$ com

$$\begin{aligned} + : \mathbb{R}_+^n \times \mathbb{R}_+^n &\to \mathbb{R}_+^n \\ ((x_1,\ldots,x_n),(y_1,\ldots,y_n)) &\mapsto (x_1y_1,\ldots,x_ny_n) \end{aligned}$$

e

$$\begin{aligned} \cdot : \mathbb{R} \times \mathbb{R}_+^n &\to \mathbb{R}_+^n \\ (\alpha,(y_1,\ldots,y_n)) &\mapsto (x_1^\alpha,\ldots,x_n^\alpha) \end{aligned}$$

é um espaço vetorial[27]. Além disso, o elemento neutro é

[27] Mostre as oito propriedades.

$$0 = (1,1,\ldots,1)$$

e para todo $v = (x_1,\ldots,x_n)$ vale que

$$-v = \left(\frac{1}{x_1},\ldots,\frac{1}{x_n}\right).$$

Uma base desse espaço é dada por

$$\mathcal{B} = \{(e,1,\ldots,1),(1,e,\ldots,1),\ldots,(1,1,\ldots,e)\}$$

e as coordenadas de um vetor $v = (x_1,\ldots,x_n)$ nessa base são

$$v = \begin{bmatrix} \ln x_1 \\ \vdots \\ \ln x_n \end{bmatrix}_{\mathcal{B}}.$$

Transformações Lineares

Sejam V, W espaços vetoriais. Dizemos que funções

$$T : V \to W$$

satisfazendo

$$T(\alpha v + v') = \alpha T(v) + T(v')$$

para todo $\alpha \in \mathbb{R}$ e $v, v' \in V$ são *transformações lineares*. Embora essa restrição pareça simples, ela impõe diversas propriedades sobre T. Primeiro note que $T(0) = 0$ pois

$$0 = T(0) - T(0) = T(0+0) - T(0) = T(0) + T(0) - T(0) = T(0).$$

Mais surpreendente: para definir uma transformação linear de V em W basta definir os valores que ela assume em uma base $v_1, \ldots, v_n$ de V. Ou seja, o valor de $T(v)$ está definido para todo $v \in V$ uma vez que se sabe $T(v_1), \ldots, T(v_n)$. Basta observar que se

$$v = \alpha_1 v_1 + \cdots + \alpha_n v_n,$$

então

$$T(v) = T\left(\sum_{i=1}^{n} \alpha_i v_i\right) = \sum_{i=1}^{n} \alpha_i T(v_i).$$

Podemos escrever T matricialmente uma vez que fixamos uma base para V e W. Seja $v_1, \ldots, v_n$ a base de V e $w_1, \ldots, w_m$ a base de W. Seja

$$T(v_i) = \begin{bmatrix} \beta_{i1} \\ \vdots \\ \beta_{im} \end{bmatrix}_{w_1,\ldots,w_m},$$

então dado $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$ vale que

$$T(v) = \sum_{i=1}^{n} \alpha_i T(v_i) = \begin{bmatrix} \beta_{11} & \cdots & \beta_{1n} \\ \vdots & \ddots & \vdots \\ \beta_{1m} & \cdots & \beta_{nm} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}.$$

Ou seja, podemos obter as coordenadas de $T(v)$ usando multiplicando as coordenadas de v por uma matriz cujas colunas são as coordenadas de $T(v_i)$. Isso mostra que, uma vez que fixamos uma base para V e uma para W, uma transformação linear pode ser vista como uma matriz.

A volta também vale, se V tem dimensão n e W tem dimensão m então uma matriz $m \times n$ pode ser vista como uma transformação linear uma vez que fixamos as bases de V e de W.

Núcleo e Imagem

Sejam V, W espaços vetoriais de dimensão finita e $T : V \to W$ uma transformação linear. Existem dois subespaços[28] relevantes associados a T, são eles o núcleo e a imagem. O núcleo é definido como

$$N(T) = \{v \in V : T(v) = 0\} \subset V$$

[28] Mostre que são subespaços.

e a imagem como

$$Im(T) = \{T(v) : v \in V\} \subset W.$$

Note que se $N(T) = \{0\}$, então T é injetiva, pois caso contrário haveriam $v, v' \in V$ tais que $v - v' \neq 0$, $T(v) = T(v')$ e, portanto, $v - v' \in N(T)$. Por outro lado, se $Im(T) = W$, então a transformação é sobrejetiva. Um fato relevante sobre transformações lineares é o seguinte resultado:

Teorema 4 (Teorema do núcleo e da imagem ou Teorema do posto e nulidade)**.** *Se $T : V \to W$ é uma transformação linear e V tem dimensão finita, então*

$$dim(V) = dim(N(T)) + dim(Im(T)).$$

Demonstração. Seja $v_1, \ldots, v_r$ uma base de $N(T)$. Podemos completar essa base com vetores $v_{r+1}, \ldots, v_n$ para obter uma base de V. Mostraremos que $T(v_{r+1}), \ldots, T(v_n)$ é uma base de $Im(T)$ e dai teremos mostrado o que desejávamos.

Primeiro mostramos que $T(v_{r+1}), \ldots, T(v_n)$ gera $Im(T)$. Tome $w \in Im(T)$, então $w = T(v)$ para algum v. Tem-se

$$v = \sum_{i=1}^{n} \alpha_i v_i$$

e, portanto,

$$T(v) = \sum_{i=1}^{n} \alpha_i T(v_i).$$

Mas como $T(v_i) = 0$ para todo $i \leq r$ resta que

$$T(v) = \sum_{i=r+1}^{n} \alpha_i T(v_i).$$

Resta então mostrar que $T(v_{r+1}), \ldots, T(v_n)$ é linearmente independente. Ou seja, queremos mostrar que

$$\sum_{i=r+1}^{n} \alpha_i T(v_i) = 0 \Leftrightarrow \alpha_{r+1} = \cdots = \alpha_n = 0.$$

O lado $(\Leftarrow)$ é óbvio. Mostramos então o lado $(\Rightarrow)$. Observe que

$$\sum_{i=r+1}^{n} \alpha_i T(v_i) = 0 \Leftrightarrow T\left(\sum_{i=r+1}^{n} \alpha_i v_i\right) = 0 \Leftrightarrow \sum_{i=r+1}^{n} \alpha_i v_i \in N(T).$$

Por outro lado,

$$\sum_{i=r+1}^{n} \alpha_i v_i \in < v_{r+1}, \ldots, v_n > .$$

Portanto, $\sum_{i=r+1}^{n} \alpha_i v_i = 0$ e como $v_{r+1}, \ldots, v_n$ são linearmente independentes tem-se $\alpha_{r+1} = \cdots = \alpha_n = 0$. $\square$

Embora o teorema do núcleo e da imagem pareça uma novidade, ele não é. É possível interpretar esse teorema em termos da decomposição LU. Assuma que $dim(V) = n$. Ao fixar uma base para V e uma para W podemos ver uma transformação $T : V \mapsto W$ como uma matriz M_T. Ao fazer a decomposição LU teremos r pivôs e $n - r$ linhas nulas em U. Vale que

$$r = dim(Im(T)) \text{ e } n - r = dim(N(T)).$$

Por esse motivo, o teorema do núcleo e da imagem é conhecido também como teorema do posto e nulidade.

Algumas transformações especiais

Vamos explorar agora algumas transformações lineares interessantes.

Rotações e números complexos

O primeiro caso que vamos estudar é das matrizes de rotação. Seja θ um ângulo, e $v = \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2$. Ao rotacionar v por θ obtemos

$$\begin{bmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \end{bmatrix}.$$

Podemos reescrever o vetor acima como

$$\begin{bmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \end{bmatrix} = x\begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} + y\begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}.$$

Portanto, uma rotação é uma transformação linear e pode ser representada, na base canônica, por

$$R_\theta := \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

Note também que qualquer matriz da forma

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix},$$

com $a, b \in \mathbb{R}$, pode ser vista como uma rotação multiplicada por um escalar não negativo, basta observar que

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \sqrt{a^2 + b^2}\begin{bmatrix} \frac{a}{\sqrt{a^2+b^2}} & -\frac{b}{\sqrt{a^2+b^2}} \\ \frac{b}{\sqrt{a^2+b^2}} & \frac{a}{\sqrt{a^2+b^2}} \end{bmatrix}$$

e então existe $\theta = \theta(a, b)$ tal que $\cos\theta = \frac{a}{\sqrt{a^2+b^2}}$ e $\sin\theta = \frac{b}{\sqrt{a^2+b^2}}$. Logo,

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \sqrt{a^2 + b^2}\, R_{\theta(a,b)}.$$

Por outro lado, todo número complexo pode ser escrito como $re^{i\theta}$ onde $r \geq 0$ é um real e $e^{i\theta} = \cos\theta + i\sin\theta$. Isso sugere que podemos mapear os números complexos nas matrizes da forma

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix}.$$

Para formalizar essa ideia definimos

$$V = \left\{ \begin{bmatrix} a & -b \\ b & a \end{bmatrix}, a, b \in \mathbb{R} \right\},$$

que pode ser facilmente visto como um espaço vetorial de dimensão 2 com as operações de soma e produto usuais.

Por outro lado, os números complexos ($\mathbb{C}$) também podem ser vistos como um espaço vetorial de dimensão 2 com as operações de soma e produto por escalar (por real) usuais, basta tomar a base $\{1, i\}$. Podemos mapear um espaço no outro com a seguinte transformação linear:

$$\begin{aligned} T : V &\to \mathbb{C} \\ \begin{bmatrix} a & -b \\ b & a \end{bmatrix} &\mapsto a + bi \end{aligned}$$

Note também que essa transformação preserva o produto e que, em V, o produto é comutativo. Ou seja[29], dados $A, B \in V$,

[29] Mostre!

$$T(BA) = T(AB) = T(A)T(B) = T(B)T(A).$$

Matriz de mudança de base

Outra família de matrizes importantes são as matrizes de mudança de base. Essas matrizes permitem levar as coordenadas de um vetor em uma base nas coordenadas de um vetor em outra base.

Seja V um espaço vetorial de dimensão finita n e $\mathcal{B} = (v_1, \ldots, v_n)$, $\mathcal{B}' = (v_1', \ldots, v_n')$ duas bases. Sejam

$$v_i = \begin{bmatrix} \alpha_{1i} \\ \vdots \\ \alpha_{ni} \end{bmatrix}_{\mathcal{B}'}, \; i \in [n].$$

Tome

$$M_{\mathcal{B}}^{\mathcal{B}'} = \begin{bmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & \ddots & \vdots \\ \alpha_{n1} & \cdots & \alpha_{nn} \end{bmatrix},$$

então se x são as coordenadas de um vetor v na base $\mathcal{B}$ valerá que $y = M_{\mathcal{B}}^{\mathcal{B}'} x$ são as coordenadas do mesmo vetor na base $\mathcal{B}'$.

Exercícios

Exercício 7. *Mostre que $v_1, \ldots, v_n \in V$ são linearmente independentes se, e somente se,*

$$< w_1, \ldots, w_r > \subsetneq < v_1, \ldots, v_n >$$

para todo $\{w_1, \ldots, w_r\} \subsetneq \{v_1, \ldots, v_n\}$.

Exercício 8 (Soma direta). *Seja V um espaço vetorial e W, W' subespaços de V. Assuma que $\{w_1, \ldots, w_r\}$ é base de W e que $\{w'_1, \ldots, w'_s\}$ é base de W'. Defina*

$$W + W' = \{w + w' : w \in W,\ w' \in W'\}.$$

(i) Mostre que $dim(W + W') \leq r + s$.

(ii) Mostre que se todo elemento $v \in W + W'$ pode ser escrito de uma única forma como soma de elementos $w \in W$ e $w' \in W'$[30] *então*

$$\{w_1, \ldots, w_r\} \cup \{w'_1, \ldots, w'_s\}$$

é base de $W + W'$.

[30] Nesse caso, dizemos que $W + W'$ é uma soma direta e denotamos $W \oplus W'$

Exercício 9. *Ache uma base e diga qual é a dimensão dos seguintes espaços vetoriais:*

(i) o espaço $\mathbb{R}_n[x]$ dos polinômios com coeficientes reais e grau menor que n.

(ii) o espaço das funções $f : S \to \mathbb{R}$, onde S é um conjunto finito com n elementos.

(iii) o espaço das sequências $(a_1, a_2, a_3, \ldots)$ que obedecem a regra

$$a_{k+2} = a_k + a_{k+1}\ \forall k \geq 1.$$

Exercício 10. *Seja V um espaço vetorial de dimensão finita e $v \in V$. Mostre que $v = 0$ se, e somente se,*

$$v = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}_{\mathcal{B}}$$

para toda base $\mathcal{B}$ de V.

Exercício 11. *Diga se cada conjunto de vetores abaixo é uma base de $\mathbb{R}_4[x]$ e, se for, encontre as coordenadas de $8x^3 + 2x - 1$*

(i) $\{x^3, x^2, x, 1\}$;

(ii) $\{x^3 + x^2, x^2 + x, x + 1, 1 + x^3\}$.

Exercício 12. *Dê exemplos de transformações lineares $T : \mathbb{R}^n \to \mathbb{R}^m$ (escolhendo n e m) tais que:*

(i) *T leva qualquer conjunto linearmente independente em V em um outro conjunto linearmente independente em W, mas não leva nenhum conjunto gerador de V em conjunto gerador de W.*

(ii) *T leva algum conjunto linearmente independente em um conjunto linearmente dependente.*

Exercício 13. *Sejam V, W, Z espaços vetoriais de dimensões m, n, p, respectivamente. Sejam $T : V \to W$ e $S : W \to Z$ transformações lineares.*

(i) *Fixadas bases para V, W, Z, quais as dimensões das matrizes M_T e M_S que representam as transformações T e S, respectivamente?*

(ii) *Mostre que a composição $S \circ T : V \to Z$ é uma transformação linear;*

(iii) *Mostre que fixada uma escolha de bases T é dada pela matriz M_T e S pela matriz M_S, mostre que $S \circ T$ é dada pela matriz produto $M_S M_T$.*

Exercício 14. *Seja T uma transformação linear e seja M sua matriz em uma base qualquer. Mostre que T possui inversa (como função) se, e somente se, M possui inversa (como matriz).*

Exercício 15. *Mostre que se $v_1, \ldots, v_n$ é um conjunto linearmente independente e $N(T) = \{0\}$, então $T(v_1), \ldots, T(v_n)$ é linearmente independente.*

Exercício 16. *Sejam n, m inteiros positivos. Encontre ou mostre que não existe:*

(i) *uma transformação linear injetiva $T : \mathbb{R}^n \to \mathbb{R}^m$ quando $n > m$;*

(ii) *uma transformação linear sobrejetiva $T : \mathbb{R}^n \to \mathbb{R}^m$ quando $n < m$;*

(iii) *uma transformação linear bijetiva $T : \mathbb{R}^n \to \mathbb{R}^n$.*

Exercício 17. *Seja $V = \mathbb{R}^2$. Encontre a matriz de mudança de base de*

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$$

para

$$\mathcal{B}' = \left\{ \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$$

e também a matriz de mudança de base de $\mathcal{B}'$ para $\mathcal{B}$.

Exercício 18 (Sistemas lineares pelo espaço das colunas)**.** *Seja A uma matriz $n \times m$. Mostre que:*

(i) *O sistema* $Ax = b$ *tem solução se, e somente se,* $b \in < Ae_1, \ldots, Ae_m >$.

(ii) *Se* $n > m$ *então sempre existirá* b *tal que* $Ax = b$ *não tem solução.*

(iii) *Se* $n < m$ *então existem infinitas soluções para* $Ax = 0$.

(iv) *Dado* v *tal que* $Av = b$ *então*

$$\{x : Ax = b\} = \{v + v' : Av' = 0\}.$$

(v) *O número de linhas linearmente independentes e de colunas linearmente independentes de* A *é o mesmo.*

Ortogonalidade: projeções, bases e coordenadas

Nesse capítulo continuaremos nossa investigação sobre bases, subespaços e coordenadas. Mas agora fazendo uso das noções de distância e ângulo. Embora as ideias desse capítulo sejam válidas em maior generalidade, vamos nos restringir ao espaço $V = \mathbb{R}^n$.

Projeções

Sejam W um subespaço de V e $v \in V$. Em geral, não vale que $v \in W$. Mas podemos nos perguntar qual o elemento $w \in W$ que mais se aproxima de v. Esse elemento é chamado de projeção de v em W, denotamos por

$$\mathrm{proj}_W(v) := \underset{w \in W}{\arg\min} \|v - w\|_2^2.$$

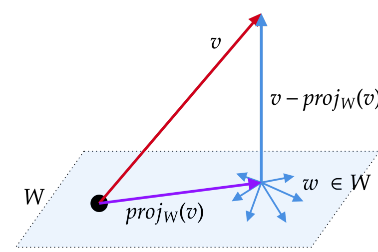

Figura 9: Ilustração da projeção de um vetor $v \in \mathbb{R}^3$ em um subespaço bidimensional W.

Podemos usar um pouco de geometria (ou alternativamente de cálculo) para obter uma expressão para $\mathrm{proj}_W(v)$. Primeiro notamos que como esse vetor minimiza a distância com v e, portanto, deve haver um ângulo reto entre $v - \mathrm{proj}_W(v)$ e o subespaço W. Em outras palavras,

$$\langle w, v - \mathrm{proj}_W(v) \rangle = 0 \ \forall w \in W.$$

A equação acima necessita que verifiquemos um ângulo reto com todos os vetores $w \in W$, mas para isso basta verificar um ângulo reto com todos os vetores de um conjunto gerador (não é necessário que seja uma base) $w_1, \ldots, w_r$ de W[31]. Ou seja, basta que

$$\langle w_i, v - \mathrm{proj}_W(v) \rangle = 0 \ \forall i \in [r].$$

[31] Prove!

Seja $A = \begin{bmatrix} w_1 & w_2 & \cdots & w_r \end{bmatrix}$ a matriz cujas colunas são os vetores $w_1, \ldots, w_r$. A equação acima é equivalente à seguinte identidade entre vetores:

$$A^T(v - \mathrm{proj}_W(v)) = 0$$

Que pode ser rearranjada no seguinte sistema linear:

$$A^T(\mathrm{proj}_W(v)) = A^T v.$$

O vetor $\text{proj}_W(v)$ também deve pertencer a W, ou seja, deve ser escrito da forma

$$\text{proj}_W(v) = Ax$$

para alguma escolha de coordenadas $x \in \mathbb{R}^r$. Portanto, para encontrar a projeção basta resolver o sistema linear

$$A^T Ax = A^T v$$

e então substituir x em Ax. Note que se $w_1, \ldots, w_r$ não é base então teremos infinitas soluções, já se for base a solução será única.

De fato, se $w_1, \ldots, w_r$ for uma base de W, então $A^T A$ tem inversa[32] e podemos escrever

$$x = (A^T A)^{-1} A^T v \tag{4}$$

e, portanto,

$$\text{proj}_W(v) = A(A^T A)^{-1} A^T v. \tag{5}$$

[32] De maneira geral vale que $A^T A$ tem o mesmo posto que A. Veremos isso nos exercícios.

Revisitando bases e coordenadas

Voltamos ao problema anterior, mas no caso em que $V = W$. Nesse caso a resposta é óbvia, $\text{proj}_V(v) = v$. Assumindo que a matriz A é construída tomando uma base de V como colunas, então A e A^T tem inversa. A equação (4) se resume a

$$x = (A^T A)^{-1} A^T v = A^{-1} (A^T)^{-1} A^T v = A^{-1} v,$$

ou seja, x não passa das coordenadas de v na base dada pelas colunas de A. De fato, x é a solução de $Ax = v$.

Em outras palavras, **para encontrar as coordenadas de um vetor numa base qualquer ou mesmo para obter a projeção de um vetor em um subespaço qualquer é necessário resolver um sistema linear.**

Embora a eliminação Gaussiana seja eficiente em resolver sistemas lineares, seu custo ainda é consideravelmente alto (da ordem de n^3 operações) e é possível construir bases para as quais não é necessário efetuar eliminação Gaussiana para computar coordenadas.

Assuma que $v_1, \ldots, v_n$ é uma base de V e que o vetor $v \in V$ seja escrito como

$$v = \sum_{i=1}^{n} \alpha_i v_i.$$

Vale que

$$\langle v, v_j \rangle = \sum_{i=1}^{n} \alpha_i \langle v_i, v_j \rangle.$$

Se $v_1, v_2, \ldots, v_n$ for escolhida tal que

$$\langle v_i, v_j \rangle = \begin{cases} 1, & \text{se } i = j \\ 0, & \text{c.c.} \end{cases}$$

então valerá que

$$\langle v, v_j \rangle = \sum_{i=1}^{n} \alpha_i \langle v_i, v_j \rangle = \alpha_j.$$

Em outras palavras, as coordenadas de v na base $v_1, v_2, \dots, v_n$ são dadas por $\langle v, v_1 \rangle, \langle v, v_2 \rangle, \dots, \langle v, v_n \rangle$. Essa discussão motiva a seguinte definição:

Definição 5 (Bases ortonormais). *Seja W um subespaço de V. Uma base $v_1, \dots, v_r$ de W é ortonormal se*

$$\langle v_i, v_j \rangle = \begin{cases} 1, & \text{se } i = j \\ 0, & \text{c.c.} \end{cases}.$$

Um exemplo típico de base ortogonal é a base canônica $e_1, \dots, e_n$. Mas dada uma base qualquer é possível construir uma base ortonormal usando o processo de Gram-Schmidt.

Gram-Schmidt

Nosso objetivo é partir de uma base $v_1, \cdots, v_r$ de um subespaço W de V e obter uma nova base $w_1, \dots, w_r$ do mesmo subespaço, mas que seja ortonormal.

Começamos tomando

$$w_1 = \frac{v_1}{\|v_1\|}.$$

Agora dado $1 \leq s < r$ assumimos que $w_1, \dots, w_s$ é base ortonormal do subespaço gerado por $v_1, \dots, v_s$. Definimos

$$w_{s+1} = \frac{v_{s+1} - \sum_{i=1}^{s} \langle v_{s+1}, w_i \rangle w_i}{\|v_{s+1} - \sum_{i=1}^{s} \langle v_{s+1}, w_i \rangle w_i\|}$$

Observe que $w_1, \cdots, w_{s+1}$ é base ortonormal do subespaço gerado por $v_1, \dots, v_{s+1}$ [33]. Daí segue que $w_1, \dots, w_r$ é base ortonormal de W.

[33] Prove!

Exercícios

Exercício 19. *Seja A uma matriz qualquer. Mostre que*

$$N(A) = N(A^T A)$$

e conclua que A e $A^T A$ possuem o mesmo posto.

Exercício 20. *Seja $v_1, \dots, v_n$ uma base ortonormal de $\mathbb{R}^n$. Mostre que a matriz $V = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$ formada por vetores coluna satisfaz $V^{-1} = V^T$.*

Exercício 21. *Seja $v_1, \ldots, v_n$ uma base ortonormal de V e W subespaço gerado por $v_1, \ldots, v_r$ para algum $r \leq n$. Mostre que*

$$proj_W(v) = \sum_{i=1}^{r} \langle v, v_i \rangle v_i.$$

Exercício 22. *Obtenha uma base ortogonal do subespaço de $\mathbb{R}^4$ gerado por*

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \ e \ \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}.$$

Área, Volume e...?

Se pegamos dois vetores em $\mathbb{R}^2$ conseguimos definir uma paralelogramo, já três vetores em $\mathbb{R}^3$ definem um paralelepípedo. O interessante é que dois vetores linearmente dependentes em $\mathbb{R}^2$ definem um paralelogramo de área zero! O mesmo vale para o volume de um paralelepípedo definido por vetores linearmente dependentes. Ou seja, nesses casos, para checar se 2 vetores em $\mathbb{R}^2$ são linearmente independentes basta checar se a área do paralelogramo é não nula, para checar se 3 vetores em $\mathbb{R}^3$ são linearmente independentes basta checar se o volume do paralelepípedo é não nulo.

Nosso objetivo neste capítulo é generalizar essas ideias para $\mathbb{R}^n$. Pra isso, vamos generalizar as noções de área e volume. É nesse contexto que vai surgir o determinante.

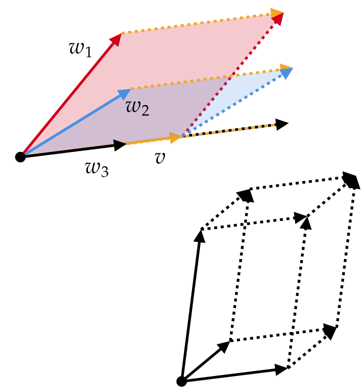

Figura 10: Em cima: fixamos o vetor $v \in \mathbb{R}^2$ e mostramos o paralelogramo formado por diferentes escolhas de w_i, $i = 1, 2, 3$. Note que v, w_3 são linearmente dependentes e verifica-se um paralelogramo sem área. Embaixo: ilustração de 3 vetores em $\mathbb{R}^3$ formando um paralelepípedo.

Generalizando área e volume para $\mathbb{R}^n$

Queremos associar um valor real para um conjunto de n vetores em $\mathbb{R}^n$. E queremos que esse valor carregue um sentido similar ao sentido de área e volume. Além disso, gostaríamos que nos casos $n = 2, 3$ esse valor encontrasse a área e o volume, respectivamente. Formalmente, fixado n, queremos encontrar uma função

$$\begin{aligned} D : \mathbb{R}^n \times \cdots \times \mathbb{R}^n &\to \mathbb{R} \\ (v_1, \ldots, v_n) &\mapsto D(v_1, \ldots, v_n). \end{aligned}$$

A função D precisa satisfazer algumas propriedades para que "carregue um sentido similar ao sentido de área e volume". Acontece que existem três propriedades básicas que acabam por definir unicamente a função D. São elas:

(i) Se dois vetores são iguais então D deve se anular. Em outras palavras, se existem $i \neq j$ tais que $v_i = v_j$, então

$$D(v_1, \ldots, v_n) = 0.$$

Em duas dimensões essa propriedade diz que dois vetores iguais formam um paralelogramo de área nula. Em três dimensões, note que o espaço gerado por v_1, v_2, v_3 quando dois vetores coincidem tem dimensão no máximo 2. E, portanto, não faz sentido que tenha volume.

(ii) A área do hipercubo formado pela base canônica deve ser sempre 1. Ou seja,

$$D(e_1, \ldots, e_n) = 1.$$

Note que isso ocorre em dimensões 2 e 3.

(iii) Por fim, essa função deve ser linear em cada coordenada. Ou seja, dados $v, w \in \mathbb{R}^n$ e $\alpha \in \mathbb{R}$, então

$$\begin{aligned} D(v_1, \ldots, v_{i-1}, v + \alpha w, v_{i+1}, \ldots, v_n) =& D(v_1, \ldots, v_{i-1}, v, v_{i+1}, \ldots, v_n) \\ &+ \alpha D(v_1, \ldots, v_{i-1}, w, v_{i+1}, \ldots, v_n) \end{aligned}$$

para todo $i \in [n]$.

Essa propriedade talvez seja a menos intuitiva das três, mas é elucidada pela Figura 11 no caso $n = 2$. Desenhos análogos são possíveis para $n = 3$.

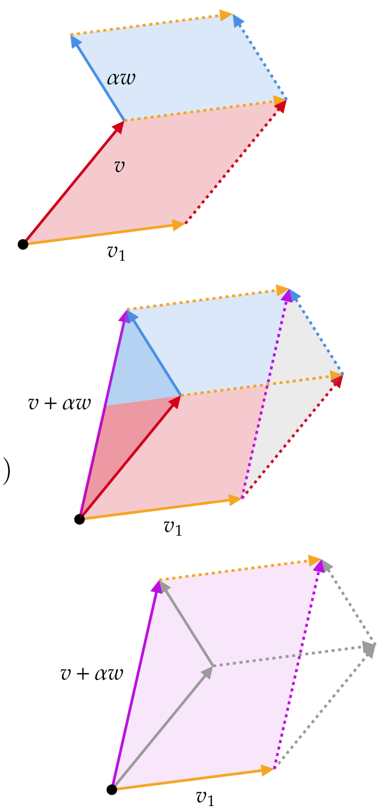

Figura 11: Ao somar a área dos paralelogramos dados por v, v_1 e por $\alpha w, v_1$ obtemos a área do paralelogramo dado por $v + \alpha w, v_1$.

Existe algo estranho acontecendo: é possível que D assuma valores negativos mesmo quando $n = 2, 3$. Justificaremos um valor negativo notando que D carrega não apenas a informação sobre área ou volume, que é dada por $|D|$, mas também informação sobre a orientação dos vetores. Discutiremos isso em detalhes no texto que segue. Antes disso, vamos elaborar um pouco mais sobre as propriedades que qualquer função satisfazendo $(i) - (iii)$ deve também satisfazer.

Lema 3. *Qualquer função D satisfazendo $(i) - (iii)$ é tal que*

- *D é alternada: ao trocar dois vetores de lugar trocamos também o sinal de D, ou seja,*

  $$D(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) = -D(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_n)$$

  para todo $i \neq j$.

- *Se $v_1, \ldots, v_n$ é linearmente dependente, então $D(v_1, \ldots, v_n) = 0$.*

Demonstração. Começamos pela primeira propriedade. Como apenas os vetores v_i e v_j são trocados denotaremos $D(v_i, v_j)$ no lugar de $D(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n)$. Usando as propriedades (i) e (iii) temos:

$$\begin{aligned} D(v_i, v_j) &= D(v_i, v_j) + D(v_j, v_j) \\ &= D(v_i + v_j, v_j) \\ &= D(v_i + v_j, v_j) - D(v_i + v_j, v_i + v_j) \\ &= -D(v_i + v_j, v_i) \\ &= -D(v_j, v_i). \end{aligned}$$

Para a segunda propriedade vamos assumir, sem perda de generalidade, que $v_1 = \alpha_2 v_2 + \cdots + \alpha_n v_n$, usando as propriedades (iii) e (i) obtemos

$$D(v_1, \ldots, v_n) = D\left(\sum_{i=2}^{n} \alpha_i v_i, v_2, \ldots, v_n\right) = \sum_{i=2}^{n} \alpha_i D\left(v_i, v_2, \ldots, v_n\right) = 0.$$

$\square$

Agora precisamos lidar com as duas questões que deixamos em aberto:

- Como interpretar o sinal de D? O que é orientação?
- Mostrar que existe, de fato, uma única função D satisfazendo as propriedades $(i) - (iii)$.

A resposta para o primeiro problema não é dada via formalismo matemático, afinal estamos em busca de uma interpretação. Vamos começar pela primeira questão e buscar ganhar mais entendimento sobre como deve ser a função D que estamos buscando. Após construir um bom candidato para D vamos mostrar que esse candidato de fato satisfaz as propriedades desejadas e é único.

Permutações e Orientação

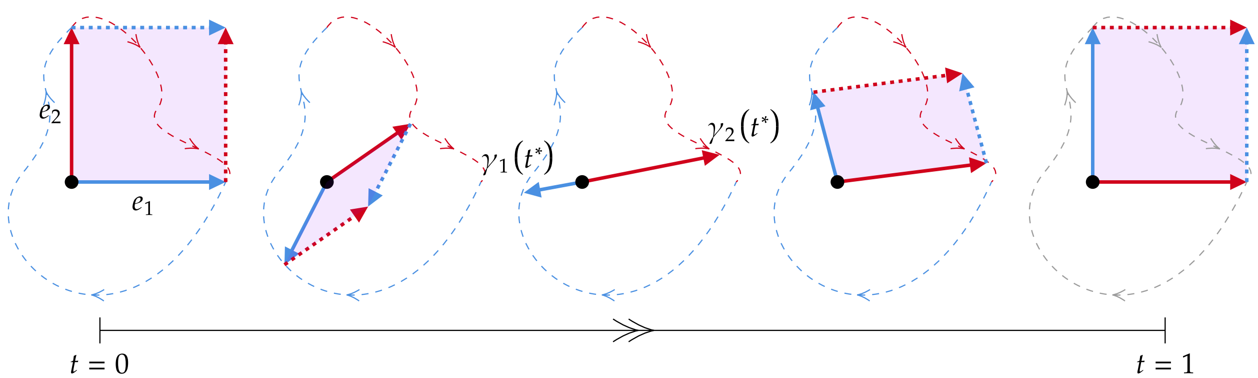

Figura 12: Ao distorcer continuamente dois vetores para que eles troquem de posição teremos sempre haverá um momento em que o paralelogramo definido por eles terá área nula, nesse momento dizemos que houve uma mudança de orientação.

Para começar a entender a presença de um sinal negativo em D e formalizar a ideia de orientação começaremos olhando para $n = 2$. Pela propriedade (ii), sabemos que $D(e_1, e_2) = 1$ e pelo lema anterior temos $D(e_2, e_1) = -1$.

A mudança de sinal pode ser entendida olhando para a Figura 12. Tomamos duas curvas contínuas γ_1, γ_2 satisfazendo $\gamma_1(0) = \gamma_2(1) = e_1$ e $\gamma_2(0) = \gamma_1(1) = e_2$. Por simplicidade, podemos assumir que a união dessas curvas forma um circuito fechado sem auto-interseções. Essas curvas deformam os vetores da base canônica continuamente e terminam por trocar a posição desses vetores conforme o tempo evolui de $t = 0$ para $t = 1$. É possível mostrar (e é intuitivo) que existe um (único) tempo t^* no qual o paralelogramo formado por $\gamma_1(t^*)$ e $\gamma_2(t^*)$ tem área zero! É nesse momento que a orientação deixa de ser positiva e se torna negativa.

Essa mesma intuição pode ser generalizada para $n > 2$. Faremos agora essa generalização, enquanto também formalizamos a noção de orientação fazendo uso de propriedades de permutações. Se

$$p : [n] \to [n]$$

é uma bijeção, então dizemos que p é uma permutação dos elementos de n. Vamos pegar emprestado (sem provar) o seguinte fato sobre permutações:

Lema 4. *Toda permutação p pode ser escrita como*

$$p = \beta_k \circ \cdots \circ \beta_1 \tag{6}$$

onde cada β_i é uma permutação que troca apenas dois elementos de lugar. Além disso, se também vale

$$p = \beta'_{k'} \circ \cdots \circ \beta'_1,$$

onde β'_i são permutações que trocam apenas dois elementos de lugar, então k e k' tem a mesma paridade.

Esse fato nos permite definir a paridade ou sinal de uma permutação:

Definição 6. *Seja p uma permutação escrita da forma (6), então a paridade ou sinal de p é denotado por $\sigma(p)$ e satisfaz*

$$\sigma(p) := (-1)^k.$$

Logo, podemos ver uma permutação de n elementos como várias permutações de dois elementos. E o argumento da Figura 12 continua válido, basta repeti-lo sequencialmente para cada permutação de dois elementos. Mais ainda: para cada permutação de dois elementos vamos trocar o sinal uma vez, justificando o seguinte resultado:

Lema 5. *Se D satisfaz as propriedades $(i) - (iii)$ e p é uma permutação, então vale que*

$$D\left(e_{p(1)}, e_{p(2)}, \ldots, e_{p(n)}\right) = \sigma(p).$$

Demonstração. Segue da discussão acima, mas formalizar esse resultado partindo de (6) pode ser um bom exercício. □

Existência e unicidade de D

Com a discussão que já tivemos é fácil encontrar um candidato para D. Seja

$$v_j = \sum_{i=1}^{n} v_{ij} e_i$$

para todo $j \in [n]$. Usando a propriedade (iii) temos:

$$\begin{aligned} D(v_1, \ldots, v_n) &= D\left(\sum_{i=1}^{n} v_{i1} e_i, \ldots, \sum_{i=1}^{n} v_{in} e_i\right) \\ &= \sum_{f:[n]\to[n]} D\left(v_{f(1)1} e_{f(1)}, \ldots, v_{f(n)n} e_{f(n)}\right). \end{aligned}$$

Agora note que toda f que não for uma permutação é tal que $f(i) = f(j)$ para $i \neq j$, de maneira que, nesses casos, D se anula. Portanto, basta somar sobre as permutações:

$$D(v_1, \ldots, v_n) = \sum_{p \text{ permutação}} D\left(v_{p(1)1} e_{p(1)}, \ldots, v_{p(n)n} e_{p(n)}\right).$$

Pela discussão anterior e pela propriedade (iii) conseguimos

$$D\left(v_{p(1)1} e_{p(1)}, \ldots, v_{p(n)n} e_{p(n)}\right) = \sigma(p)\left(v_{p(1)1} \cdots v_{p(n)n}\right)$$

e, daí,

$$D(v_1, \ldots, v_n) = \sum_{p \text{ permutação}} \sigma(p)\left(v_{p(1)1} \cdots v_{p(n)n}\right). \tag{7}$$

Perceba que chegamos na fórmula acima para D usando apenas as propriedades $(i) - (iii)$. Isso não garante que a formula acima nos dá uma função D satisfazendo as três propriedades, mas garante que essa é a única candidata possível. Segue então que:

Teorema 5. *A única função que satisfaz as propriedades $(i) - (iii)$ é*

$$D(v_1, \ldots, v_n) = \sum_{p \text{ permutação}} \sigma(p)\left(v_{p(1)1} \cdots v_{p(n)n}\right).$$

Demonstração. A unicidade é garantida uma vez que verificarmos que a formula acima satisfaz as propriedades $(i) - (iii)$. Deixamos como exercício verificar as propriedades. □

Por fim, observe que uma matriz quadrada $n \times n$ pode ser vista como uma matriz cujas colunas são vetores e, portanto, podemos associar uma noção de volume para uma matriz usando a função D que construímos:

Definição 7. *Seja A uma matriz $n \times n$ da forma*

$$A = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix},$$

onde $v_1, \ldots, v_n \in \mathbb{R}^n$. Definimos o determinante de A como

$$det(A) := D(v_1, \ldots, v_n).$$

O determinante de uma matriz tem uma interpretação geométrica natural: é fator de escala da transformação $v \mapsto Av$, ou seja, o quanto A distorce o cubo definido pela base canônica. A Figura 13 ilustra essa intuição no caso $n = 3$, onde o determinante é o volume da imagem do cubo unitário.

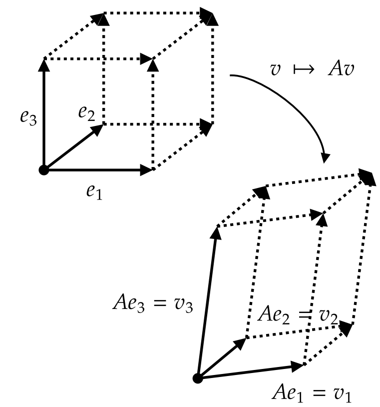

Figura 13: Uma matriz A de tamanho 3×3 pode ser vista como uma transformação linear que leva o cubo definido pela base canônica no paralelepípedo definido pelos vetores v_1, v_2, v_3 que correspondem as colunas de A.

Propriedades do determinante

Nós agora elaboramos algumas propriedades do determinante:

- Se $A = (a_{ij})_{n\times n}$ é uma matriz quadrada, então $det(A^T) = det(A)$.

  Para mostrar essa propriedade podemos usar a fórmula do determinante:

  $$det(A) = \sum_{p \text{ permutação}} \sigma(p) \left(a_{1p(1)} \cdots a_{np(n)}\right)$$

  Somar sobre todas as permutações é a mesma coisa que somar sobre o inverso de todas as permutações, logo:

  $$det(A) = \sum_{p \text{ permutação}} \sigma(p^{-1}) \left(a_{p^{-1}(1)1} \cdots a_{p^{-1}(n)n}\right)$$

  Como a paridade de p e de p^{-1} é a mesma[34], vale:

  $$det(A) = \sum_{p \text{ permutação}} \sigma(p) \left(a_{p(1)1} \cdots a_{p(n)n}\right) = det(A^T).$$

- Se $A = BC$ e B, C são matrizes quadradas, então $det(A) = det(B)det(C)$.

  Para mostrar esse resultado vamos fazer uso da seguinte propriedade: $det(EC) = det(E)det(C)$ sempre que E é uma matriz elementar[35].

  Primeiro note que se B tem linhas linearmente dependentes, então A também terá e valerá $0 = det(A) = det(B)det(C)$.

  Se as linhas de B são linearmente independentes, o processo de eliminação Gaussiana pode ser executado para obter

  $$B = E_k \ldots E_1 I_n,$$

[34] Mostre!

[35] Essa propriedade é deixada como exercício.

onde E_i é operação elementar para todo i. Logo:

$$A = E_k \dots E_1 I_n C = E_k \dots E_1 C$$

Escrevemos

$$\begin{aligned} det(A) &= det(E_k \dots E_1 C) \\ &= det(E_k) det(E_{k-1} \dots E_1 C) \\ &= \cdots \\ &= det(E_k) \cdots det(E_1) det(C) \end{aligned}$$

e usamos que $det(B) = det(E_k) \cdots det(E_1)$ para obter

$$det(A) = det(B) det(C).$$

- Se S é inversível, então $det(S^{-1}) = \frac{1}{det(S)}$.
  Basta usar a propriedade anterior e fazer
  $$1 = det(I) = det(SS^{-1}) = det(S) det(S^{-1}).$$
- Em especial, segue que se S é inversível, então $det(SAS^{-1}) = det(A)$.

O resultado acima permite estender a definição de determinante de matriz para transformações lineares:

Definição 8. *Seja V um espaço vetorial de dimensão finita n e $T : V \to V$ uma transformação linear. Definimos*

$$det(T) = det(A)$$

onde A é a representação matricial de T fixado um sistema de coordenadas qualquer.

A definição acima faz sentido pois dadas duas possíveis matrizes A, B representando T em duas bases quaisquer, podemos usar as matrizes de mudança de base, que são inversíveis, para alternar entre as bases. Note também que essa definição permite falar sobre o determinante não apenas de transformações em $\mathbb{R}^n$, mas em qualquer espaço vetorial de dimensão finita.

Exercícios

Exercício 23. *Mostre que a fórmula (7) satisfaz as propriedades $(i) - (iii)$.*

Exercício 24. *Nesse exercício mostraremos que o determinante satisfaz o princípio de Cavalieri. Sejam $v_1, \dots, v_n \in \mathbb{R}^n$ e $w \in span(v_2, v_3, \dots, v_n)$. Mostre que*

$$D(v_1, \dots, v_n) = D(v_1 + w, v_2, \dots, v_n).$$

Exercício 25. *Mostre que o determinante de uma matriz triangular (superior ou inferior) é dado pelo produto das entradas da diagonal principal.*

Exercício 26. *Use a fórmula (7) para mostrar que*

(i) $det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = ab - cd$

(ii) *e que se* $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$, *então*

$$\begin{aligned} det\,(A) = \; & a_{11}(a_{22}a_{33} - a_{23}a_{32}) \\ & - a_{12}(a_{21}a_{33} - a_{23}a_{31}) \\ & + a_{13}(a_{21}a_{32} - a_{22}a_{31}). \end{aligned}$$

Nos próximos exercícios vamos explorar as relações entre o determinante e a decomposição PA=LU.

Exercício 27. *Seja A uma matriz quadrada. Usando as propriedades* $(i) - (iii)$, *mostre que:*

(i) *Se E é uma transformação elementar de troca de linha $L_i \longleftrightarrow L_j$, $i \neq j$, então*

$$det(EA) = det(E)det(A)$$

e $det(E) = -1$.

(ii) *Se E é uma transformação elementar de multiplicação de linha por escalar $L_i \longleftarrow \alpha L_i$, então*

$$det(EA) = det(E)det(A)$$

e $det(E) = \alpha$.

(iii) *Se E é uma transformação elementar de somar linha por múltiplo $L_i \longleftarrow L_i + \alpha L_j$, $i \neq j$, então*

$$det(EA) = det(E)det(A)$$

e $det(E) = 1$.

Exercício 28. *Seja $PA = LU$ a decomposição de A como no Teorema 1. Mostre que:*

- *Se U possui algum zero na diagonal, então* $det(A) = 0$.
- *O determinante de P é dado pela quantidade de trocas de linha feitas durante a decomposição* $PA = LU$.

- *Se U não possui zeros na diagonal principal então $det(A) = det(L)det(P)$. Note que $det(L)$ é apenas o produto das entradas da diagonal de L.*
- *Conclua que $det(A) \neq 0$ se, e somente se, A tem inversa.*

Observe que o exercício anterior também fornece um algoritmo para o cálculo do determinante: basta fazer Eliminação Gaussiana. Como já vimos o processo de Eliminação Gaussiana tem da ordem de n^3 operações. Por outro lado, a fórmula (7) envolve a soma de $n!$ termos, um custo computacional impraticável.

Autovalores e autovetores

Se $A = (a_{ij})_{ij}$ é uma matriz diagonal, então calcular Ax é trivial: basta multiplicar a i-ésima entrada de x por a_{ii}. Por exemplo:

$$\begin{bmatrix} 7 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 7x_1 \\ -3x_2 \\ -x_3 \end{bmatrix}.$$

Sabemos que uma mesma transformação linear pode ter diferentes representações em diferentes bases. Uma pergunta natural é: dada uma transformação $T : V \to V$, existe alguma base onde a matriz que representa T é diagonal?

Transformações e matrizes diagonalizáveis

Definição 9. *Dizemos que uma transformação linear $T : V \to V$ é diagonalizável quando existe uma base $v_1, \ldots, v_n$ na qual a matriz de T é diagonal.*

Dizemos também que uma matriz A é diagonalizável quando existe S inversível tal que $S^{-1}AS$ é diagonal.

Seja A uma matriz diagonalizável, então $A = SDS^{-1}$ para alguma matriz diagonal D com diagonal $\lambda_1, \lambda_2, \ldots, \lambda_n$. Olhando para S como uma matriz de colunas

$$S = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$$

temos que $Se_i = v_i$ para todo i. Logo, $S^{-1}v_i = e_i$. Isso significa que

$$\begin{aligned} Av_i &= SDS^{-1}v_i \\ &= SDe_i \\ &= \lambda_i Se_i \\ &= \lambda_i v_i. \end{aligned}$$

Ou seja, se uma matriz A é diagonalizável então existe uma base de vetores $v_1, \ldots, v_n$ e respectivas constantes $\lambda_1, \ldots, \lambda_n$ tais que

$$Av_i = \lambda_i v_i.$$

Note que a volta também vale, se existe uma base $v_1, \dots, v_n$ e constantes $\lambda_1, \dots, \lambda_n$ tais que $Av_i = \lambda_i$ então podemos construir S como anteriormente e $S^{-1}AS$ será diagonal já que

$$S^{-1}ASe_i = S^{-1}Av_i = S^{-1}(\lambda_i v_i) = \lambda_i S^{-1}v_i = \lambda_i e_i.$$

Dai tiramos nosso primeiro teorema:

Teorema 6. *Uma matriz A é diagonalizável se, e somente se, existe uma base de vetores $v_1, \dots, v_n$ e constantes $\lambda_1, \dots, \lambda_n$ tais que $Av_i = \lambda v_i$.*

Demonstração. Veja discussão acima. □

Os vetores $v_1, \dots, v_n$ e as constantes $\lambda_1, \dots, \lambda_n$ tem nome e serão nosso principal objeto de estudo nesse capítulo.

Definição 10. *Dizemos que $v \neq 0$ é autovetor de A com autovalor correspondente λ quando*

$$Av = \lambda v.$$

A definição acima permite enunciar o teorema anterior de maneira mais compacta: **uma matriz é diagonalizável quando possui uma base de autovetores**.

Encontrando autovalores e autovetores

A igualdade $Av = \lambda v$ é, na verdade, um sistema linear. Uma maneira de reescrever esse sistema é

$$(\lambda I - A)v = 0.$$

Estamos interessados em valores de λ para os quais é possível achar soluções não triviais do sistema acima. Ou seja, estamos interessados nos valores de λ para os quais $det(\lambda I - A) = 0$.

Definição 11. *Dada uma matriz A definimos seu polinômio característico como sendo $p_A(x) = det(xI - A)$.*

Queremos, então, encontrar as raízes do polinômio característico de A. Mas antes de prosseguir, observamos duas propriedades do polinômio característico:

Teorema 7. *O polinômio característico de uma matriz A qualquer satisfaz as seguintes propriedades:*

- *Se $B = S^{-1}AS$ para alguma matriz S inversível, então $p_A = p_B$;*
- *Ele é mônico de grau n e, portanto, tem n raízes (considerando multiplicidade).*

Demonstração. A prova da primeira propriedade segue observando que

$$\begin{aligned} det(xI - A) &= det(xI - S^{-1}BS) \\ &= det(S^{-1}xIS - S^{-1}BS) \\ &= det(S^{-1}(xI - B)S) \\ &= det(xI - B). \end{aligned}$$

A prova de que é mônico de grau n é deixada como exercício, mas note que uma vez que sabemos que o polinômio tem grau n o teorema fundamental da álgebra nos diz que ele possui n raízes (não necessariamente distintas[36]). □

Temos então que as raízes de p_A serão os autovalores de A. Já se λ é raiz de p_A então os elementos de $N(\lambda I - A)$ serão nos autovetores associados à λ.

Definição 12. *Seja λ uma raiz do polinômio característico p_A de uma matriz A. Dizemos que a multiplicidade de λ como raiz de p_A é a multiplicidade algébrica de λ, já a multiplicidade geométrica de λ é definida como sendo $dim(E_\lambda)$, onde $E_\lambda = N(\lambda I - A)$ é dito o autoespaço associado à λ.*

Por exemplo, a matriz

$$B = \begin{bmatrix} 7 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

possui polinômio característico $p_B(x) = (x-7)(x-7)(x-1)$, com raízes 7 e 1. O autovalor 7 tem multiplicidade algébrica 2 e o autovalor 1 tem multiplicidade algébrica 1. Por outro lado,

$$dim(N(7I - B)) = dim(span(e_1, e_2)) = 2$$

e

$$dim(N(1I - B)) = dim(span(e_3)) = 1.$$

Logo, o autovalor 7 tem multiplicidade geométrica 2, já o autovalor 1 tem multiplicidade geométrica 1.

Nesse momento, já conseguimos construir uma estratégia para diagonalizar uma matriz:

- Encontramos os autovalores $\lambda_1, \ldots, \lambda_m$ de A.
- Encontramos os autoespaços associados $E_{\lambda_1}, \ldots, E_{\lambda_m}$ e também obtemos bases para cada um desses espaços.
- Se for possível tomar $v_1, \ldots, v_n$ como sendo uma base de $\mathbb{R}^n$ onde cada v_i pertence ao autoespaço de algum autovalor, então basta construir a matriz S e a matriz D como discutido inicialmente.

[36] A multiplicidade de uma raiz é o número de vezes que ela aparece ao escrevermos um polinômio como produto de fatores de grau um. Por exemplo, o polinômio $x^3 - 7x^2 + 8x + 16$ tem raízes -1 com multiplicidade um e 4 com multiplicidade dois já que pode ser escrito como $(x+1)(x-4)(x-4)$. Note também que o teorema fundamental da álgebra garante a existência de raízes apenas nos complexos, mas não vamos nos preocupar com isso no momento.

- Se não for possível... não sabemos o que fazer.

Tente usar a estratégia acima para diagonalizar as seguintes matrizes:

(i) $\begin{bmatrix} -4 & -6 \\ 3 & 5 \end{bmatrix}$

(ii) $\begin{bmatrix} 3 & 0 & 0 \\ -3 & 4 & 9 \\ 0 & 0 & 3 \end{bmatrix}$

(iii) $\begin{bmatrix} 3 & 12 & -21 \\ -1 & -6 & 13 \\ 0 & -2 & 6 \end{bmatrix}$

(iv) $\begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}$

(v) $\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$

A estratégia deve funcionar normalmente nos três primeiros casos, o quarto também funciona[37], mas envolve operar no corpo dos complexos. Por fim, a estratégia deve falhar no quinto exemplo.

[37] Note que o quarto exemplo é uma matriz de rotação, é intuitivo que essa matriz não tenha autovalores reais.

Vamos ver o que ocorre no ultimo exemplo. Tomamos

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

então

$$p_B(x) = det\left(\begin{bmatrix} x-2 & 1 \\ 0 & x-2 \end{bmatrix}\right) = (x-2)(x-2).$$

Obtemos então um único autovalor, $\lambda = 2$, com multiplicidade algébrica dois. Por outro lado,

$$E_\lambda = N(2I - B) = N\left(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right) = span(e_1)$$

e portanto o autovalor 2 tem multiplicidade geométrica um. Dessa forma, será impossível obter uma base de $\mathbb{R}^2$ formada por autovetores, já que temos um único autovalor e seus autovetores formam um espaço unidimensional. Note que, em todos os exemplos acima, a multiplicidade geométrica foi menor ou igual à multiplicidade algébrica. O seguinte lema diz que o contrário nunca ocorre:

Lema 6. *Seja λ um autovalor de A, a multiplicidade geométrica de λ é sempre menor ou igual que sua multiplicidade algébrica.*

Demonstração. A prova desse lema baseia-se no fato de que o polinômio característico é invariante por mudanças de base. Seja $v_1, \ldots, v_k$ uma base de $N(\lambda I - A)$ e complete a base para que $v_1, \ldots, v_n$ seja uma base de A, onde A é uma matriz $n \times n$. A representação de A na base $v_1, \ldots, v_n$ é uma matriz da forma

$$\begin{bmatrix} \lambda I_k & \star \\ 0_{(k-n)\times k} & \star \end{bmatrix}_{n\times n}$$

onde $\star$ são valores quaisquer. Observe que

$$p_A(x) = det\left(xI - \begin{bmatrix} \lambda I_k & \star \\ 0_{(k-n)\times k} & \star \end{bmatrix}_{n\times n} \right)$$

é um polinômio divisível por $(x-\lambda)^k$. Como k é a multiplicidade geométrica de λ então mostramos que a multiplicidade algébrica de λ é ao menos k. □

O lema anterior restringe um pouco nossas esperanças de conseguir diagonalizar qualquer matriz: como a multiplicidade geométrica é sempre limitada pela multiplicidade algébrica, então se algum autovetor possui multiplicidade geométrica maior que sua multiplicidade algébrica, nunca conseguiremos obter uma base com n autovetores. O seguinte teorema mostra que, de fato, essa condição é necessária e suficiente.

Teorema 8. *Uma matriz A é diagonalizável se, e somente se, para todo autovalor λ de A vale que a multiplicidade algébrica e geométrica de λ são iguais.*

Demonstração. Esse teorema segue da discussão acima e do Exercício 31 abaixo. □

Potência e exponencial de matrizes

Nessa sessão vamos mostrar duas aplicações da diagonalização de matrizes e definir duas operações envolvendo matrizes. Nessa sessão assumimos que A é diagonalizável e pode ser escrita como

$$A = SDS^{-1}$$

para alguma matriz inversível S e uma matriz diagonal

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & \lambda_n \end{bmatrix}.$$

Definição 13. *Se A é uma matriz quadrada e k um inteiro definimos a k-ésima potência de A, denotada por A^k, como sendo:*

- *I, se $k = 0$;*
- *$A \times A \times A \cdots \times A$ (produto k vezes), se $k > 0$;*
- *$A^{-1} \times A^{-1} \times A^{-1} \cdots \times A^{-1}$ (produto $-k$ vezes), se $k < 0$ e A é inversível;*

Também definimos a exponencial de A, denotada e^A, como sendo:

$$e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k = I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \dots.$$

A operação A^k é bem definida, mas a operação e^A envolve um somatório de infinitos termos e poderia não estar bem definida. De fato, ocorre que e^A é uma operação bem definida em geral, mas não discutiremos isso nesse curso. Nesse momento, veremos apenas como a diagonalização pode ser utilizada para realizar essas operações, que serão úteis posteriormente.

Primeiro note que

$$D^k = \begin{bmatrix} \lambda_1^k & 0 & \cdots & 0 \\ 0 & \lambda_2^k & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \lambda_n^k \end{bmatrix} \text{ e } e^D = \begin{bmatrix} e^{\lambda_1} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & e^{\lambda_n} \end{bmatrix}.$$

Se $k > 0$ temos

$$A^k = (SDS^{-1})(SDS^{-1})(SDS^{-1})\dots(SDS^{-1}) = SD^kS^{-1}$$

e, analogamente,

$$A^{-k} = (SD^{-1}S^{-1})(SD^{-1}S^{-1})(SD^{-1}S^{-1})\dots(SD^{-1}S^{-1}) = SD^{-k}S^{-1}.$$

Como a exponencial é uma soma de potências segue que

$$e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k = \sum_{k=0}^{\infty} \frac{1}{k!} (SDS^{-1})^k = S\left(\sum_{k=0}^{\infty} \frac{1}{k!} D^k\right) S^{-1} = Se^DS^{-1}.$$

Ou seja, a diagonalização permite transformar as operações de potência e exponenciação de matrizes nas operações de potência e exponenciação dos autovalores.

Exercícios

Exercício 29. *Dizemos que as matrizes A e B são simultaneamente diagonalizáveis quando uma mesma matriz S faz com que $S^{-1}AS$ e $S^{-1}BS$ sejam diagonais. Sejam A e B diagonalizáveis, mostre que A e B são simultaneamente diagonalizáveis se, e somente se, $AB = BA$.*

Exercício 30. *Seja A uma matriz $n \times n$, mostre que p_A é um polinômio mônico de grau n, ou seja, que $p_A(x) - x^n$ é um polinômio de grau estritamente menor que n.*

Exercício 31. *Seja A uma matriz com autovalores $\lambda_1, \ldots, \lambda_k$ (listados unicamente independentemente da sua multiplicidade algébrica). Sejam $E_{\lambda_1}, \ldots, E_{\lambda_k}$ os autoespaços associados. Mostre que*

$$dim\left(span\left(\bigcup_{i=1}^{k} E_{\lambda_i}\right)\right) = \sum_{i=1}^{k} dim(E_{\lambda_i}).$$

Ou, na linguagem do Exercício 8, a soma

$$E_{\lambda_1} \oplus E_{\lambda_2} \oplus \cdots \oplus E_{\lambda_k}$$

é direta.

Exercício 32. *Mostre que uma matriz quadrada A é inversível se, e somente se, nenhum dos seus autovalores é zero.*

Exercício 33 (Discos de Gershgorin)**.** *Seja $A = (a_{ij})_{n\times n}$ uma matriz cujas entradas tomam valores em $\mathbb{C}$. Defina*

$$R_i = \sum_{j \neq i} |a_{ij}|$$

e seja $D(a_{ii}, R_i)$ o disco fechado de raio R_i centrado em a_{ii}. Mostre que todo autovalor de A pertence a pelo menos um disco $D(a_{ii}, R_i)$.

Dica: olhe para a i-ésima coordenada da igualdade $Ax = \lambda x$. Tome $i \in [n]$ tal que $|x_i| \geq |x_j|$ para todo $j \in [n]$.

Exercício 34. *Seja $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ uma matriz com entradas reais.*

(i) *Mostre que $p_A(x) = x^2 - tr(A)x + det(A)$.*

(ii) *Mostre que se $tr^2(A) > 4\,det(A)$, então A tem dois autovalores reais distintos.*

(iii) *Mostre que se $tr^2(A) < 4\,det(A)$, então A tem dois autovalores complexos.*

(iv) *Mostre que se $tr^2(A) = 4\,det(A)$, então A possui um único autovalor $\lambda = \frac{tr(A)}{2}$ com multiplicidade algébrica dois. Mostre que, nesse caso, A é diagonalizável se, e somente se, $a = d$ e $b = c = 0$.*

Exercício 35. *Uma matriz 3×3 possui autovalores $0, 1$ e 2. Responda as perguntas abaixo ou ache exemplos mostrando que as informações acima não são suficientes:*

1. *Qual o posto de B?*

2. *Qual o determinante de $B^T B$?*

3. *Quais os autovalores de $B^T B$?*

4. *Quais os autovalores de $(B^2 + I)^{-1}$?*

Exercício 36. *Dizemos que uma matriz quadrada P é uma projeção quando $P^2 = P$. Mostre que os únicos autovalores de uma projeção são 0 e 1. Além disso, mostre que é possível construir uma base de autovetores.*

Exercício 37. *Seja A uma matriz diagonalizável e*

$$p_A(x) = a_n x^n + \cdots + a_1 x + a_0$$

seu polinômio característico. Mostre que[38]

$$a_n A^n + \cdots + a_1 A + a_0 = 0.$$

[38] De fato, esse resultado vale para qualquer matriz A. E uma das possíveis demonstrações passa por um aproximar matrizes não diagonalizáveis por matrizes diagonalizáveis e "tomar o limite".

Exercício 38. *Seja n ímpar. Mostre que toda matriz real $n \times n$ possui pelo menos um autovalor real.*

Os exercícios abaixo são mais avançados e podem ser ignorados. Eles estão aqui apenas a critério de curiosidade.

Exercício 39. *Sejam $\lambda_1, \ldots, \lambda_n$ autovalores de A (aqui representados sem unicidade) satisfazendo $|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|$. Sejam também $v_1, \ldots, v_n$ uma base de autovetores de A satisfazendo $Av_i = \lambda_i v_i$ para todo $i \in [n]$. Tome $v_0 \notin span(v_2, \ldots, v_n)$ e defina*

$$v_{k+1} = \frac{Av_k}{\|Av_k\|_2} \ \forall k \geq 0.$$

Argumente que[39]

$$\lim_{k \to \infty} v_k \in span(v_1).$$

[39] Esse exercício não é trivial e depende de noções básicas de análise para ser formalizado. Tente apenas construir um argumento informal.

Exercício 40. *Seja V o espaço vetorial das funções $f : \mathbb{R} \to \mathbb{R}$ infinitamente diferenciáveis (note que esse espaço tem dimensão infinita). Defina também $T : V \to V$ como sendo $T(f) = f'$, ou seja, T é a derivada. Mostre que T é uma transformação linear. Use que*

$$Te^{\alpha x} = \alpha e^{\alpha x} \ \forall \alpha \in \mathbb{R}$$

para mostrar que T possui infinitos autovalores.

O teorema espectral

No capítulo anterior investigamos quando é possível encontrar uma base onde uma matriz A pode ser escrita como uma matriz diagonal. Não fizemos nenhuma exigência sobre a matriz A, ou seja, bastava obter S inversível tal que SAS^{-1} fosse diagonal. Neste capítulo vamos estudar quando é possível obter uma base ortonormal na qual a matriz A é diagonal, em outras palavras, queremos que a matriz S satisfaça $S^{-1} = S^T$.

Trabalhando no corpo dos complexos

Até agora buscamos trabalhar apenas no corpo dos reais, definimos espaços vetoriais sobre $\mathbb{R}$, falamos sobre transformações lineares usando o corpo dos reais e, na maioria dos casos, diagonalizamos matrizes com autovalores e autovetores reais. Mesmo assim, a teoria desenvolvida até agora não depende da escolha do corpo dos reais, as definições de espaço vetorial, base e transformação linear seguem válidas se trocarmos os escalares reais por escalares complexos.

Neste capítulo trabalharemos no corpo dos complexos, já que isso permitirá desenvolver nossa teoria em maior generalidade. Para isso precisamos de atualizar algumas das definições vistas até agora.

Definição 14 (O produto escalar em $\mathbb{C}$). *Se $v, w \in \mathbb{C}^n$, definimos*

$$\langle v, w \rangle = \sum_{i=1}^{n} v_i \bar{w}_i,$$

onde $\bar{z}$ é o conjugado[40] *de z.*

[40] O conjugado de um número complexo é o número complexo de mesmo componente real, mas com o sinal do componente imaginário trocado. Por exemplo, o conjugado de $1 + 8i$ é $1 - 8i$.

Note que o produto escalar em $\mathbb{C}$ concorda com o produto escalar em $\mathbb{R}$, além disso, como $|z|^2 = z\bar{z} \in \mathbb{R}$ para todo $z \in \mathbb{C}$ tem-se que $\langle v, v \rangle \in \mathbb{R}$ para todo $v \in \mathbb{C}^n$. Ou seja, $\langle v, v \rangle$ generaliza também a norma Euclidiana.

Matrizes simétricas ou auto-adjuntas

Definição 15. *Dada uma matriz A com entradas complexas, definimos sua adjunta como sendo $A^* = \bar{A}^T$, ou seja, a transposta do conjuntado entrada* *por entrada de A. Uma matriz A com entradas complexas é dita simétrica ou auto-adjunta quando $A^* = A$.*

As matrizes auto-adjuntas ou mesmo as matrizes normais aparecem frequentemente em aplicações como estatística, grafos ou mecânica quântica. Alguns exemplos são dados nos exercícios. A propriedade de interesse em matrizes auto-adjuntas é que dados $v, w \in \mathbb{C}^n$, se A é auto-adjunta, então

$$\langle Av, w\rangle = \langle v, Aw\rangle.$$

Para ver isso note que

$$\langle Av, w\rangle = (Av)^T\bar{w} = v^TA^T\bar{w} = v^T\overline{\bar{A}^Tw} = \langle v, A^*w\rangle = \langle v, Aw\rangle.$$

Lema 7. *Matrizes auto-adjuntas possuem somente autovalores reais.*

Demonstração. Seja $\lambda \in \mathbb{C}$ e $v \neq 0$ tal que $Av = \lambda v$. Vale que

$$\lambda\langle v, v\rangle = \langle \lambda v, v\rangle = \langle Av, v\rangle = \langle v, Av\rangle = \bar{\lambda}\langle v, v\rangle.$$

Logo, $\bar{\lambda} = \lambda$. Daí, $\lambda \in \mathbb{R}$. □

No caso das matrizes diagonalizáveis sabíamos que possuir autovetores de autovalores distintos eram linearmente independentes. Já para matrizes auto-adjuntas conseguimos obter perpendicularidade!

Lema 8. *Se A é auto-adjunta, então os autovetores de diferentes autovalores são perpendiculares.*

Demonstração. Sejam $\lambda_1 \neq \lambda_2$ e $v_1, v_2 \neq 0$ tais que $Av_1 = \lambda_1 v_1$ e $Av_2 = \lambda_2 v_2$. Tem-se

$$\lambda_1\langle v_1, v_2\rangle = \langle Av_1, v_2\rangle = \langle v_1, Av_2\rangle = \bar{\lambda_2}\langle v_1, v_2\rangle$$

e como $\lambda_1, \lambda_2 \in \mathbb{R}$ a igualdade acima só ocorre se $\langle v_1, v_2\rangle = 0$. □

As duas propriedades acima já permitem, em alguns casos, diagonalizar matrizes tomando S ortonormal[41]:

Lema 9. *Seja A matriz auto-adjunta com entradas complexas. Se todos os autovalores de A possuem multiplicidade algébrica 1, então $A = SDS^*$, onde D é uma matriz diagonal com entradas reais e S é ortonormal.*

Demonstração. Usamos os dois lemas anteriores. Basta tomar $v_1, \ldots, v_n$ os autovetores associados aos autovalores $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$. Sabemos que $\langle v_i, v_j\rangle = 0$ para todo $i \neq j$. Tomando

$$S = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$$

e D a matriz diagonal com diagonal $\lambda_1, \lambda_2, \ldots, \lambda_n$ é fácil ver[42] que $SS^* = S^*S = I$ e que $A = SDS^*$. □

[41] Dizemos que uma matriz S com entradas complexas é ortonormal quando $SS^* = S^*S = I$.

[42] Pense com calma, note que é a construção é a mesma que fizemos no capítulo anterior.

O lema garante apenas que D possui entradas reais, mas não garante S possui entradas reais. Felizmente, quando A é uma matriz com entradas reais é possível tomar S uma matriz com entradas reais. Veremos isso nos exercícios.

O próximo lema vai nos ajudar a generalizar o teorema acima para quaisquer matrizes simétricas. A ideia básica é tentar somar uma pequena perturbação que ainda preserve a propriedade de simetria, mas seja o suficiente para garantir que a matriz perturbada tem apenas autovalores de multiplicidade algébrica um.

Lema 10 (Lema do agito)**.** *Seja A matriz auto-adjunta. É possível obter matrizes A_ε, $\varepsilon > 0$ tais que*

- *A_ε é auto-adjunta para todo $\varepsilon > 0$;*
- *Todos os autovalores de A_ε possuem multiplicidade algébrica um;*
- $\lim_{\varepsilon \to 0} A_\varepsilon = A$.

Demonstração. Mostraremos por indução no tamanho da matriz A. Se A é 1×1 então basta tomar $A_\varepsilon = A$. Assumimos que seja verdade para toda matriz de tamanho $(n-1) \times (n-1)$ e mostraremos para matrizes $n \times n$. Seja v_1 um autovetor de A associado ao autovalor λ. Tome v_1 garantindo que $\langle v_1, v_1 \rangle = 1$. A matriz $A' := A + \varepsilon v_1 \bar{v}_1{}^T$ tem autovetor v_1 associado ao autovalor $\lambda + \varepsilon$. Sejam $v_2, \ldots, v_n$ vetores tais que $v_1, \ldots, v_n$ forme uma base ortonormal de $\mathbb{C}^n$, tomando S a matriz de colunas $v_1, \ldots, v_n$ temos

$$A' = S \begin{bmatrix} \lambda + \varepsilon & C \\ 0 & B \end{bmatrix} S^*$$

para alguma escolha de B, C. Logo:

$$\begin{bmatrix} \lambda + \varepsilon & C \\ 0 & B \end{bmatrix} = S^* A' S \Rightarrow \begin{bmatrix} \lambda + \varepsilon & C \\ 0 & B \end{bmatrix} = \begin{bmatrix} \lambda + \varepsilon & 0 \\ C^* & B^* \end{bmatrix}.$$

Portanto, $C = 0$ e $B = B^*$. O teorema segue aplicando o passo indutivo na matriz B e aproximando-a por B_ε. Tomamos então

$$A_\varepsilon = S \begin{bmatrix} \lambda + \varepsilon & 0 \\ 0 & B_\varepsilon \end{bmatrix} S^*.$$

A primeira e a segunda propriedade segue por construção, para a primeira propriedade observe que

$$\begin{aligned}
\lim_{\varepsilon \to 0} A_\varepsilon &= \lim_{\varepsilon \to 0} S \begin{bmatrix} \lambda + \varepsilon & 0 \\ 0 & B_\varepsilon \end{bmatrix} S^* \\
&= S \begin{bmatrix} \lim_{\varepsilon \to 0}(\lambda + \varepsilon) & 0 \\ 0 & \lim_{\varepsilon \to 0} B_\varepsilon \end{bmatrix} S^* \\
&= S \begin{bmatrix} \lambda & 0 \\ 0 & B \end{bmatrix} S^* \\
&= A.
\end{aligned}$$

□

Teorema 9 (Teorema espectral)**.** *Dada uma matriz auto-adjunta A com entradas complexas, existe uma matriz diagonal D com entradas reais e uma matriz ortonormal S tais que*

$$A = SDS^*.$$

Demonstração. Primeiro usamos o lema do agito para perturbar a matriz A e obter, para todo $\varepsilon > 0$, uma matriz A_ε auto-adjunta e cujos autovalores tem todos multiplicidade um. Seja S_ε e D_ε tais que $A_\varepsilon = S_\varepsilon D_\varepsilon S_\varepsilon^*$ como no Lema 9. Para finalizar a demonstração precisaremos de um argumento de compacidade para mostrar que é possível obter uma S ortogonal como o limite (de certa maneira) das S_ε e, usando que $\lim A_\varepsilon = A$ obter também D tal que $A = SDS^*$. □

Exercises

Exercício 41. *Diagonalize as seguintes matrizes:*

(i) $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \\ 3 & 3 & 0 \end{bmatrix}.$

(ii) $\begin{bmatrix} i & 3i & 1 \\ 3i & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$

(iii) $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}.$

Exercício 42. *Encontre todos os autovetores e autovalores da matriz*

$$A = \begin{bmatrix} 2222 & 2 & 3 & 4 & 5 \\ 2 & 2225 & 6 & 8 & 10 \\ 3 & 6 & 2230 & 12 & 15 \\ 4 & 8 & 12 & 2237 & 20 \\ 5 & 10 & 15 & 20 & 2246 \end{bmatrix}.$$

Exercício 43. *Seja A uma matriz simétrica com entradas reais e assuma que todos os autovalores de A possuem multiplicidade algébrica 1. Mostre que:*

(i) *Se $v \in \mathbb{C}^n$ é autovetor de A associado ao autovalor λ, então $\frac{v+\bar{v}}{2} \in \mathbb{R}^n$ e $\frac{v-\bar{v}}{2i} \in \mathbb{R}^n$ também são autovetores associados à λ.*

(ii) *Conclua que no Lema 9, para A como descrita no enunciado desse exercício, é possível escrever $A = SDS^T$. Onde D é diagonal com entradas reais e S satisfaz $SS^T = S^TS = I$ e tem apenas entradas reais.*

Exercício 44. *Use o exercício anterior para concluir que matrizes reais simétricas podem ser diagonalizadas tomando uma base ortonormal de autovalores reais e autovetores reais.*

Exercício 45. *Dado $p \in [0,1]$, defina*

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

Obtenha os autovalores e autovetores de A.

Exercício 46. *Um grafo finito G é um par (V, E) onde V é um conjunto finito qualquer e E é um conjunto de pares $\{v, v'\}$, $v, v' \in V$. Seja $V = \{v_1, \dots, v_n\}$. Definimos o Laplaciano de um grafo como sendo a matriz $A = (a_{ij})_{n\times n}$ onde*

$$a_{ij} = \begin{cases} -1 \text{ se } \{v_i, v_j\} \in E,\ i \neq j \\ 0 \text{ se } \{v_i, v_j\} \notin E,\ i \neq j \\ -\sum_{k\neq i} a_{ik} \text{ se } i = j. \end{cases}$$

Mostre que o Laplaciano de qualquer grafo é sempre simétrico. Construa um grafo com $|V| = 4$ e $|E| = 6$ e calcule seus autovetores e autovalores.

Exercício 47. *Sejam A, B, S matrizes complexas tais que $A = SBS^*$ e $SS^* = S^*S = I$. Mostre que A é auto-adjunto se, e somente se, B é auto-adjunto.*

Exercício 48. *Seja $\mathcal{M}_{n\times n}(\mathbb{C})$ o conjunto das matrizes complexas de tamanho $n \times n$. Mostre que*

(i) *$\mathcal{M}_{n\times n}(\mathbb{C})$ é um espaço vetorial de dimensão n^2 sobre o corpo $\mathbb{C}$.*

(ii) *$\mathcal{M}_{n\times n}(\mathbb{C})$ é um espaço vetorial de dimensão $2n^2$ sobre o corpo $\mathbb{R}$.*

(iii) *As matrizes auto-adjuntas complexas não são um subespaço vetorial de $\mathcal{M}_{n\times n}(\mathbb{C})$ sobre o corpo $\mathbb{C}$.*

(iv) *As matrizes auto-adjuntas complexas são um subespaço vetorial de $\mathcal{M}_{n\times n}(\mathbb{C})$ sobre o corpo $\mathbb{R}$. Ache a dimensão.*

Parte II

Aplicações

k-means

Objetivo

O problema de agrupamento (*clustering* em inglês) consiste em particionar um conjunto $x_1, x_2, \ldots, x_n$ de pontos em $\mathbb{R}^d$ em k partes[43]. Podemos ver uma partição como uma função

$$\sigma : [n] \to [k]$$

que associa para x_i à $\sigma(i)$-ésima parte.

Uma vez que temos uma partição σ podemos obter o centro de cada grupo (ou *cluster*) fazendo

$$\mu_j = \frac{1}{|\{i : \sigma(i) = j\}|} \sum_{i:\sigma(i)=j} x_i \text{ para todo } j \in [k].$$

Para o problema de k-means queremos encontrar $\sigma^\star$ tal que

$$\sigma^\star \in \underset{\sigma:[n]\to[k]}{\arg\min} \sum_{i=1}^{n} \|x_i - \mu_{\sigma(i)}\|_2^2.$$

[43] Normalmente k é muito menor que n.

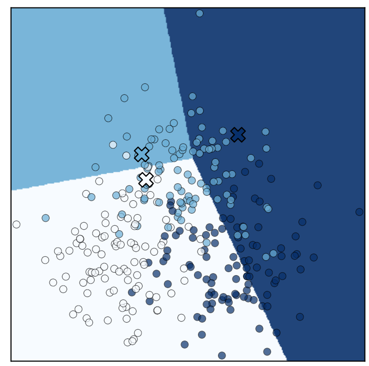

Figura 14: Exemplo de inicialização do *k-means*, no exemplo temos 3 classes e $k = 3$.

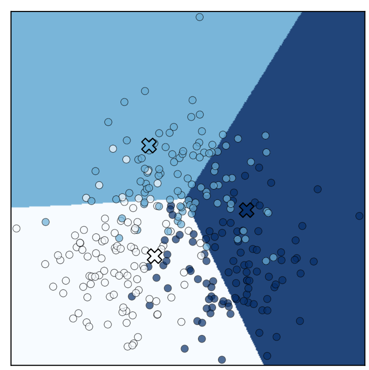

Figura 15: O mesmo exemplo anterior, mas agora após uma iteração com um passo de atribuição e um de atualização de centros.

Algoritmo clássico

O algoritmo clássico do *k-means*, também conhecido como *naive k-means*, é uma abordagem iterativa para encontrar uma solução aproximada para o problema de *clustering*. Começamos fixando um número k de partes. Após a inicialização, o algoritmo alterna dois passos principais: atribuição de pontos aos *clusters* e atualização dos centros dos *clusters*.

(i) **Inicialização**: para inicializar o algoritmo escolhemos algum particionamento σ aleatoriamente (veja exemplo na Figura 14).

(ii) **Atribuição de Pontos aos Clusters**: Dada uma partição atual σ, o primeiro passo é atribuir cada ponto x_i ao *cluster* cujo centro é o mais próximo, ou seja:

$$\sigma(i) = \underset{j\in[k]}{\arg\min} \|x_i - \mu_j\|_2^2.$$

Isso significa que cada ponto é associado ao *cluster* cujo centro é o mais próximo em termos de distância Euclidiana.

(iii) **Atualização dos Centros dos Clusters**: Uma vez que os pontos foram atribuídos aos *clusters*, o próximo passo é recalcular os centros dos *clusters* com base na nova partição. Os novos centros são dados por:

$$\mu_j = \frac{1}{|\{i : \sigma(i) = j\}|} \sum_{i:\sigma(i)=j} x_i \quad \text{para todo } j \in [k].$$

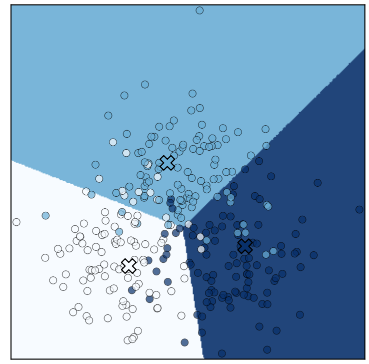

Figura 16: Resultado final após algumas iterações.

Esse processo de atribuição de pontos (i) e atualização dos centros (ii) é repetido iterativamente até que não haja alterações significativas nas atribuições dos pontos ou até que um número máximo de iterações seja atingido. O exemplo da Figura 14 é continuado nas Figuras 15 e 16 e mostra o algoritmo se estabilizando com a passagem das iterações.

A escolha da inicialização dos centros dos clusters pode afetar o desempenho do algoritmo. Uma prática comum é inicializar os centros aleatoriamente a partir dos dados ou utilizando algum método heurístico. O algoritmo pode convergir para um mínimo local, e a escolha do número de clusters k também é crítica. Diversas variações e melhorias foram propostas para lidar com essas questões.

Um exemplo com código

Agora veremos como o *k-means* se comporta em um conjunto de dados não artificial. Os dados que usaremos serão imagens de dígitos $(0, 1, 2, 3, \ldots, 9)$ escritos à mão e digitalizados como imagens em escala de cinza (16 bits) com $8x8$ pixels.

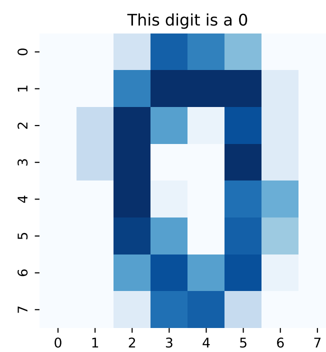

Figura 17: Saída do Código 1.

Podemos então imaginar que essa base de dados é composta de matrizes $8x8$ cujas entradas são inteiros entre 0 e 15. Mas também podemos ver essas matrizes como um único vetor de 64 coordenadas. Nesse caso, a distância entre duas imagens é simplesmente a norma Euclideana da diferença entre essas imagens.

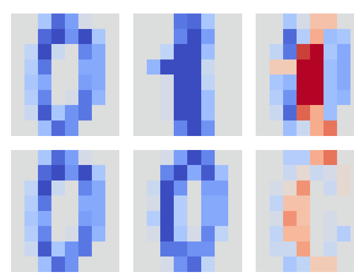

Figura 18: Exemplo de diferença entre dois dígitos. Nas duas linhas a imagem da direita é a diferença das outras duas.

O Código 1 importa as bibliotecas, carrega os dados e permitem visualizar um dígito, que é apresentado na Figura 17. Na Figura 18 é possível observar como a norma Euclideana pode ser utilizada para obter uma distância entre duas representações de dígitos. A diferença entre um dígito zero e o dígito um foi de 59.6, já entre o dígito zero e outro dígito zero foi de 23.7.

```
# Import libs
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import pandas as pd
```

```
from sklearn.datasets import load_digits
from sklearn.cluster import KMeans
from sklearn.decomposition import PCA
from sklearn.preprocessing import OneHotEncoder

# Set a nice figure size
plt.figure(figsize=(4, 4))

# Load data
data, labels = load_digits(return_X_y=True)
(n_samples, n_features), n_digits = data.shape, np.unique(labels).size

# Show a random digit
i = np.random.choice(n_samples)

sns.heatmap(data[i].reshape(8,8), cbar=False, cmap="Blues")
plt.title(f"This digit is a {labels[i]}")
plt.show()
```

Código 1: Importando bibliotecas e visualizando um dígito aleatório

Após executar o *k-means* podemos associar para cada cluster uma classe tomando a classe como a classe mais comum no cluster. O Código 2 faz isso e produz uma matriz relacionando para cada dígito os clusters em que suas representações aparecerem (em proporção)[44].

```
# Apply kmeans
kmeans = KMeans(init="k-means++", n_clusters=n_digits, n_init=4)
kmeans.fit(data)

# Assign cluster labels
encoder = OneHotEncoder(sparse_output=False, categories='auto')
oh_kmeans_clusters = encoder.fit_transform(kmeans.labels_.reshape(-1, 1))
oh_labels = encoder.fit_transform(labels.reshape(-1, 1))
confusion = oh_labels.T @ oh_kmeans_clusters

# Plot digits vs clusters
sns.heatmap(confusion[:,np.argmax(confusion, axis=1)], cbar=False, cmap="Blues")
plt.title("Digits vs clusters")
plt.xlabel("Digit")
plt.ylabel("Cluster")
plt.show()
```

Código 2: Executando o k-means e visualizando digitos e clusters

Por fim, o Código 3 produz a Figura 20 usando o método de redução de dimensionalidade PCA[45].

```
# Apply PCA to data and cluster centers
reduced_data = PCA(n_components=2).fit_transform(
    np.vstack([data, kmeans.cluster_centers_])
)

# Create a scatter plot of the 2D PCA result
data_points = pd.DataFrame({
    "PC1": reduced_data[:-10,0],
    "PC2": reduced_data[:-10,1],
    "Digit": labels
})
```

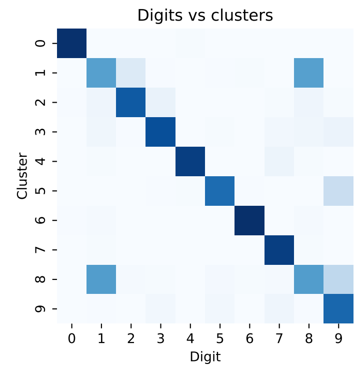

Figura 19: Saída do Código 2.

[44] O que explica que o cluster do dígito 8 tenha uma quantidade significativa de dígitos 1 ou 9? E que existam alguns noves na classe do dígito 5?

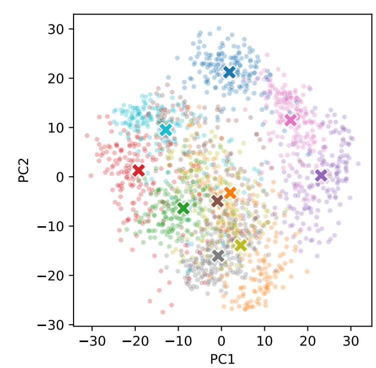

Figura 20: Saída do Código 3.

[45] O PCA busca achar uma transformação linear entre o espaço original, no caso $\mathbb{R}^{64}$, e o espaço final, no caso $\mathbb{R}^2$, que distorça o mínimo possível as distancias no espaço original. Permitindo visualizar os dados.

```python
cluster_centers = pd.DataFrame({
    "PC1": reduced_data[-10:,0],
    "PC2": reduced_data[-10:,1],
    "Digit": np.argmax(confusion, axis=0)
})

sns.scatterplot(data_points, x="PC1", y="PC2", hue="Digit", palette="tab10", marker="o", s=15, legend=False, alpha=.3)
sns.scatterplot(cluster_centers, x="PC1", y="PC2", hue="Digit", palette="tab10", marker="X", s=120, legend=False)
plt.show()
```

Código 3: Visualização em 2 dimensões

Sistemas de equações diferenciais

Neste capítulo, vamos aplicar os conceitos de diagonalização de matrizes para resolver (alguns) sistemas de equações diferenciais. Começamos discutindo o que são equações diferenciais e entendendo o que seriam esses sistemas.

O que é uma equação diferencial?

Estamos acostumados com equações da forma $x + 4 = 7$ ou $(x + 1)^2 = 4$, onde propriedades de um valor desconhecido x nos são dadas e nos permitem obter x ou, ao menos, restringir os possíveis valores de x. Equações diferenciais são equações que descrevem o comportamento de uma função em termos das suas derivadas e variáveis. Por exemplo, podemos nos perguntar "qual é a função que tem a si mesma como derivada?"ou seja, podemos buscar a função $f(x)$ tal que

$$f'(x) = f(x).$$

Nesse caso, as soluções são da forma $f(x) = Ce^x$, para um C qualquer.

Uma equação diferencial pode envolver também derivadas de ordem maior que um. Dizemos que a ordem de uma equação diferencial é a ordem da maior derivada envolvida nessa equação. Por exemplo, a equação do oscilador harmônico é da forma

$$mx''(t) + kx(t) = 0$$

onde $x(t)$ é a posição, m é a massa e k uma constante. Nesse caso, temos uma equação de segunda ordem devido ao termo $x''(t)$.

Também podemos construir sistemas de equações diferenciais. Por exemplo. Um exemplo clássico são os sistemas presa-caçador, seja $\alpha(t)$ é a população de um predador e $\beta(t)$ é a população de uma presa. Suponha que só existem essas duas populações em um certo ambiente. A população de presas deve crescer conforme elas se reproduzem e cair conforme as presas são capturadas. Ou seja,

$$\beta'(t) = r_\beta \beta(t) - c_\beta \beta(t)\alpha(t),$$

onde r_β é a taxa de reprodução das presas e c_β pode ser vista como uma taxa de captura das presas pelos predadores. Analogamente,

$$\alpha'(t) = r_\alpha \alpha(t)\beta(t) - d_\alpha \alpha(t),$$

onde r_α é a taxa de reprodução dos predadores e d_α a taxa de decrescimento da população de predadores.

Nesse cenário, gostaríamos de resolver o sistema

$$\begin{cases} \alpha'(t) = r_\alpha \alpha(t)\beta(t) - d_\alpha \alpha(t) \\ \beta'(t) = r_\beta \beta(t) - c_\beta \beta(t)\alpha(t) \end{cases}.$$

Nesse capítulo, vamos trabalhar especificamente na resolução de sistemas da seguinte forma:

$$\begin{cases} x_1'(t) = a_{11}x_1(t) + \cdots + a_{1n}x_n(t) \\ \vdots \\ x_n'(t) = a_{n1}x_1(t) + \cdots + a_{nn}x_n(t) \end{cases}.$$

Esses sistemas são ditos lineares, homogêneos, de primeira ordem com coeficientes constantes. Podemos representar o sistema anterior como

$$x'(t) = Ax$$

onde $x(t)$ tem coordenadas $x_i(t)$, $i \in [n]$, e $A = (a_{ij})_{n\times n}$.

Problemas de valor inicial

Note que qualquer função da forma $f(x) = Ce^x$ é solução de $f'(x) = f(x)$. Por outro lado, se sabemos que $f(0) = 2$, então a solução é completamente determinada por $f(x) = 2e^x$. De fato, nesse capítulo, vamos buscar soluções de

$$\begin{cases} x'(t) = Ax(t) \\ x(0) = x_0 \end{cases} \tag{8}$$

onde $x_0 \in \mathbb{R}^n$. Nesse caso, sempre existe solução e ela é única[46].

[46] Caso queira mais detalhes, busque pelo Teorema de Picard-Lindelöf

O truque da diagonalização

Se a matriz A em (8) pode ser diagonalizada tomando $A = SDS^{-1}$, então podemos escrever

$$x'(t) = SDS^{-1}x(t)$$

e multiplicando por S^{-1} obtemos

$$S^{-1}x'(t) = SDS^{-1}x(t).$$

Tomando $y(t) = S^{-1}x(t)$ e $y_0 = S^{-1}x_0$ chegamos em

$$\begin{cases} y'(t) = Dy(t) \\ y(0) = y_0 \end{cases}$$

que é uma série de equações da forma $y_i'(t) = \lambda_i y_i(t)$ e com $y_i(0)$ dado pela i-ésima coordenada de y_0. Logo, a solução é

$$y_i(t) = y_i(0)e^{\lambda_i t}\,\forall i \in [n].$$

Em notação matricial temos

$$y(t) = e^{Dt}y_0,$$

multiplicando por S obtemos:

$$x(t) = Se^{Dt}y_0 = Se^{Dt}S^{-1}x_0 = e^{At}x_0.$$

Um exemplo simples

Vamos tentar resolver o seguinte problema de valor inicial:

$$\begin{cases} x'(t) = -3x(t) + 4y(t) \\ y'(t) = -x(t) + 2y(t) \\ x(0) = y(0) = 1 \end{cases}$$

Primeiro reescrevemos como

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix}' = \begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} \text{ e } \begin{bmatrix} x(0) \\ y(0) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Podemos diagonalizar

$$\begin{bmatrix} -3 & 4 \\ -1 & 2 \end{bmatrix} = \underbrace{\begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix}}_{S} \underbrace{\begin{bmatrix} -2 & 0 \\ 0 & 1 \end{bmatrix}}_{D} \underbrace{\begin{bmatrix} \frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{4}{3} \end{bmatrix}}_{S^{-1}}$$

e então calcular:

$$\begin{aligned} \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} &= Se^{Dt}S^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} e^{-2t} & 0 \\ 0 & e^t \end{bmatrix} \begin{bmatrix} \frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{4}{3} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} e^{-2t} & 0 \\ 0 & e^t \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ e^t \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} e^t. \end{aligned}$$

No exemplo acima algo curioso acontece: a condição inicial é dada por um autovetor associado ao autovalor 1 e a solução da equação é dada por um vetor na direção desse mesmo autovalor, mas com uma constante multiplicativa e^t. Esse fato é geral.

Como visto, ao tomar $y(t) = S^{-1}x(t)$ e $y_0 = S^{-1}x_0$ podemos escrever $y'(t) = Dy(t)$ e obter uma solução da forma $y(t) = e^{Dt}y_0$. Dado um vetor v qualquer, o valor de $S^{-1}v$ não passa das coordenadas de v na base dada pelas colunas de S. Sejam

$$S = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \text{ e } y_0 = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix},$$

então $x(t) = Sy(t) = Se^{Dt}y_0$ pode ser reescrita como

$$x(t) = v_1c_1e^{\lambda_1 t} + \cdots + v_nc_ne^{\lambda_n t}. \tag{9}$$

O exemplo acima não passa do caso em que $c_1 = 0$, $c_2 = 1$ e $\lambda_2 = 1$.

O oscilador harmônico

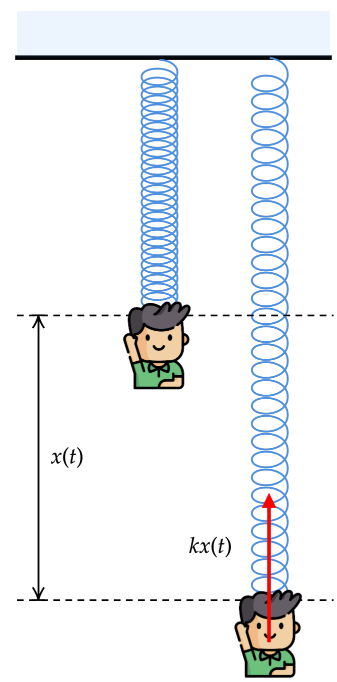

Figura 21: Aluno do IMPA Tech extremamente comprometido com os experimentos de física resolve testar os limites de um sistema massa-mola se dependurando.

Um aluno do IMPA Tech, extremamente comprometido com a disciplina de física experimental, resolve ser ele mesmo a massa em um sistema de massa-mola. E claro, faz isso no espaço para que não haja interferência da força da gravidade. Seja $x(0) = 0$ a posição inicial da mola, ou seja, antes do aluno se conectar à mola. A lei de Hooke diz que a força exercida pela mola no aluno quando a mola tem intensidade $kx(t)$ [47] e vai na direção contrária ao sinal de $x(t)$. Suponha também que ao se conectar na mola e pular o aluno tem uma velocidade inicial $v(0) = 1$. Usando que a aceleração é a segunda derivada da posição e a velocidade é a primeira obtemos

$$\begin{cases} mx''(t) = -kx(t) \\ x(0) = 0 \\ x'(0) = 1 \end{cases}.$$

Esse problema de valor inicial não pode ser escrito diretamente na forma (8), mas podemos tomar $y(t) = x'(t)$ e construir um novo sistema

$$\begin{cases} x'(t) = y(t) \\ y'(t) = -\frac{k}{m}x(t) \\ x(0) = 0 \\ y(0) = 1 \end{cases},$$

[47] Onde k é uma constante positiva relacionada à elasticidade da mola.

que é da forma (8). Ou seja, reduzimos a ordem da equação diferencial aumentando o número de equações no sistema.

Tomando $\alpha = \frac{k}{m} > 0$ podemos escrever:

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix}' = \underbrace{\begin{bmatrix} 0 & 1 \\ -\alpha & 0 \end{bmatrix}}_{A} \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} \text{ e } \begin{bmatrix} x(0) \\ y(0) \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Ao diagonalizar a matriz A obtemos

$$\begin{bmatrix} 0 & 1 \\ -\alpha & 0 \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{i}{\sqrt{\alpha}} & -\frac{i}{\sqrt{\alpha}} \\ 1 & 1 \end{bmatrix}}_{S} \underbrace{\begin{bmatrix} -i\sqrt{\alpha} & 0 \\ 0 & i\sqrt{\alpha} \end{bmatrix}}_{D} \underbrace{\begin{bmatrix} -i\frac{\sqrt{\alpha}}{2} & \frac{1}{2} \\ i\frac{\sqrt{\alpha}}{2} & \frac{1}{2} \end{bmatrix}}_{S^{-1}},$$

ou seja, os autovalores de A são complexos! Por outro lado, sabemos que a solução de $x(t)$ precisa ser um número real, caso contrário nosso querido colega vai acabar vivendo num mundo imaginário... Vamos prosseguir mesmo assim.

A solução deve ser da forma

$$v_1 c_1 e^{\lambda_1 t} + v_2 c_2 e^{\lambda_2 t},$$

onde $\lambda_1 = -i\sqrt{\alpha}$ e $\lambda_2 = i\sqrt{\alpha}$. Além disso,

$$S^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$$

e portanto $c_1 = c_2 = \frac{1}{2}$. Daí,

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} \frac{i}{\sqrt{\alpha}} \\ 1 \end{bmatrix} \frac{e^{-it\sqrt{\alpha}}}{2} + \begin{bmatrix} -\frac{i}{\sqrt{\alpha}} \\ 1 \end{bmatrix} \frac{e^{it\sqrt{\alpha}}}{2}.$$

Nos interessa apenas a posição $x(t)$, que é dada por

$$x(t) = \frac{i}{2\sqrt{\alpha}} e^{-it\sqrt{\alpha}} - \frac{i}{2\sqrt{\alpha}} e^{it\sqrt{\alpha}}.$$

Usando que

$$e^{i\theta} = \cos\theta + i\sin\theta,$$

obtemos

$$\begin{aligned}
x(t) &= \frac{i}{2\sqrt{\alpha}} e^{-it\sqrt{\alpha}} - \frac{i}{2\sqrt{\alpha}} e^{it\sqrt{\alpha}} \\
&= \frac{i}{2\sqrt{\alpha}} \left(\cos(-t\sqrt{\alpha}) + i\sin(-t\sqrt{\alpha})\right) - \frac{i}{2\sqrt{\alpha}} \left(\cos(t\sqrt{\alpha}) + i\sin(t\sqrt{\alpha})\right) \\
&= \frac{i}{2\sqrt{\alpha}} \left(\cos(t\sqrt{\alpha}) - i\sin(t\sqrt{\alpha})\right) - \frac{i}{2\sqrt{\alpha}} \left(\cos(t\sqrt{\alpha}) + i\sin(t\sqrt{\alpha})\right) \\
&= \frac{i}{2\sqrt{\alpha}} \left(-i\sin(t\sqrt{\alpha})\right) - \frac{i}{2\sqrt{\alpha}} \left(i\sin(t\sqrt{\alpha})\right) \\
&= \frac{1}{\sqrt{\alpha}} \sin(t\sqrt{\alpha}).
\end{aligned}$$

Reescrevendo em termos do problema inicial tem-se

$$x(t) = \sqrt{\frac{m}{k}} \sin\left(t\sqrt{\frac{k}{m}}\right).$$

Uma análise similar forneceria

$$y(t) = -\cos\left(t\sqrt{\frac{k}{m}}\right).$$

Curiosamente, os valores obtidos para $x(t)$ são reais para todo $t \in \mathbb{R}$ [48]. Além disso, o par $(x(t), y(t))$ fornece uma solução que evolui girando sobre uma elipse conforme t varia [49]. Comparando esse exemplo com o exemplo anterior conseguimos observar alguns comportamentos interessantes:

- Autovalores negativos correspondem à direções cuja influência na solução vai exponencialmente à zero conforme $t \to \infty$;
- Autovalores positivos correspondem à direções cuja influência cresce exponencialmente;
- Autovalores imaginários correspondem à rotações.

Essas observações são gerais e podem ser mostradas analisando a equação (9).

[48] Isso não é uma coincidência, prove que isso sempre é verdade quando A é uma matriz com entradas reais.

[49] Escolha um valor para $\frac{k}{m}$ e desenhe soluções para alguns valores de t. Interprete essa solução em termos do problema: o que aconteceu com nosso colega extremamente interessado em física?