

# Álgebra Linear

Repositório NewDen

---

## Sumário

<b>1</b>	<b>Conceitos Elementares</b>	<b>2</b>
1.1	Espaços Vetoriais . . . . .	2
1.2	Matrizes . . . . .	4
1.3	Sistemas Lineares . . . . .	7
1.4	Decomposição LU . . . . .	8
<b>2</b>	<b>Explorando Espaços Vetoriais</b>	<b>15</b>
2.1	Bases e Coordenadas . . . . .	15
2.2	Transformações Lineares . . . . .	16
2.3	Produto Interno e Normas . . . . .	18
2.4	Determinante . . . . .	21
<b>3</b>	<b>Ortogonalidade</b>	<b>26</b>
3.1	Vetores Ortogonais . . . . .	26
3.2	Autovalores e Autovetores . . . . .	28
3.3	Teorema Espectral . . . . .	31

---

# 1 Conceitos Elementares

Álgebra Linear é o estudo de coisas que se comportam com linearidade, em especial com o uso de vetores, matrizes, transformações lineares e temas tangentes a isso. De início, o que precisamos é o conhecimento geral sobre esses conceitos e relações entre eles.

## 1.1 Espaços Vetoriais

O nosso principal objeto de estudo na álgebra linear é o **vetor**, e sua definição é: Um vetor é um elemento de um espaço vetorial. Note que essa definição não nos ajuda muito, pois ela se apoia em uma outra definição ainda desconhecida: um espaço vetorial. Por sorte, esse possui uma definição mais convincente.

Um **espaço vetorial** é um espaço  $V$  equipado de duas operações soma e produto por escalar  $(+, \cdot)$ , onde  $+: V \times V \rightarrow V$  e  $\cdot: \mathbb{R} \times V \rightarrow V$  onde valem todas as oito propriedades a seguir:

1. **Comutatividade da soma:**  $u, v \in V \rightarrow u + v = v + u$
2. **Associatividade da soma:**  $u, v, w \in V \rightarrow (u + v) + w = u + (v + w)$
3. **Elemento neutro da soma:**  $\exists 0 \in V$  t.q.  $u \in V \rightarrow u + 0 = u$
4. **Inverso aditivo:**  $u \in V \rightarrow \exists (-u) \in V$  t.q.  $u + (-u) = 0$
5. **Distributividade da soma:**  $u, v \in V, \alpha \in \mathbb{R} \rightarrow \alpha(u + v) = \alpha u + \alpha v$
6. **Distributividade do produto por escalar:**  $u \in V, \alpha, \beta \in \mathbb{R} \rightarrow (\alpha + \beta)u = \alpha u + \beta u$
7. **Associatividade do produto por escalar:**  $u \in V, \alpha, \beta \in \mathbb{R} \rightarrow \alpha(\beta u) = (\alpha\beta)u$
8. **Elemento neutro do produto:**  $u \in V \rightarrow 1 \cdot u = u$

Note que no fundo, essas propriedades significam: um espaço vetorial é um espaço que a soma e produto são bem comportadas e lineares. Além disso, podemos fazer mais algumas definições: Se  $V$  é um espaço tal que  $(V, +, \cdot)$  é um espaço vetorial e  $W \subset V$ , então  $(W, +, \cdot)$  também é um espaço vetorial e é um **subespaço vetorial** de  $V$ .

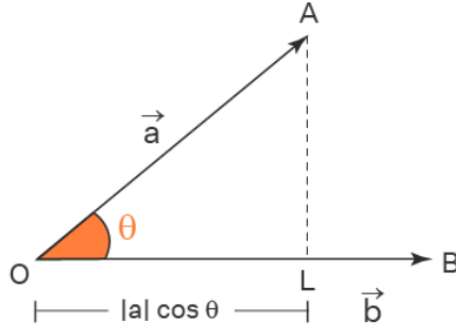
Note que  $\mathbb{R}^n$  é um espaço vetorial. Além disso, se  $v \in \mathbb{R}^n$ , podemos tratá-lo como uma lista de números  $v = (v_1, \dots, v_n)$ , onde a soma e produto são feitos coordenada a coordenada, assim como a igualdade e conceitos similares. Também podemos desenhar um vetor  $v$  no  $\mathbb{R}^n$  como uma seta indo da origem ao ponto de coordenadas  $(v_1, \dots, v_n)$ .

Veremos que essa representação tem um significado mais profundo do que parece, e não só para o  $\mathbb{R}^n$ , mas só um pouco mais a frente.

Podemos definir mais algumas operações muito relevantes para lidar com vetores:

- **Norma euclidiana:**  $\|v\|_2 = \sqrt{v_1^2 + v_2^2 + \dots + v_n^2}$
- **Produto escalar:**  $\langle v, w \rangle = \sum_{i=1}^n v_i w_i$

Esses operadores possui muitas propriedades importantes e/ou interessantes, como  $\langle a, b \rangle = \|a\|_2 \cdot \|b\|_2 \cdot \cos(\theta)$ , onde  $\theta$  é o ângulo entre os dois vetores no  $\mathbb{R}^n$ . Se representarmos os dois vetores  $a, b$  no espaço desenhando o ângulo entre eles, teremos o seguinte desenho:



Nele, veja que podemos aplicar o teorema de Pitágoras e concluir que  $AL^2 = OA^2 - LO^2$  e também que  $AL^2 + LB^2 = BA^2$ . Como  $OA = \|a\|_2$ , podemos usar trigonometria para concluir que  $AL = \|a\|_2 \cdot \sin(\theta)$  e  $LO = \|a\|_2 \cdot \cos(\theta)$ , e então,  $LB = \|b\|_2 - \|a\|_2 \cdot \cos(\theta)$ . Além disso,  $BA = \|b - a\|_2$ . Com isso, podemos aplicar o teorema de Pitágoras do triângulo  $ALB$ , concluindo:

$$\|a\|_2^2 \cdot \sin^2(\theta) + \|b\|_2^2 + \|a\|_2^2 \cdot \cos^2(\theta) - 2\|a\|_2\|b\|_2 \cdot \cos(\theta) = \|b - a\|_2^2$$

Que pode ser simplificada para:

$$\|a\|_2^2 + \|b\|_2^2 - 2\|a\|_2\|b\|_2 \cdot \cos(\theta) = \|b - a\|_2^2$$

Além disso, note que:

$$\|b - a\|_2^2 = \sum_{i=1}^n (b_i - a_i)^2 = \|b\|_2^2 + \|a\|_2^2 - 2\langle a, b \rangle$$

E igualando as duas expressões, podemos concluir que:

$$\langle a, b \rangle = \|a\|_2 \cdot \|b\|_2 \cdot \cos(\theta)$$

Esse desenho também nos diz um pouco sobre a **projeção** de um vetor sobre o outro: o segmento  $LO$  é a projeção de  $a$  sobre  $b$ , que pode ser interpretada como a sombra que um vetor faz sobre o outro, mas teremos mais tempo para explorar isso mais a frente.

Vamos agora falar um pouco mais de subespaços. Sabemos a definição de subespaços, mas como eles se comportam no geral? Vamos pensar em alguns exemplos do que são e o que não são subespaços:

- Seja  $E$  o espaço vetorial das funções reais de uma variável (veja que ele cumpre as propriedades). O subconjunto  $C^k$  das funções  $k$  vezes diferenciáveis é um subespaço vetorial de  $E$ , assim como o subconjunto  $C^\infty$ , das funções infinitamente diferenciáveis.
- Seja  $P$  o espaço dos polinômios de grau menor ou igual a  $m$ . Veja que o espaço dos polinômios de grau exatamente  $m$  não é subespaço dele, pois dois polinômios de grau  $m$  somados podem formar um polinômio de grau menor, que sai do espaço definido.
- Seja  $F$  um espaço vetorial que tem um vetor  $v$ . Então, o conjunto dos múltiplos de  $v$  também é um subespaço vetorial de  $F$ .

Então podemos ver que nem todo exemplo de parte intuitiva de um espaço vetorial forma um subespaço vetorial, mas no geral, temos bons subespaços vetoriais para analisar.

Seja  $E$  um espaço vetorial e  $F_1, F_2$  subespaços vetoriais de  $E$ . Se  $F_1 \cap F_2 = \{0\}$ , representamos a união dos subespaços  $F_1 + F_2$  como  $F_1 \oplus F_2$ , e isso é chamado de **soma direta** de  $F_1$  e  $F_2$ .

Dito isso, dizer que  $F = F_1 \oplus F_2$  é equivalente a dizer que  $\forall w \in F$ , podemos escrever  $w$  de forma única como  $w = v_1 + v_2$ , onde  $v_1 \in F_1$  e  $v_2 \in F_2$ . Para demonstrar isso, note que:

- $F = F_1 \oplus F_2 \rightarrow F_1 \cap F_2 = \{0\}$ . Além disso, suponha que exista  $w = v_1 + v_2 = u_1 + u_2$ , onde  $u_1, v_1 \in F_1$  e  $u_2, v_2 \in F_2$ . Então,  $x = v_1 - u_1 = u_2 - v_2$ . Note que  $u_1 - v_1 \in F_1$  e também  $u_2 - v_2 \in F_2$ , logo,  $x \in F_1 \cap F_2$ , portanto,  $x = 0$ . Isso mostra que só há uma forma de escrevermos  $w$ .
- Se só há uma forma de escrevermos  $w = v_1 + v_2$  para qualquer  $w$ , tome  $w \in F_1 \cap F_2$ . Note que  $w = w + 0 = 0 + w$ , onde  $w, 0 \in F_1$  e  $0, w \in F_2$ . Pela unicidade, devemos ter  $0 = w$ , mostrando então que  $F = F_1 \oplus F_2$  pois  $F_1 \cap F_2 = \{0\}$ .

Note que com isso, podemos aplicar esse resultado para várias coisas, onde uma delas é que dividir o  $\mathbb{R}^n$  em hiperplanos que só se intersectam em um ponto forma uma soma direta. Um exemplo de divisão é dividir ele nas retas que tem uma das coordenadas variável e as restantes fixas iguais a zero: é fácil ver que só há uma forma de construir um vetor como soma dos vetores em outros subespaços, pois cada subespaço é responsável por alterar uma coordenada.

## 1.2 Matrizes

Agora, veremos outro elemento crucial do estudo de vetores: as **matrizes**. Uma matriz é graficamente uma tabela  $N \times M$ , com  $N$  linhas e  $M$  colunas. No geral, uma matriz também pode ser vista como uma junção de  $M$  vetores coluna, e essa interpretação também nos ajuda em muitas contas.

Podemos fazer operações com matrizes: a **soma**, que é feita entrada a entrada correspondente e também a **multiplicação por escalar**, novamente, feita entrada a entrada. Veja que essas operações são análogas as operações vetoriais, justamente pois matrizes e vetores estão fortemente relacionados.

Além dessas operações, também há uma operação exclusiva de matrizes: a **multiplicação de matrizes**. Para duas matrizes  $A \in \mathbb{R}^{N \times M}$ ,  $B \in \mathbb{R}^{M \times Q}$  serem multiplicáveis, devemos ter  $M = P$ , e o resultado será uma matriz  $N \times Q$  onde podemos calcular as entradas da matriz produto  $AB$

com a expressão  $AB_{ij} = \sum_{k=1}^M A_{ik}B_{kj}$ . Segue um exemplo de multiplicação de matrizes (verifique).

$$\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ 2 & 6 \end{bmatrix}$$

Além disso tudo, seja  $X = \{v_1, v_2, \dots, v_n\}$  um conjunto de  $n$  vetores de um espaço vetorial  $V$  e  $Y = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$  um conjunto de escalares. A expressão  $\alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n$  é chamada de **combinação linear** de  $X$ .

Indo além da multiplicação de matrizes por definição explícita, perceba que uma multiplicação de matrizes pode ser interpretada como **várias combinações lineares de vetores**.

Podemos interpretar que o  $i$ -ésimo vetor-coluna na matriz da direita são os coeficientes que devemos multiplicar cada vetor coluna da matriz da esquerda para que a soma desses vetores coluna multiplicados pelos coeficientes seja o  $i$ -ésimo vetor coluna da matriz final.

Da mesma forma, podemos interpretar que a  $i$ -ésima linha da matriz da esquerda representa os coeficientes dos vetores-linha da matriz da direita para que a combinação linear deles seja o  $i$ -ésimo vetor-linha da matriz resultado. Lembre-se muito bem dessas duas interpretações, pois serão muito valiosas mais a frente.

A multiplicação de matrizes é uma operação associativa, mas não é comutativa:  $A(BC) = (AB)C$ , mas  $AB \neq BA$ . Prove a associatividade e dê um contra exemplo para a não comutatividade.

Há ainda uma outra definição que, por ora não será muito usada, mas tem grande relevância no estudo das equações diferenciais: a **exponencial de matriz**. Ela é definida com o mesmo formato da série de Taylor da exponencial, mas cujos termos  $x^k$  são as potências de matrizes  $A^k$ , onde uma potência de matriz é definida como uma sequência de multiplicações de matriz. No geral, a formula da exponencial de matriz é: 
$$e^A = \sum_{n=1}^{\infty} \frac{A^n}{n!}.$$

Há ainda uma lista de definições que importam e serão vistos com muita frequência ao decorrer do documento:

- **Matriz Transposta:**  $A^T$ , sendo a matriz obtida ao trocarmos as linhas de  $A$  por suas colunas, isto é,  $A_{ij} = A_{ji}^T$ . Uma matriz é **simétrica** se  $A = A^T$ .
- **Matriz Conjugada:**  $\overline{A}$ , onde a operação feita é tomar o conjugado complexo de todas as entradas da matriz caso ela seja uma matriz de números complexos, mantendo as dimensões da matriz.
- **Matriz Transposta Conjugada:**  $A^*$ , e essa definição é a combinação das duas definições de cima ao mesmo tempo:  $A^* = \overline{A^T}$ . Uma matriz é **hermitiana** se  $A = A^*$ .
- **Matriz Identidade:**  $I$ . É uma matriz quadrada que possui todas entradas iguais a zero, exceto a diagonal que possui valores 1.
- **Matriz Inversa:**  $A^{-1}$ , sendo essa uma matriz tal que  $AA^{-1} = A^{-1}A = I$ . Uma matriz  $A_E$  é **inversa pela esquerda** de  $A$  se  $A_E A = I$ , mas não necessariamente  $AA_E = I$ . Da mesma forma, uma matriz  $A_D$  é **inversa pela direita** de  $A$  se  $AA_D = I$ , mas não necessariamente  $A_D A = I$ .
- **Traço de uma matriz:**  $tr(A)$ , é a soma dos elementos da diagonal da matriz.

Também podemos lidar com matrizes tratando pedaços delas como **blocos**. Vamos imaginar por exemplo uma matriz  $4 \times 4$ . Ela pode ser interpretada como uma matriz de  $2 \times 2$  blocos de tamanho  $2 \times 2$ , da seguinte forma:

$$\begin{bmatrix} a_{11} & a_{12} & b_{11} & b_{12} \\ a_{21} & a_{22} & b_{21} & b_{22} \\ c_{11} & c_{12} & d_{11} & d_{12} \\ c_{21} & c_{22} & d_{21} & d_{22} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

Esses blocos de matrizes podem ser operados como matrizes menores na multiplicação de matrizes e produzirão os mesmos resultados, desde que as dimensões das multiplicações sejam condizentes.

Vamos pensar de uma forma generalizada para tentar verificar essa propriedade: Suponha que uma matriz  $A$  seja dividida em blocos da seguinte forma:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1m} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2m} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nm} \end{bmatrix} = \begin{bmatrix} B_{11} & B_{12} & \dots & B_{1q} \\ B_{21} & B_{22} & \dots & B_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ B_{p1} & B_{p2} & \dots & B_{pq} \end{bmatrix}$$

Suponha que exista outra matriz de blocos  $C$  e queremos fazer a multiplicação de matrizes  $AC$  mas operando nos blocos, e não entrada a entrada. Para isso ser possível, veja que a multiplicação de linha e coluna de matrizes deve ser possível na escala de blocos e todo produto  $B_{ik}D_{kj}$  deve ser factível. Verificando as necessidades das dimensões das matrizes, concluímos que:

- O número de linhas de blocos na matriz  $C$  deve ser igual ao número de colunas de blocos na matriz  $D$ .
- O número de colunas dos blocos  $B$  na mesma coluna de blocos mas em linhas diferentes deve ser o mesmo.
- O número de linhas dos blocos  $B$  na mesma linha de blocos mas em colunas diferentes deve ser o mesmo.
- O número de linhas dos blocos  $D$  na mesma linha de blocos mas em colunas diferentes deve ser o mesmo.
- O número de colunas dos blocos  $D$  na mesma coluna de blocos mas em linhas diferentes deve ser o mesmo.
- O número de colunas dos blocos  $B$  na coluna de blocos  $k$  deve ser igual ao número de linhas dos blocos  $D$  na linha de blocos  $k$ .

Note que as condições 2 até 5 implicam somente que **os blocos das duas matrizes devem estar alinhados verticalmente e horizontalmente**, e assim também será a matriz resultado. Ao cumprir essas propriedades, é possível multiplicar as matrizes. Agora, vamos olhar para o bloco  $E_{xy}$  da matriz resultado da multiplicação  $AC$ , e mais especificamente para entrada qualquer  $ij$  de  $E$  que pertence ao bloco  $E_{xy}$ . Sabemos que  $E_{xy}$  foi obtida por:

$$\sum_{k=1}^q B_{xk}D_{ky}$$

E portanto, sendo  $\alpha_{xk}$  o número de colunas de  $B_{xk}$ , a entrada  $ij$  de  $E$  em  $E_{xy}$  é dada por:

$$\sum_{k=1}^q (B_{xk}D_{ky})_{ij} = \sum_{k=1}^q \sum_{u=1}^{\alpha_{xk}} (B_{xk})_{iu} \cdot (D_{ky})_{uj}$$

Um detalhe é que, na expressão acima,  $(B_{xk})_{iu}$  não se refere a entrada  $iu$  de  $B_{xk}$ , mas sim à entrada  $i'u$ , onde  $i'$  é o índice da linha de  $B_{xk}$  que corresponde a linha  $i$  da matriz completa  $E$ . O mesmo está sendo considerado para  $j$  em  $D$ .

Dito isso, veja que a expressão acima é completamente equivalente a multiplicar a linha  $i$  de  $A$  e a coluna  $j$  de  $C$ , mostrando que para toda entrada, as duas formas de fazer essa multiplicação são equivalentes, e com isso, concluímos que representar matrizes em blocos é um método válido caso os blocos sejam **alinhados** e de dimensões **compatíveis**.

### 1.3 Sistemas Lineares

Sistemas lineares são conjuntos com equações que usam um determinado conjunto de variáveis, e resolvê-los significa encontrar valores para as variáveis que tornem as equações verdadeiras. Alguns exemplos de sistema linear são  $\{2x + y = 3, x + y = 4\}$ , que possui duas equações e as variáveis  $x$  e  $y$ , e também  $\{x + y + z = 1, 2x + y + z = 4, x + 7y + z = 7, 4x + 4y = 9\}$ , com quatro equações e três variáveis.

Podemos representar sistemas lineares com matrizes: uma matriz contém todos os coeficientes, onde cada linha representa os coeficientes de uma equação. Essa matriz multiplica outra matriz (um vetor, que é uma matriz de uma só coluna) à sua direita: um vetor cujas entradas são as variáveis correspondentes, na mesma ordem dos coeficientes. Esse produto é igualado a outro vetor, representando os resultados das equações.

Por exemplo, os sistemas anteriores podem ser representados assim:

$$\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix} \quad \text{e} \quad \begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & 1 \\ 1 & 7 & 1 \\ 4 & 4 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 7 \\ 9 \end{bmatrix}$$

Essa representação formal permite uma maneira mais eficiente de trabalhar com sistemas. Mas será que ela também nos ajuda a resolvê-los?

No geral, existe sim uma forma generalizada de resolver sistemas lineares a partir da representação por matrizes. Trata-se de um algoritmo chamado **Eliminação Gaussiana**, que permite fazer inferências sobre as soluções de um sistema linear. Um sistema pode ter nenhuma, uma ou infinitas soluções - pois se houver duas, há infinitas combinações lineares entre elas que também são soluções.

O algoritmo consiste em aplicar operações que não alteram as soluções do sistema, com o objetivo de transformar a matriz  $A$  em uma matriz triangular superior, que pode ser resolvida facilmente com retrossubstituição. Uma matriz triangular superior é aquela cujos elementos abaixo da diagonal principal são todos zero.

Um sistema linear pode ser representado como  $Ax = y$ , onde  $A$  é a matriz dos coeficientes,  $x$  o vetor de variáveis e  $y$  o vetor dos resultados, por exemplo:

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

As operações devem ser aplicadas tanto à matriz  $A$  quanto ao vetor  $y$ , mas não ao vetor de variáveis. As operações válidas são:

- $M(i, \alpha)$ : multiplicar a linha  $i$  por um escalar  $\alpha \neq 0$ .
- $P(i, j)$ : trocar as linhas  $i$  e  $j$  de posição.
- $S(i, \alpha, j)$ : somar  $\alpha$  vezes a linha  $j$  à linha  $i$ .

Essas operações preservam o conjunto de soluções do sistema:

- Em  $M(i, \alpha)$ , como multiplicamos ambos os lados da equação por  $\alpha$ , as soluções se mantêm.
- Em  $P(i, j)$ , apenas trocamos a ordem das equações.
- Em  $S(i, \alpha, j)$ , se uma solução resolve as equações  $i$  e  $j$ , também resolve a nova linha  $i$ .

Portanto, nenhuma solução é perdida com essas operações. O algoritmo então consiste em, começando da linha 1 e coluna 1, seguir os seguintes passos:

1. Se todos os valores da coluna forem zero, passe para a próxima coluna.
2. Caso contrário, mova uma linha com um valor diferente de zero para a linha atual, torne o valor 1 multiplicando a linha por um escalar e subtraia múltiplos dessa linha das demais linhas para zerar o restante da coluna.
3. Avance para a próxima linha e coluna, repetindo o processo.

Ao final, obtemos uma matriz triangular superior, e podemos resolver o sistema por **retrossubstituição**. A matriz resultante é chamada de **matriz escalonada**. Os primeiros elementos não nulos de cada linha são chamados de **pivôs**, e suas colunas de **colunas pivô**. O número total de pivôs é chamado de **posto** da matriz, que também é o número de colunas linearmente independentes.

O posto de uma matriz é o número de colunas linearmente independentes de  $A$ .

Dado que a Eliminação Gaussiana transforma um sistema  $Ax = b$  em outro equivalente  $Ux = c$ , podemos estudar as soluções observando  $U$  e  $c$ . As possíveis situações são:

- Se uma linha de  $U$  for composta apenas por zeros e o valor correspondente de  $c$  for diferente de zero, o sistema não tem solução.
- Se uma linha de  $U$  for composta por zeros e o valor de  $c$  também for zero, a equação é redundante.
- Se alguma coluna de  $U$  não tiver pivô (isto é, o número de pivôs for menor que o número de variáveis), então o sistema tem **infinitas soluções**, pois existe uma variável livre.
- Se todas as colunas tiverem pivôs, então o sistema tem uma **única solução**, obtida por retrossubstituição.

Portanto, a Eliminação Gaussiana não apenas nos permite resolver sistemas lineares, mas também entender a estrutura de suas soluções. Logo a frente, veremos que essa abordagem matricial nos dá ferramentas poderosas que vão além da resolução de sistemas, nos levando a resultados importantes sobre a natureza das próprias matrizes.

## 1.4 Decomposição LU

Vimos que o algoritmo de Eliminação Gaussiana, mas não conseguimos obter de fato um método reutilizável: para todo sistema linear, temos que refazer todo o algoritmo, mesmo que mudemos apenas algo simples como o vetor  $b$ . Além disso, esse método não nos dá objetos úteis que sabemos trabalhar em álgebra linear, como vetores ou matrizes. Veremos agora uma variação mais abrangente desse algoritmo que resolve esses problemas.

Podemos interpretar as três operações descritas anteriormente como matrizes, que chamaremos de **matrizes elementares**. Por exemplo, para  $m = n = 5, i = 4, j = 2$ , temos as matrizes:



$$P(i, j) = \begin{bmatrix} 1 & & & \\ & & 1 & \\ & 1 & & \\ & & & 1 \end{bmatrix} \quad M(i, \alpha) = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & \alpha \end{bmatrix} \quad S(i, \alpha, j) = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ \alpha & & & 1 \end{bmatrix}$$

No geral, a estrutura dessas operações como matrizes é:

- A **matriz elementar de permutação**  $E_P(i, j)$  é uma matriz identidade  $n \times n$  com as linhas  $i, j$  trocadas.
- A **matriz elementar de multiplicação**  $E_M(i, \alpha)$  é uma matriz identidade com a entrada  $i$  da diagonal tendo o valor  $\alpha$ .
- A **matriz elementar de soma de linha**  $E_S(i, \alpha, j)$  é uma matriz identidade com a entrada da linha  $i$  coluna  $j$  igual a  $\alpha$ .

Todas essas operações são conceitualmente invertíveis, o que nos induz a pensar que as matrizes também possuem uma inversa. E de fato, esse é o caso:

- A **matriz inversa elementar de permutação**  $L_P(i, j)$  é uma matriz identidade  $n \times n$  com as linhas  $i, j$  trocadas.
- A **matriz inversa elementar de multiplicação**  $L_M(i, \alpha)$  é uma matriz identidade com a entrada  $i$  da diagonal tendo o valor  $1/\alpha$ .
- A **matriz inversa elementar de soma de linha**  $L_S(i, \alpha, j)$  é uma matriz identidade com a entrada da linha  $i$  coluna  $j$  igual a  $-\alpha$ .

Note então que poderíamos agora descrever nosso algoritmo de Eliminação Gaussiana como uma grande multiplicação de matrizes:

$$Ax = b \rightarrow E_k \dots E_2 E_1 Ax = E_k \dots E_2 E_1 b$$

E com isso, podemos descrever nosso algoritmo de decomposição LU. Para cada uma das linhas de  $A$ , sua **posição** é o número de linhas que estão acima dessa linha, e seu índice é uma constante definida como sua posição inicial na matriz  $A$ , sem alterações. O mesmo vale para as colunas, mas com a primeira variação do algoritmo que faremos, seu índice é sempre igual a sua posição.

Considere  $V_E$  como um conjunto ordenado que guardará operações de multiplicação e soma utilizadas, enquanto  $V_P$  guardará as operações de permutação utilizadas, ou seja, toda vez que usarmos uma operação, ela será adicionada em algum dos conjuntos.

Definimos também que as linhas e colunas podem estar **livres** ou **ocupadas**. Inicialmente, todas as linhas e colunas estão livres.

Agora, repetimos o seguinte processo, até encerrarmos o algoritmo:

1. Suponha que estamos na iteração  $i$ . Seja  $j$  o menor índice de uma coluna livre que não tenha todas as suas linhas livres com valor 0. Defina que toda coluna de índice menor que  $j$  agora está ocupada. Se não existir  $j$ , então encerramos o algoritmo.

2. Seja  $l$  o índice da linha livre que possui o valor máximo na coluna  $j$ . Se não existir linha livre, então encerramos o algoritmo. Agora, usamos (ou não) uma operação de permutação para permutar a linha de índice  $l$  com a linha na posição  $i$ .
3. Para todas as outras linhas que possuam um valor na coluna  $j$  diferente de zero, usaremos operações de soma com a linha de índice  $l$  para zerar essas linhas na coluna  $j$ . Após essas operações, nossa coluna  $j$  tem um único pivô, e caso necessário, podemos usar uma operação de multiplicação para torná-lo com valor 1.
4. Após finalizarmos o algoritmo, o  $k$ -ésimo pivô está na linha  $k$ , e portanto, nossa matriz resultante (que chamaremos de  $U$ ) é uma matriz triangular superior. Agora, considere  $P$  inicialmente como uma matriz identidade e para cada uma das operações de permutação  $P_i$  em  $V_P$ , multiplique  $P$  por  $P_i$  pela esquerda, obtendo nossa matriz  $P$  final.
5. Assim, considere a nossa matriz  $U_0 = PA$ . Agora, para todas as operações  $O_i$  de  $V_E$ , considere como  $E_i$  a matriz equivalente a operação  $O_i$ , considerando a posição de cada linha de índice  $k$  como a sua posição em  $U_0$ . Seja  $E = E_q \dots E_2 E_1 I$ . Como todas as matrizes que compõem  $E$  possuem inversa, então  $E$  possui inversa, que chamaremos de  $L = L_1 L_2 \dots L_q$ .
6. Note que como guardamos nossas operações em  $V_E$  considerando os índices das linhas, então para qualquer permutação inicial da matriz  $A$ , aplicar a matriz  $E$  chegará ao menos a uma permutação de diferença da nossa matriz  $U$ , e como nossa matriz de permutação  $P$  consiste das mesmas permutações feitas durante o nosso processo para gerar  $U$ , então  $EPA = U$ . Usando que  $L$  é inversa de  $E$ , podemos ainda multiplicar por  $L$  pela esquerda e obtemos  $PA = LU$ , que é nosso resultado final.
7. Pela nossa estratégia de usar o maior valor como pivô, também garantimos estabilidade numérica, e essa técnica é chamada de **pivoteamento parcial**.

Assim, temos que esse algoritmo prova o seguinte teorema:

**Teorema (Decomposição Lower-Upper):** Seja  $A^{m \times n}$  uma matriz qualquer. Existem matrizes  $P, L, U$  satisfazendo as seguintes propriedades:

- $P^{m \times m}$  é uma matriz de permutação.
- $L^{m \times m}$  é uma matriz triangular inferior com entradas binárias na diagonal.
- $U^{m \times n}$  é uma matriz triangular superior.

E além disso, vale que  $PA = LU$ .

O teorema também funciona caso passarmos a condição de  $L$  ter entradas binárias (0 ou 1) como pivôs para a matriz  $U$ .

A principal diferença da situação descrita para essa é o fato de usar ou não a operação de multiplicação de escalar para normalizar os pivôs de  $U$ .

Agora, vamos explorar um código em Python implementando o algoritmo da decomposição LU com pivoteamento parcial (que é o algoritmo que foi descrito até então), explorando os detalhes e sutilezas por trás dele. Existem outros tipos de pivoteamento para a decomposição LU, como o **pivoteamento total**, mas não exploraremos ele nesse momento.

## Seção 1: Inicialização

---

```
import numpy as np

compare = np.frompyfunc(lambda a,b: np.abs(a-b) < 1e-12, 2, 1)

def LU_Decomposition(A):
    n = len(A); m = len(A[0]); nL = 0
    index = np.arange(n, dtype=int)
    where = np.arange(n, dtype=int)

    P = np.eye(n)
    L = np.eye(n)
    U = A.copy()
    operations = []
```

---

Na seção 1, do código, primeiro importamos a biblioteca NumPy e criamos uma função de comparação para usarmos com ponto flutuante adequadamente.

Também criamos a função principal **LU\_Decomposition**, que tem como parâmetro apenas a matriz  $A$  dada. Com base nessa matriz, extraímos suas dimensões, tendo  $n$  linhas e  $m$  colunas, além de inicializar as matrizes  $P, L, U$ , o array que contém o índice  $index[i]$  da linha que atualmente está na posição  $i$  da matriz  $U$  e uma lista com as operações a serem realizadas, que representará o conjunto ordenado  $V_E$ . Também declaramos o array  $where$ , que representa a operação inversa do array  $index$ :  $index[i] = j \leftrightarrow where[j] = i$ .

Note a existência da variável  $nL$ , que representa a posição da próxima linha livre na matriz  $U$ . Uma linha está ocupada se sua posição é menor que  $nL$ , e o mesmo vale para  $nC$  com as colunas, que será criada na seção 2.

## Seção 2: Iteração nas colunas

---

```
for nC in range(m):
    if np.all(compare(U[nL:, nC], 0)):
        continue
    l = np.argmax(np.abs(U[nL:, nC])) + nL

    if nL != l:
        U[[nL, l]] = U[[l, nL]]
        P[[nL, l]] = P[[l, nL]]
        index[nL], index[l] = index[l], index[nL]
        where[index[nL]], where[index[l]] =
            where[index[l]], where[index[nL]]
```

---

Nesse código, começamos a iterar pelas colunas. Primeiro, asseguramos de que avançamos pelas colunas até o momento que exista uma coluna livre não nula, sendo esse nosso  $j = nC$ .

Também encontramos  $l$  como a posição da linha de valor máximo entre as colunas livres. Se  $l \neq nL$ , veja que precisamos permutar em  $U$ , adicionar em  $V_P$  que equivale a permutar em  $P$  e também atualizar os índices que estão nas posições correspondentes, que é o que fazem as próximas linhas.

### Seção 3: Correção da coluna nC

---

```
alpha = U[nL+1:, nC] / U[nL, nC]
for i, x in enumerate(alpha, start=nL+1):
    operations.append([index[i], x, index[nL]])

for i in range(nL+1, n):
    U[i] -= alpha[i-nL-1]*U[nL]

nL += 1
if (nL == n):
    break
```

---

Essa seção tem o objetivo de zerar a entrada na coluna nC das outras linhas. Para isso, ela guarda os coeficientes das operações no array alpha, guarda as operações feitas em operations e por fim faz a subtração das linhas diretamente, incrementando nL ao acabar.

Ao acabar todo o laço de repetição de nC, veja que a matriz  $U$  já será triangular superior como queríamos. Por consequência, já fizemos todas as permutações necessárias, então a matriz  $P$  também está correta.

Agora, temos que reconstruir a matriz  $L$ :

### Seção 4: Construção de L

---

```
for op in operations:
    i = where[op[0]]
    j = where[op[2]]
    alpha = op[1]
    L[:, j] += alpha * L[:, i]

return P, L, U
```

---

Nesse cenário, efetuamos as multiplicações de matrizes pela direita para reconstruir  $L$ , pois  $E_k \dots E_2 E_1 L_1 L_2 \dots L_k = I$ . Note que as matrizes elementares  $L$  são apenas uma soma de múltiplo de uma linha, então multiplicar diretamente as matrizes com o operador @ não é a forma mais eficiente de fazer isso.

Por isso, fizemos a operação equivalente de forma manual: somar  $\alpha$  vezes a **coluna**  $j$  à **coluna**  $i$ , e veja que estamos operando na coluna pois por mais que a representação original seja nas linhas, estamos fazendo agora apenas multiplicações de matrizes **pela direita**, que operam nas colunas.

Observe o uso do vetor where: como as linhas foram permutadas, as posições que elas estavam no momento que criamos a operação não importam, só queremos saber quais as linhas com certeza tem o valor que precisamos para a operação.

Como isso foi decidido com o vetor de índices e where é o inverso dele, conseguimos recuperar essa informação obtendo a posição atual do índice final guardado.

Observe também que a operação elementar de multiplicação por escalar **não é necessária** para a decomposição LU nesse caso: como já vimos, ela pode ser feita apenas se desejarmos que  $U$  tenha pivôs binários ao invés de  $L$ .

Por fim, o código apenas retorna as matrizes obtidas  $P, L, U$ . Discussões sobre estabilidade e complexidade do código não cabem a essa parte do documento: possivelmente podem ser abordadas em álgebra linear numérica, se necessário.

Além de tudo isso, usando a decomposição LU, podemos obter resultados que nos dizem quando matrizes possuem ou não inversa por algum dos lados, relacionando as partes da decomposição com as condições de inversão de matrizes:

**Teorema (Inversão de Matrizes):** Seja  $A^{m \times n}$  uma matriz qualquer e  $P, L, U$  uma decomposição LU dela. As seguintes afirmações são verdadeiras:

- $A$  tem inversa pela esquerda se, e somente se,  $U$  tem inversa pela esquerda.
- $A$  tem inversa pela direita se, e somente se,  $U$  tem inversa pela direita.
- Se  $n > m$ , então  $U$  não tem inversa pela esquerda.
- Se  $m > n$ , então  $U$  não tem inversa pela direita.
- Se  $n \geq m$  e a diagonal de  $U$  não tem entradas nulas, então  $U$  tem inversa pela esquerda.
- Se  $m \geq n$  e a diagonal de  $U$  não tem entradas nulas, então  $U$  tem inversa pela direita.

#### Demonstração:

Para o item 1, suponha que  $A$  tem inversa pela esquerda. Como  $A = P^{-1}LU$ , isso equivale a  $A^{-1}A = A^{-1}P^{-1}LU$ , ou seja,  $A^{-1}P^{-1}L$  é inversa de  $U$  pela esquerda. Agora, suponha que  $U$  tenha pela inversa pela esquerda. Então,  $L^{-1}PA = U \rightarrow U^{-1}L^{-1}PA = I$ , que implica que  $U^{-1}L^{-1}P$  é inversa de  $A$  pela esquerda. A demonstração do item 2 segue de forma análoga.

Para o item 3, se  $n > m$ , significa que há mais linhas do que colunas, e faremos agora a eliminação Gaussiana de  $A$ . Isso força que exista uma linha (suponha que seja a linha de índice  $i$ ) de zeros em  $U$ . Assim, para qualquer matriz  $V$  que multiplicar  $U$  pela direita, o produto  $UV$  terá apenas entradas 0 na linha  $i$ , ou seja, é impossível  $UV$  ser identidade, que implica ser impossível existir inversa pela direita para  $U$ .

A demonstração do item 4 segue similarmente, mas fazendo a eliminação gaussiana de  $A^T$ . Note que isso implicaria ser impossível existir inversa pela direita de  $A^T$ , mas como  $A^T B = (B^T A)^T$ , se existir uma matriz qualquer  $B^T$  inversa pela esquerda de  $A$ , então como  $A^T B = (B^T A)^T$ ,  $A^T$  tem inversa pela direita, que concluímos ser absurdo, portanto, se  $n > m$ ,  $A$  não tem inversa pela esquerda, e pelo item 1,  $U$  também não tem.

Para o item 5, veja que como  $A$  tem  $n$  pivôs, suponha que todos eles sejam pivôs de valor 1, já que nosso algoritmo consegue tranquilamente gerar uma matriz  $U$  dessa forma. Então, sendo  $v_1 = [v_{11}, v_{12}, \dots, v_{1m}]_{1 \times m}$ , sabemos que  $v_1 U = [\beta_1, \beta_2, \dots, \beta_n]_{1 \times n}$ , e por  $U$  ser triangular superior com posto completo, sabemos que o produto para obter  $\beta_i$  é uma equação com  $i$  variáveis (cada variável sendo um elemento de  $v_1$ ). Portanto, podemos definir  $\beta_1 = 1$ ,  $\beta_i = 0$  com  $i > 1$  e usar substituição retroativa para resolver o sistema, conseguindo no fim obter  $v_1$  tal que  $v_1 U = [1, 0, \dots, 0]_{1 \times n}$ .

Agora, faremos indução: Suponha que exista uma matriz  $V_{k \times m}$  tal que  $VU$  seja equivalente as primeiras  $k$  linhas da matriz  $I$ . Então, adicione a  $V$  uma nova linha  $v_{k+1}$ . Agora, a  $k + 1$ -ésima linha de  $VU$  é igual a  $v_{k+1}U = [\beta_1, \beta_2, \dots, \beta_n]_{1 \times n}$ .

Novamente, sabemos que o produto para obter  $\beta_i$  é uma equação com ao menos 1 e no máximo  $i$  variáveis, e então, podemos definir  $\beta_{k+1} = 1$  com  $i \neq k + 1$  e usar substituição retroativa para resolver o sistema, obtendo que a  $k + 1$ -ésima linha de  $VU$  é igual a  $k + 1$ -ésima linha de  $I$ .

Com base nisso, concluímos que se  $U$  tem  $n$  pivôs e  $n \leq m$ , então  $U$  possui inversa pela esquerda. A demonstração do item 6 é bastante similar, onde apenas vamos construindo coluna a coluna ao invés de linha a linha.

E com isso, conseguimos provar um resultado muito interessante: Nosso teorema diz que para termos inversa pela direita e inversa pela esquerda simultaneamente, devemos ter  $n \leq m$  e  $m \leq n$ , além da decomposição  $LU$  possuir posto cheio. Ou seja, a única forma de termos inversa pela esquerda e direita simultaneamente é se  $A$  for uma matriz quadrada de colunas linearmente independentes. Além disso: se  $A$  possuir inversa pela esquerda  $E$  e pela direita  $D$ , então  $D = (EA)D = EAD = E(AD) = E$ , logo,  $E = D$ . Ou seja, apenas matrizes quadradas podem possuir inversa pelos dois lados e ela é a mesma para os dois lados.

## 2 Explorando Espaços Vetoriais

A essência da álgebra linear é trabalhar com espaços vetoriais, matrizes e resolver sistemas lineares. Veremos com o tempo que no fim, tudo se resume a problema relacionados a esses conceitos. Ainda assim, dados os conceitos elementares, podemos explorar outras definições e verificar muitas coisas que nos ajudarão no nosso estudo de álgebra linear.

### 2.1 Bases e Coordenadas

Interpretar matrizes como combinações lineares é algo muito poderoso justamente pois as combinações lineares tem uma grande importância na álgebra linear, tanto que o conjunto de todas as combinações lineares de um conjunto de vetores  $X$  tem nome: ele é chamado de **span** de  $X$ , e é denotado por:

$$\text{span}(X) = \langle v_1, v_2, \dots, v_n \rangle = \left\{ \sum_{i=1}^n \alpha_i v_i : \alpha_i \in \mathbb{R} \right\}$$

Veja que a notação central parece com a notação de produto interno, mas não confunda os dois: tente analisar pelo contexto se estiver em dúvida. O span de um conjunto de vetores de um espaço vetorial  $V$  é um subespaço vetorial de  $V$ . No entanto, em certos casos, alguns vetores de um conjunto podem ser desnecessários para gerar o span.

Por exemplo, o span de  $\{(1, 0), (0, 1), (1, 1)\}$  gera  $\mathbb{R}^2$ , mas apenas dois dos três vetores já são suficientes para isso. Nesse caso, dizemos que há **dependência linear** entre os vetores.

Mais precisamente, um conjunto  $X = \{v_1, v_2, \dots, v_n\}$  é dito **linearmente dependente** se algum vetor pode ser escrito como combinação linear dos demais. Caso contrário, o conjunto é **linearmente independente** (LI).

Um conjunto é LI se, e somente se, a equação  $\sum_{i=1}^n \alpha_i v_i = 0$  possui como única solução  $\alpha_i = 0$  para todo  $i$ . Suponha uma solução diferente: considere um vetor com coeficiente não nulo, normalize-o para ter coeficiente  $-1$  e mova-o para o outro lado da equação. Isso implicaria que um vetor é combinação linear dos outros, o que contradiz a independência.

Um fato interessante é que, se um vetor pode ser escrito como combinação linear de outros vetores com todos os coeficientes não nulos, então qualquer um dos vetores **com coeficientes não nulos** pode ser escrito como combinação linear dos demais **que também tenham coeficientes não nulos** nessa combinação linear. Com isso, surge a ideia de remover o máximo de vetores possível mantendo o mesmo span. O conjunto resultante ao chegar em um conjunto linearmente independente é chamado de **base**.

Uma **base** de um espaço vetorial  $V$  é um conjunto de vetores  $X = \{v_1, v_2, \dots, v_n\}$  tal que  $X$  é linearmente independente e  $\text{span}(X) = V$ .

Todo espaço vetorial gerado por um número finito de vetores possui uma base, e todas as bases têm o mesmo número de elementos. Para ver isso, considere um conjunto  $V = \{v_1, v_2, \dots, v_n\}$ : se ele for LI, é base. Caso contrário, remova vetores dependentes até obter um subconjunto LI.

Suponha que  $A = \{a_1, a_2, \dots, a_n\}$  e  $B = \{b_1, b_2, \dots, b_m\}$  sejam duas bases de um mesmo espaço vetorial  $V$ , com  $n < m$ . Então, cada  $b_k \in B$  pode ser escrito como:

$$b_k = \sum_{j=1}^n \beta_{jk} a_j$$

Logo, uma combinação linear qualquer dos vetores de  $B$  pode ser escrita como:

$$\sum_{k=1}^m \alpha_k b_k = \sum_{k=1}^m \alpha_k \left( \sum_{j=1}^n \beta_{jk} a_j \right) = \sum_{j=1}^n \left( \sum_{k=1}^m \alpha_k \beta_{jk} \right) a_j$$

Se essa soma for nula, veja que por ser uma combinação linear de vetores linearmente independentes, deveríamos ter exclusivamente  $\alpha_i = 0 \forall i$ . Porém, também vale que os  $n$  somatórios que somam termos  $\alpha_k \beta_{jk}$  forem todos nulos, isso também vale. Como isso é um sistema linear com  $n$  equações e  $m$  variáveis  $\alpha_i$  e  $n < m$ , ele possui infinitas soluções, mostrando então que na verdade a combinação linear de vetores de  $B$  não é  $LI$ , então  $B$  não é uma base.

Todas as bases então tem o mesmo tamanho, e esse número é chamado de **dimensão** do espaço vetorial. Em particular, o espaço  $\mathbb{R}^n$  tem dimensão  $n$  e uma de suas bases mais importantes é a **base canônica**, cujos vetores são as colunas da matriz identidade  $n \times n$ .

Além disso, qualquer conjunto linearmente independente pode ser completado até formar uma base. Se o conjunto ainda não gera o espaço, existe algum vetor fora do span. Adicione-o. Repetindo esse processo, eventualmente o conjunto atingirá tamanho  $n$  e será uma base. Caso contrário, existirão bases de tamanhos diferentes, o que é impossível.

Seja  $W = \{v_1, v_2, \dots, v_n\}$  uma base de um espaço vetorial  $V$ . Para todo vetor  $u \in V$ , existem escalares  $\alpha_1, \alpha_2, \dots, \alpha_n$  tais que:

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

Esses escalares são chamados de **coordenadas** de  $u$  na base  $W$ , e denotamos por:

$$u = [\alpha_1, \alpha_2, \dots, \alpha_n]_W$$

As coordenadas de um vetor são únicas. Se existissem dois conjuntos distintos de coordenadas que representassem  $u$ , então sua diferença geraria uma combinação linear não trivial que anula  $u$ , o que contradiz o fato de que  $W$  é uma base (LI).

Dada uma base  $V = \{v_1, \dots, v_n\}$  e outra base  $W = \{w_1, \dots, w_n\}$ , existe uma **matriz de mudança de base** de  $V$  para  $W$ . Essa matriz  $A = [v_1]_W \dots [v_n]_W$  tem como colunas as coordenadas de cada vetor  $v_i$  da base  $V$  na base  $W$ . Para qualquer vetor com coordenadas  $p$  na base  $V$ , o vetor  $Ap$  fornece suas coordenadas na base  $W$ .

No geral, se quisermos mudar de uma base  $B$  para uma base  $A$ , basta multiplicarmos pela esquerda por uma matriz que descreve  $B$  na base  $A$ .

## 2.2 Transformações Lineares

Um conceito com uma importância extrema em álgebra linear é o de transformações lineares, que nos permite ter uma nova interpretação sobre conceitos que já vimos anteriormente. Dizemos que uma transformação linear  $T : U \rightarrow V$  é uma função que recebe como entrada vetores de um espaço vetorial  $U$  e leva-os em vetores de um espaço vetorial  $V$ , e além disso, vale a seguinte propriedade:



- $T(\alpha u + v) = \alpha T(u) + T(v)$

Se definirmos os valores dessa transformação para os vetores que formam uma base de  $U$ , então ela está definida para todo o espaço de  $U$ , e isso foi ganho pela propriedade de ser linear:

$$\bullet \quad T(v) = T\left(\sum_{i=1}^n \alpha_i u_i\right) = \sum_{i=1}^n \alpha_i T(u_i)$$

Transformações assim podem ser vistas como multiplicação por matriz pela esquerda (ao fixar uma base), onde uma transformação que leva de um espaço vetorial de dimensão  $m$  para um de dimensão  $n$  é correspondente a uma matriz  $m \times n$ . Para encontrarmos a matriz, basta escolhermos uma base  $u_1, u_2, \dots, u_m$  de  $U$ , achar a imagem dessa base  $w_1 = T(u_1), w_2 = T(u_2), \dots, w_m = T(u_m)$  e considerar a seguinte matriz de vetores coluna como a matriz da transformação linear em relação a base  $U$ :

$$A = \begin{bmatrix} | & | & \dots & | \\ w_1 & w_2 & \dots & w_m \\ | & | & \dots & | \end{bmatrix}_{u_1, u_2, \dots, u_m}$$

Algumas vezes, denotamos por  $\mathcal{L}(E, F)$  o conjunto das transformações lineares de  $E \rightarrow F$ .

As transformações lineares em  $\mathcal{L}(E, E) = \mathcal{L}(E)$  são chamadas de **operadores lineares** do espaço  $E$ , enquanto transformações lineares  $\phi : E \rightarrow \mathbb{R}$  são chamadas de **funcionais lineares** de  $E$ . Escrevemos  $\mathcal{L}(E, \mathbb{R})$  como  $E^*$ , e chamamos esse de **espaço vetorial dual** de  $E$ .

Além de todos os conceitos vistos, existe um teorema capaz de relacionar as dimensões do núcleo e da imagem de uma transformação linear. Considere agora que estamos lidando com transformações lineares  $T : U \rightarrow V$ .

O **núcleo** de uma transformação linear definido como  $N(T) = \{v \in U \mid T(v) = 0\}$ , isto é, o conjunto dos vetores que são levados no vetor zero. A **imagem**, sendo mais intuitiva, é definida como  $Im(T) = \{v \in V \mid \exists u \text{ t.q. } T(u) = v\}$ , que é o conjunto de vetores alcançáveis no contradomínio.

**Teorema (Núcleo e Imagem):** Seja  $T : U \rightarrow V$  uma transformação linear onde as dimensões de  $U$  e  $V$  são finitas. Então, sendo  $N(T)$  e  $Im(T)$  o núcleo e a imagem de  $T$  respectivamente, vale que:

$$\bullet \quad Dim(N(T)) + Dim(Im(T)) = Dim(U)$$

**Demonstração:**

Seja  $u_1, u_2, \dots, u_r$  uma base de  $N(T)$ . Enquanto existir um vetor em  $U$  ainda não gerado por esses vetores, podemos sempre completar essa base com outros vetores linearmente independentes aos anteriores de modo que  $u_{r+1}, u_{r+2}, \dots, u_n$  agora seja uma base de  $U$ . Assim, basta provarmos que os vetores  $T(u_{r+1}), \dots, T(u_n)$  são uma base de  $Im(T)$ .

Se  $v \in Im(T)$ , então  $\exists u \in U$  onde  $T(u) = v$ . Sabemos que  $u_1, \dots, u_n$  é uma base de  $U$ , logo, escrevemos  $v$  como:

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

Para simplificar a notação, considere somatórios de índice  $i$  sem limites com limites de de 1 até  $n$ , de índice  $j$  com limites de 1 até  $r$  e de índice  $k$  com limites de  $r+1$  até  $n$ .

Veja que  $T(u) = \sum \alpha_i T(u_i) = \sum \alpha_k T(u_k)$ , pois  $\sum \alpha_j T(u_j) = \sum \alpha_j \cdot 0 = 0$ . Portanto, a imagem dos vetores adicionados é suficientes para gerar  $Im(T)$ , bastando verificar se  $T(u_{r+1}), \dots, T(u_n)$  são vetores linearmente independentes.

Se  $\sum \beta_k T(u_k) = 0$ , então temos que  $T(\sum \beta_k u_k) = 0$ , logo,  $\sum \beta_k u_k \in N(T)$ . Note que se não for verdade que  $\beta_k = 0 \ \forall k$ , então o vetor  $\sum \beta_k u_k$  é gerado pelos vetores  $u_1, \dots, u_r$ , mas isso é um absurdo pois isso é uma combinação linear de vetores linearmente independentes à base  $u_1, \dots, u_r$ .

Veja que, então, sabemos que  $\sum \beta_k u_k = 0 \rightarrow \beta_k = 0 \ \forall k$ , que implica que então,  $T(u_{r+1}), \dots, T(u_n)$  são vetores linearmente independentes, e portanto, uma base de  $Im(T)$ . Isso mostra que  $Dim(N(T)) + Dim(Im(T)) = r + (n - r) = n = Dim(U)$ .

Veja que podemos interpretar esse teorema como uma conservação de dimensões: Se temos inicialmente  $n$  dimensões e perdemos  $r$  para o núcleo, pois todas foram levadas em zero, então nos restam  $n - r$  dimensões na imagem. O teorema do núcleo e imagem também é conhecido por outros nomes, como **teorema do posto e nulidade**, pois ele associa a dimensão da imagem da matriz ao posto dela, enquanto a nulidade é o espaço equivalente ao núcleo.

O posto é também a quantidade de pivôs da matriz na sua forma escalonada: se houvessem mais pivôs, temos um absurdo pelas colunas-pivô serem linearmente independentes, e se houvesse menos, então conseguiríamos reconstruir a matriz fazendo as operações inversas nas colunas-pivô, mostrando que o posto na verdade seria menor.

## 2.3 Produto Interno e Normas

Até aqui vimos sobre o produto escalar. Agora, vamos estar retomando esse conceito, chamando agora de **produto interno**, que generaliza o produto escalar usual.

Até aqui, lidamos somente com os números reais, mas muitos desses conceitos podem ser generalizados para os números complexos ou outros corpos, então a partir de agora, lidaremos tanto com reais quanto com complexos, denotando como  $\mathbb{F}$  um **corpo** (estrutura com operações de adição e multiplicação bem definidas) arbitrário.

Seja  $\langle x, y \rangle : V \times V \rightarrow \mathbb{F}$  uma operação e  $V$  um espaço vetorial sobre o corpo  $\mathbb{F}$ . Chamamos  $\langle x, y \rangle$  de **produto interno** se ele satisfaz as seguintes propriedades:

- $\langle \alpha u + w, v \rangle = \alpha \langle u, v \rangle + \langle w, v \rangle$  (**Linearidade no primeiro argumento**)
- $\langle u, v \rangle = \overline{\langle v, u \rangle}$  (**Simetria conjugada**)
- $\langle v, v \rangle \geq 0 \ \forall v; \ \langle v, v \rangle = 0 \iff v = 0$  (**Positividade definida**)

Uma propriedade muito usada que pode ser concluída a partir dessas é a antilinearidade do segundo argumento:  $\langle v, \alpha w \rangle = \overline{\alpha} \langle v, w \rangle$  (Prove!).

E juntamente com ele, vamos definir uma **norma**: uma medida de comprimento de um vetor. Seja  $\|x\| : V \rightarrow \mathbb{F}$  uma operação, onde  $V$  é um espaço vetorial sobre o corpo  $\mathbb{F}$ . Dizemos que  $\|x\|$  é uma **norma** se, para todo  $u, v \in V$  e todo escalar  $\alpha$ , vale que:

- $\|v\| > 0$  se  $v \neq 0$ ;  $\|v\| = 0 \iff v = 0$
- $\|\alpha v\| = |\alpha| \|v\|$
- $\|u + v\| \leq \|u\| + \|v\|$

Entre as normas vetoriais, destacam-se as **p-normas**, definidas em  $\mathbb{R}^n$  por

$$\|x\|_p = \left( \sum_{i=1}^n |x_i|^p \right)^{1/p}$$

e a  **$\infty$ -norma**

$$\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|.$$

Observe que a norma que usamos até agora é a **p-norma 2**. Agora, como essas duas definições que demos se relacionam?

Bom, podemos verificar alguns fatos importantes sobre elas, mas primeiro, precisamos relembrar algumas coisas de outras áreas, mas que vamos usar aqui.

Primeiramente, precisamos retomar o conceito dos números complexos, juntamente a algumas de suas propriedades que usaremos para as demonstrações que seguem.

Os números complexos são os números que estendem os reais:  $z = a + bi$ , com  $a, b \in \mathbb{R}$  e  $i^2 = -1$ , onde  $i$  é a **unidade imaginária**.

Apesar com o objetivo de relembrar, veremos agora o conjunto das principais propriedades que devemos saber dos números complexos:

- O conjugado de um número complexo  $z = a + bi$  é dado por  $\bar{z} = a - bi$
- A parte real de um complexo  $z = a + bi$  é descrita como  $Re(z) = a = z + \bar{z}$
- O módulo de um complexo é dado por  $|z| = \sqrt{a^2 + b^2}$
- $|z| = |-z| = |\bar{z}| = |-\bar{z}|$
- $z \cdot \bar{z} = |z|^2 = |\bar{z}|^2 = |z|^2 = |\bar{z}|^2$
- $|z_1 z_2| = |z_1| \cdot |z_2|$
- $|z_1| - |z_2| \leq |z_1 + z_2| \leq |z_1| + |z_2|$
- $|z^{-1}| = |z|^{-1}$  se  $z \neq 0$
- $\left| \frac{z_1}{z_2} \right| = \frac{|z_1|}{|z_2|}$  se  $z_2 \neq 0$
- Multiplicar  $z$  por  $i$  rotaciona  $z$  em  $90^\circ$  em sentido anti horário (usual).

E com base nelas, vamos demonstrar agora um teorema que também é importante para o que vamos usar a seguir:

**Teorema (Cauchy–Schwarz).** Seja  $V$  um espaço vetorial sobre  $\mathbb{F}$  com produto interno  $\langle \cdot, \cdot \rangle$ . Então, para todo  $u, v \in V$ :

$$|\langle u, v \rangle| \leq \sqrt{\langle u, u \rangle} \sqrt{\langle v, v \rangle}.$$

**Demonstração:**

Pela positividade, veja que a expressão  $\left\langle u - t \frac{\langle u, v \rangle}{|\langle u, v \rangle|} v, u - t \frac{\langle u, v \rangle}{|\langle u, v \rangle|} v \right\rangle$  é não negativa.

Usando a linearidade da primeira coordenada e a antilinearidade da segunda:

$$\langle u, u \rangle - t \frac{\overline{\langle u, v \rangle}}{|\langle u, v \rangle|} \langle v, u \rangle - \bar{t} \frac{\langle u, v \rangle}{|\langle u, v \rangle|} \langle u, v \rangle + t\bar{t} \frac{\langle u, v \rangle \overline{\langle u, v \rangle}}{|\langle u, v \rangle|^2} \langle v, v \rangle$$

Que por sua vez equivale a:

$$\langle u, u \rangle - t|\langle u, v \rangle| - \bar{t}|\langle u, v \rangle| + |t|^2 \langle v, v \rangle$$

Note que ao escolhermos  $t$  como um real qualquer, temos:

$$t^2 \langle v, v \rangle - 2t|\langle u, v \rangle| - \langle u, u \rangle$$

E como essa é uma expressão quadrática em  $t$  que sempre é não negativa, devemos ter  $\Delta \leq 0$ :

$$4|\langle u, v \rangle|^2 - 4\langle v, v \rangle \langle u, u \rangle \leq 0$$

Que equivale a expressão:

$$|\langle u, v \rangle|^2 \leq \langle v, v \rangle \langle u, u \rangle$$

E como ambos os lados são positivos, podemos extrair a raiz quadrada e obter:

$$|\langle u, v \rangle| \leq \sqrt{\langle v, v \rangle} \sqrt{\langle u, u \rangle}$$

Encerrando a demonstração.

Quando vimos o produto escalar pela primeira vez, vimos que usamos junto a ele a p-norma 2. Essa é a **norma induzida** por esse produto interno, e mostraremos agora que toda norma induz algum produto interno.

**Proposição:** Todo produto interno induz uma norma  $\|v\| = \sqrt{\langle v, v \rangle}$ .

**Demonstração:**

- $\|v\| \geq 0$  e  $\|v\| = 0 \iff \langle v, v \rangle = 0 \iff v = 0$ .
- $\|\alpha v\|^2 = \langle \alpha v, \alpha v \rangle = \alpha \bar{\alpha} \langle v, v \rangle = |\alpha|^2 \|v\|^2$ .
- Para a desigualdade triangular, use  $\|u + v\|^2 = \|u\|^2 + 2\Re\langle u, v \rangle + \|v\|^2$  e que  $\Re\langle u, v \rangle \leq |\langle u, v \rangle| \leq \|u\|\|v\|$ .

Um detalhe útil para ângulos: definimos

$$\cos \theta = \frac{\langle u, v \rangle}{\|u\| \cdot \|v\|}$$

que em  $\mathbb{R}$  varia em  $[-1, 1]$ .

Vamos encerrar essa discussão falando sobre **normas de matrizes**.

Matrizes  $A \in \mathbb{F}^{m \times n}$  também admitem normas. A **norma induzida** por uma norma vetorial  $\|\cdot\|$  é definida por:

$$\|A\| := \sup_{y \neq 0} \frac{\|Ay\|}{\|y\|}.$$

Uma propriedade útil é que para toda norma induzida,  $\|AB\| \leq \|A\|\|B\|$ .

Uma **norma geral** de matrizes é qualquer  $\|\cdot\|$  que em  $\mathbb{F}^{m \times n}$  satisfaz as mesmas três propriedades de comprimento, homogeneidade e desigualdade triangular definidas para as normas vetoriais.

Um exemplo importante de norma não-induzida é a **norma de Frobenius**:

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2} = \sqrt{\text{tr}(A A^*)},$$

que satisfaz  $\|AB\|_F \leq \|A\|_F \|B\|_F$  e tem interpretação como a p-norma 2 de um vetor de dimensão  $mn$ .

## 2.4 Determinante

Em um espaço vetorial, podemos definir construtos usando vetores e cópias deles para definir paralelogramos, paralelepípedos, e outros hiper sólidos. Uma característica desses construtos é que eles podem possuir uma medida, como área, volume, hipervolume, etc. Para melhor entendimento, entenda hipervolume como uma dessas medidas em uma dimensão qualquer.

Vimos também que matrizes podem ser interpretadas como transformações lineares, isto é, multiplicar por uma matriz a esquerda é uma forma de deformar o espaço, onde podemos definir todo o espaço apenas descobrindo o que ocorre com a base dele. Com isso, sabemos como entender o que acontece com os vetores em si, mas o que acontece com a medida dos construtos?

Para isso, precisamos de uma ferramenta matemática que seja capaz de medir como uma determinada transformação linear deforma o espaço e seu hipervolume. Para encontrar essa ferramenta, precisamos definir o que esperamos dela. Para interpretar melhor, tente pensar no  $\mathbb{R}^2$  e  $\mathbb{R}^3$ , com a área e volume respectivamente.

- Veja que se estamos no  $\mathbb{R}^n$  e dois dos  $n$  vetores que usamos no construto são iguais, podemos substituí-los por um só vetor com o dobro do tamanho, então nosso construto na verdade tem uma dimensão faltando, e portanto, seu hipervolume é zero.
- Para termos uma noção padrão de volume, precisamos definir um referencial, e para isso, faz sentido que definimos o hipervolume como 1 quando estivermos no pedaço do espaço mais simples possível: O hipercubo formado pelos vetores da base canônica.
- Por fim, faz sentido que essa deformação do hipervolume seja linear em cada coordenada, ou seja, dado um construto de vetores, se por exemplo triplicarmos um vetor, o hipervolume também deve triplicar. Geometricamente, veja que isso faz bastante sentido com o que ocorre com área e volume, então também deve fazer com hipervolume.

Essas noções nos permitem definir **três propriedades** que gostaríamos de esperar dessa ferramenta. Seja  $D$  essa função de medição que recebe como entrada  $n$  vetores de dimensão  $n$ . Então, temos:

- $v_i = v_j$  para algum  $i \neq j \rightarrow D(v_1, \dots, v_n) = 0$
- $D(e_1, \dots, e_n) = 1$ , onde  $e_i$  é o  $i$ -ésimo vetor da base canônica.
- $D(v_1, \dots, \alpha v_i + v_j, \dots, v_n) = \alpha D(v_1, \dots, v_i, \dots, v_n) + D(v_1, \dots, v_j, \dots, v_n)$

E se o determinante cumpre essas propriedades, podemos notar que ele também cumpre algumas outras, que podem ser obtidas combinando as propriedades definidas anteriormente:

- $\{v_1, \dots, v_n\}$  é linearmente dependente  $\rightarrow D(v_1, \dots, v_n) = 0$
- $D(v_1, \dots, v_i, \dots, v_j, \dots, v_n) = -D(v_1, \dots, v_j, \dots, v_i, \dots, v_n)$

E para mostrarmos as propriedades acima, veja que a primeira podemos mostrar de forma direta: se é linearmente independente, podemos dividir esse determinante em uma soma de várias partes dividindo a entrada  $i$  na sua combinação linear das outras, e como cada uma das partes tem dois vetores iguais, todas são zero, mostrando que a soma total também é zero. Já para a segunda propriedade, veja que todas as linhas abaixo são equivalentes:

1.  $D(\dots, v_i, \dots, v_j, \dots)$
2.  $D(\dots, v_i, \dots, v_j, \dots) + D(\dots, v_i, \dots, v_i, \dots)$
3.  $D(\dots, v_i, \dots, v_i + v_j, \dots)$
4.  $D(\dots, v_i, \dots, v_i + v_j, \dots) - D(\dots, v_i + v_j, \dots, v_i + v_j, \dots)$
5.  $D(\dots, -v_j, \dots, v_i + v_j, \dots)$
6.  $-D(\dots, v_j, \dots, v_i + v_j, \dots)$
7.  $D(\dots, v_j, \dots, -v_i - v_j, \dots)$
8.  $D(\dots, v_j, \dots, -v_i - v_j, \dots) + D(\dots, v_j, \dots, v_j, \dots)$
9.  $D(\dots, v_j, \dots, -v_i, \dots)$
10.  $-D(\dots, v_j, \dots, v_i, \dots)$

Nos números reais, uma propriedade interessante é que o sinal de  $D$  define a **orientação dos vetores**, e por mais que isso seja um conceito um pouco abstrato de se entender, ainda podemos visualizar isso no  $\mathbb{R}^2$  e no  $\mathbb{R}^3$ .

Vimos que no geral, gostaríamos muito que essa função existisse de fato, e seria ainda melhor se ela fosse única. Felizmente, essa é a realidade em que vivemos: apenas essas três propriedades são suficientes para definirmos o determinante.

Seja  $D(v_1, \dots, v_n) := \det(v_1, \dots, v_n) = \sum_{\forall p} \sigma_p \cdot v_{1p_1} \cdot v_{2p_2} \cdot \dots \cdot v_{np_n}$

Onde  $p$  é uma permutação qualquer  $(p_1, p_2, \dots, p_n)$  dos números  $(1, 2, \dots, n)$  e  $\sigma_p$  é o **sinal** da permutação: o sinal é 1 se o número de trocas  $(i, j) \rightarrow (j, i)$  partindo do conjunto  $(1, 2, \dots, n)$  para chegar até  $(p_1, p_2, \dots, p_n)$  é par, e caso seja ímpar, o sinal é  $-1$ .

Para todas as formas distintas de fazer essas trocas, é garantido que o sinal da permutação é o mesmo em todas elas: Suponha que seja possível partir de  $(1, \dots, n)$  chegar em  $(p_1, \dots, p_n)$  de duas formas distintas executando trocas  $(i, j) \rightarrow (j, i)$  onde a paridade do número de operações é diferente.

Ao fazer um dos caminhos e depois fazer o outro caminho na ordem inversa, temos uma sequência com número ímpar de trocas que leva  $(1, \dots, n)$  em  $(1, \dots, n)$ .

Dada uma permutação arbitrária, defina o **número de inversões** dela como a quantidade de pares  $(p_i, p_j)$  com  $i < j$ ,  $p_i > p_j$ . Veja que no início, esse número é zero. Ao fazer uma troca  $(i, j) \rightarrow (j, i)$ , o seguinte ocorre:

- O par  $(i, j)$  altera o número de inversões em  $\pm 1$ .
- $\forall k < \min(i, j)$ , os pares  $(i, k)$  e  $(j, k)$  seguem inalterados.
- $\forall k > \max(i, j)$ , os pares  $(i, k)$  e  $(j, k)$  seguem inalterados.
- $\forall k$  onde  $\min(i, j) < k < \max(i, j)$  onde  $p_k > \max(p_i, p_j)$  ou  $p_k < \min(p_i, p_j)$ , os pares  $(i, k)$  e  $(j, k)$  seguem inalterados.
- $\forall k$  onde  $\min(i, j) < k < \max(i, j)$  onde  $\min(p_i, p_j) < p_k < \max(p_i, p_j)$ , cada dupla de pares  $(i, k)$  e  $(j, k)$  altera a soma em  $\pm 2$ .

Dito isso, **toda troca da forma  $(i, j) \rightarrow (j, i)$  altera a paridade do número de inversões**. Com isso, é impossível existir um caminho  $(1, \dots, n) \rightarrow (1, \dots, n)$  com um número ímpar de trocas, pois o número de inversões terá o sinal inverso do que deveria ser, e portanto, o **sinal da permutação é bem definido** conforme a **paridade do número de inversões**.

Dito disso, vamos verificar se a nota função *det* obedece as propriedades criadas:

- A primeira propriedade é verdade pois se existem dois vetores  $v_i = v_j$ , para toda permutação  $p$ , existe uma outra permutação  $p'$  que é análoga, mas troca  $i, j$  de lugar, tendo então o sinal inverso. Ao somar as partes de  $p$  e  $p'$ , resultará numa soma zero, e como isso vale para todas as permutações, o somatório todo é igual a zero.
- A segunda propriedade também é verdadeira, pois o único valor sem nenhuma entrada zero é quando a permutação correspondente a  $p = (1, \dots, n)$ , que possui sinal 1, fazendo o somatório ser igual a 1.
- Para a terceira propriedade, veja que ela também vale para nossa função *det* pois aquela propriedade vale para os reais: podemos tanto multiplicar por escalar quanto somar um valor em todas as permutações possíveis, então podemos fatorar os termos que queremos e chegar na expressão separada.

Como a função *det* existe e obedece as propriedades, então temos uma função que é candidata a ser nossa medida de hipervolume. Agora, precisamos verificar se ela é única. Seja  $v_j$  um vetor arbitrário no conjunto  $V = \{v_i, i \in [n]\}$ . Veja que:

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

Pela propriedade 3, temos:

$$D(V) = D\left(\sum_{i=1}^n v_{1i} e_i, \sum_{i=1}^n v_{2i} e_i, \dots, \sum_{i=1}^n v_{ni} e_i\right).$$

E separando os somatórios pela terceira propriedade:

$$D(V) = \sum_{f:[n] \rightarrow [n]} D\left(v_{1f(1)} e_{f(1)}, \dots, v_{nf(n)} e_{f(n)}\right).$$

Veja que o somatório está iterando sobre  $f : [n] \rightarrow [n]$ , que são as  $n^n$  funções com domínio e contradomínio como o conjunto  $\{1, 2, \dots, n\}$ : cada uma corresponde a um dos  $n^n$  termos que foram separados dos somatórios. Perceba então que, se  $f(i) = f(j)$  com  $i \neq j$ , então aparecem dois vetores linearmente dependentes no termo correspondente a esse  $f$ , o que leva o determinante a ser zero, pela propriedade adicional 2 (que usa as propriedades 1 e 3). Assim, temos:

$$D(V) = \sum_{p \text{ permutação}} D(v_{1p(1)} e_{p(1)}, \dots, v_{np(n)} e_{p(n)}),$$

e então:

$$D(V) = \sum_{p \text{ permutação}} v_{1p(1)} \times v_{2p(2)} \times \dots \times v_{np(n)} \times D(e_{p(1)}, \dots, e_{p(n)}),$$

que, pela propriedade adicional 1 e também pela propriedade 2, é igual a:

$$D(V) = \sum_{p \text{ permutação}} \sigma_p \times v_{1p(1)} \times v_{2p(2)} \times \dots \times v_{np(n)}.$$

E note que isso é exatamente a nossa função suposta anteriormente. Portanto, se uma função  $D$  existir com aquelas propriedades, então ela é  $\det(V)$ , e como ela cumpre as propriedades, temos que **ela é a nossa única expressão para o determinante possível**.

Agora, seja  $\det(A)$  o determinante de uma matriz  $A \in \mathbb{F}^{n \times n}$ , representando um conjunto de  $n$  vetores de dimensão  $n$ . Vamos conferir algumas propriedades adicionais do determinante:

- $\det(A^T) = \det(A)$
- $\det(A) \neq 0 \leftrightarrow A$  possui inversa
- $\det(AB) = \det(A) \cdot \det(B)$ , onde  $B \in \mathbb{F}^{n \times n}$
- Se  $A$  tem inversa,  $\det(A^{-1}) = \det(A)^{-1}$

Podemos interpretar o determinante como a proporção de distorção do espaço, e dessa forma, todas essas propriedades fazem sentido ao tentarmos interpretar com frases significativas: “duplicar e depois triplicar o espaço é o mesmo que sextuplicar”, “se uma matriz triplica o espaço, a sua inversa o divide por três”, e coisas do gênero.

Agora, vamos demonstrar essas propriedades. A primeira em particular é simples: Para toda permutação cujos termos escolhidos são  $[(1, p_1), (2, p_2), \dots, (n, p_n)]$ , existe uma outra permutação  $[(p_1, 1), (p_2, 2), \dots, (p_n, n)]$  que multiplica exatamente os mesmos termos para  $A^T$ , tendo também o mesmo sinal. Com isso, existe uma bijeção entre os termos somados, que infere no determinante de  $A^T$  ser igual ao de  $A$ .

Para a segunda propriedade, veja que  $A$  tem linhas linearmente independentes, então seu determinante é zero pela propriedade adicional que já demonstramos. Agora, se  $A$  tem posto cheio, então podemos escrever  $U = E_k E_{k-1} \dots E_1 A$ , onde  $E_i$  são matrizes elementares de permutação ou soma. Veja ainda que:

- Se  $E$  é uma matriz elementar de permutação, então para todo termo no somatório de  $\det(EA)$  existe um correspondente em  $\det(A)$ , onde a única diferença está no sinal, divergindo pela troca de posições feita por  $E$ . Assim, temos que  $\det(EA) = \det(E) \det(A)$ .
- Se  $E$  é uma matriz elementar de soma, então  $\det(E)=1$ . Além disso,  $\det(EA) = \sum_p \sigma_p \prod_{i \neq j} a_{ip_i} \cdot (a_{jp_j} + \alpha_p \cdot a_{kp_j}) = \sum_p \sigma_p \prod_{i \neq j} a_{ip_i} \cdot (a_{jp_j}) + \alpha_p \sum_p \sigma_p \prod_{i \neq j} a_{ip_i} \cdot (a_{kp_j})$ , onde  $\alpha_p$  é o coeficiente de multiplicação da operação elementar. O somatório da direita pega os elementos  $a_{ip_i}$ , exceto com  $i = j$ , que pega o termo  $a_{kp_j}$ . Perceba que para toda permutação,  $p$ , existe outra permutação  $p'$  que troca  $p_j, p_k$  de lugar, mas que a soma nesse somatório continua intacta.



Veja que o sinal de  $p'$ , por outro lado, muda: isso implica que as somas de  $p$  e  $p'$  se anulam, resultando em todo o somatório sendo zero, restando apenas o somatório da esquerda, que é igual a  $\det(A)$ . Portanto,  $\det(EA) = \det(A) = \det(E) \det(A)$ .

Com essas duas observações,  $\det(EA) = \det(E) \det(A)$  para toda  $E$  operação elementar de soma ou permutação. Com isso, temos que  $\det(U) = \det(E_1) \det(E_2) \dots \det(E_k) \det(A) = \pm \det(A)$ . Além disso, como  $A$  é quadrada e tem posto cheio,  $U$  não tem elementos nulos na diagonal, portanto, podemos ver que  $\det(U) = \prod_i U_{ii} \neq 0$ , e então,  $\det(A) \neq 0$ . Com isso, concluímos a segunda propriedade.

Para a terceira propriedade, vamos dividir em alguns casos.

- Caso 1:  $\det(A) \neq 0$ .

Nesse caso, considere a função  $f(B) = \det(AB)/\det(A)$ . Mostraremos que essa função é o determinante de  $B$  verificando que ela cumpre as três propriedades que definem o determinante.

1. Se  $B$  tem duas colunas iguais  $b_i, b_j$ , então  $AB$  tem duas colunas iguais  $Ab_i, Ab_j$ , e pela propriedade do determinante,  $\det(AB) = 0 \rightarrow f(B) = 0$ , cumprindo a primeira propriedade do determinante.
2. Se  $B$  é a matriz identidade,  $\det(AB)/\det(A) = \det(A)/\det(A) = 1$ , comprovando a segunda propriedade do determinante.
3.  $f(\{b_1, \dots, b_{i-1}, \alpha b_i + b_j, b_{i+1}, \dots, b_n\}) = \det(\{Ab_1, \dots, Ab_{i-1}, \alpha Ab_i + Ab_j, Ab_{i+1}, \dots, Ab_n\}) / \det(A)$ , e veja que o próprio determinante nos permite separar o termo do numerador, comprovando que  $f$  também obedecerá essa propriedade.

Com isso,  $f(B)$  possui as mesmas propriedades que  $\det(B)$ , logo, podemos usar a unicidade do determinante e mostrar que  $\det(B) = \det(AB)/\det(A)$ , que implica em  $\det(AB) = \det(A) \det(B)$ .

- Caso 2:  $\det(A) = 0$ .

Para esse caso, veja que como  $\dim(\text{Im}(A)) < n$ , então  $\dim(\text{Im}(AB)) \leq \dim(\text{Im}(A)) < n$ , e portanto,  $AB$  não é inversível. Isso implica que  $\det(AB) = 0$ , e assim,  $0 = 0 \times \det(B) \rightarrow \det(AB) = \det(A) \times \det(B)$ .

E demonstrada a propriedade anterior, é fácil demonstrar a quarta:  $A \cdot A^{-1} = I$ , que implica  $\det(A) \cdot \det(A^{-1}) = \det(I)$ , mas como  $\det(I) = 1$ , então  $\det(A^{-1}) = \det(A)^{-1}$ .

### 3 Ortogonalidade

A ortogonalidade é uma forma de generalizarmos perpendicularidade, sendo um tópico muito relevante para toda a álgebra linear. Vamos começar uma com introdução a esse conceito, como encontrar vetores ortogonais e depois conferir um dos principais teoremas que veremos até aqui.

#### 3.1 Vetores Ortogonais

Retome que a projeção de um vetor  $v$  sobre um vetor  $w$  é a sombra de  $v$  sobre  $w$ . Podemos fornecer uma definição mais precisa de projeção ao considerar a projeção de um vetor  $v \in V$  sobre um espaço vetorial  $W$  subespaço de  $V$ :

$$\text{proj}_W(v) = \text{argmin}_w \{\|v - w\|^2\}$$

Essa definição significa que a projeção de  $v$  em  $W$  é o vetor  $w \in W$  mais próximo de  $v$ , isto é, que minimiza a norma das diferenças. Geometricamente, vemos que o ângulo entre  $v$  e  $W$  deve ser reto (pense por exemplo em  $W$  como um plano), e então, os vetores  $v$  e  $w$  serão perpendiculares (claro, se  $v \notin W$ ).

No caso real, o produto escalar está diretamente relacionado ao ângulo do vetor, e de fato será essa noção que vamos adotar para generalizar essa noção de perpendicularidade: dois vetores  $u, v$  são **ortogonais** se  $\langle u, v \rangle = 0$ .

Assim, veja que se tirarmos a projeção do vetor  $v$ , a sua componente perpendicular é perpendicular a todo o espaço. Podemos ver isso notando que  $\langle w, v - \text{proj}_W(v) \rangle \forall w \in W$ .

Mas para isso ser verdade para todo o espaço  $W$ , basta ver que vale para uma base  $(w_1, \dots, w_n)$  desse espaço, pois a base é capaz de construir todo o espaço, e então:

$$\langle w, v - \text{proj}_W(v) \rangle = \left\langle \sum_{i=1}^n w_i \alpha_i, v - \text{proj}_W(v) \right\rangle = \sum_{i=1}^n \alpha_i \langle w_i, v - \text{proj}_W(v) \rangle = 0$$

Agora, basta nos concentrarmos em encontrar um vetor  $u$  que seja ortogonal a todos os vetores da base de  $W$ . Para isso, veja que podemos representar o produto interno  $\langle a, b \rangle$  como  $a^*b$  (no caso real e complexo ao mesmo tempo). Então, considerando a matriz  $A$  de colunas  $w_1, \dots, w_n$ , queremos resolver o sistema  $A^*(v - \text{proj}_W(v)) = 0 \rightarrow A^*v = A^*\text{proj}_W(v)$ .

Além disso, como  $\text{proj}_W(v) \in W$ , ele pode ser escrito como  $Au$  para algum  $u$ . Assim, temos:  $A^*v = A^*Au$ . Assim, basta resolver esse sistema linear e conseguiremos encontrar a projeção: se  $A$  for uma base e não apenas um conjunto gerador, então a solução também é única, pois  $A^*A : \mathbb{F}^n \rightarrow \mathbb{F}^n$ , que implica que  $A^*A$  tem inversa e então  $u = (A^*A)^{-1}A^*v$ .

Agora, vamos para o caso  $W = V$ . Perceba que o vetor mais próximo de  $v$  será de fato  $v$ , uma afirmação até um pouco óbvia. Na verdade, ao definirmos a matriz  $A$ , sabemos que  $x$  não passa das coordenadas de  $v$  na base descrita pela matriz  $A$ , ou seja,  $Ax = v$ .

Em outras palavras, podemos perceber um fato importante em tudo isso: para encontrar as coordenadas de um vetor numa base qualquer ou mesmo para obter a projeção de um vetor em um subespaço qualquer (com produto interno canônico) é necessário resolver um sistema linear.

Porém, resolver sistemas lineares é muito caro, e a Eliminação Gaussiana que conhecemos por si só é um algoritmo da ordem de complexidade  $O(n^3)$ .

Ainda assim, é possível construir bases em que não é necessário efetuar a Eliminação Gaussiana para calcularmos suas coordenadas, e essas bases são chamadas de **bases ortonormais**.

Basicamente, nossa pergunta atual é: dado um espaço vetorial  $V$  e um subespaço  $W$ , existe uma base de  $W$  cujos vetores são todos ortogonais entre si?

Em particular, se uma base é ortogonal e a norma de todos os vetores é 1 (ou melhor: o produto interno entre qualquer vetor e ele mesmo é 1, e 0 em qualquer outro caso), então dizemos que ela é uma **base ortonormal**. Veja que se  $U$  é a matriz cujas colunas são vetores ortonormais, isto é,  $\{u_1, u_2, \dots, u_n\}$  forma uma base ortonormal, então  $U^*U = I$ . Nesse caso, dizemos que a matriz  $U$  é **unitária**.

Tome  $v \in V$ , escrito em uma base  $\{v_1, \dots, v_n\}$  como  $v = \sum_{i=1}^n \alpha_i v_i$ . Note que  $\langle v, v_j \rangle = \sum_{i=1}^n \alpha_i \langle v_i, v_j \rangle$ . Se essa base for ortonormal, temos que  $\langle v, v_j \rangle = \alpha_j$ . Isso nos mostra que **em uma base ortonormal, as coordenadas de  $v$  nessa base são**  $\{\langle v, v_1 \rangle, \langle v, v_2 \rangle, \dots, \langle v, v_n \rangle\}$ .

Se um conjunto de vetores é ortogonal, ele necessariamente é  $LI$ . Isso é uma propriedade bastante intuitiva, mas também fácil de provar: suponha que  $\{v_1, \dots, v_r\}$  seja um subconjunto da base onde  $v_r$  pode ser escrito como combinação linear dos anteriores. Assim,  $\langle v_r, v_r \rangle = \sum_{i=1}^{r-1} \alpha_i \langle v_i, v_r \rangle$ , porém, o lado esquerdo por definição é 1 e o direito é uma soma de zeros, gerando um absurdo.

Agora que sabemos o que são bases ortogonais e ortonormais, como podemos encontrar elas? O algoritmo introdutório que veremos para esse objetivo se chama **ortogonalização de Gram-Schmidt**, descrito como o seguinte algoritmo:

1. Tome uma base qualquer  $\{v_1, \dots, v_r\}$  de um espaço vetorial  $V$
2. Defina  $w_1 = \frac{v_1}{\|v_1\|}$ .
3. Defina  $w'_i = v_i - \sum_{j=1}^{i-1} \text{proj}_{w_j}(v_i)$
4. Defina  $w_i = \frac{w'_i}{\|w'_i\|}$
5. Ao repetir os passos 3 e 4  $\forall i \in [2, r]$ , teremos uma base ortonormal  $\{w_1, \dots, w_r\}$

No geral, o que esse algoritmo está fazendo é tirar a componente da projeção de um vetor da base sobre os vetores que você tem que representarão sua base ortonormal. Como a projeção foi removida,  $w_i$  se torna ortogonal aos vetores  $w_j$  anteriores, enquanto a definição de  $w'_i$  nos assegura que conseguimos recuperar a informação de  $v_i$  usando combinação linear dos vetores anteriores da sua base, e então, podemos descrever todo o espaço como combinação linear dos seus novos vetores  $w_i$ .

Como vetores ortogonais são  $LI$  e as duas bases tem o mesmo tamanho, temos de fato uma base de vetores ortonormais. Se você não desejar a normalidade, apenas a ortogonalidade, basta ignorar o passo 2 e 4: não precisa dividir pela norma.

Para sermos mais rigorosos com as propriedades, vamos demonstrar mais rigorosamente que isso é uma base ortogonal. Veja que inicialmente, é fácil ver que  $w_1$  gera o mesmo espaço que  $v_1$ . Assim, suponha que para algum  $i$ ,  $\{w_1, \dots, w_i\}$  gera o mesmo espaço que  $\{v_1, \dots, v_i\}$ .

Agora, precisamos verificar que a adição do novo vetor continua mantendo o espaço  $\{w_1, \dots, w_{i+1}\}$  gerando o mesmo espaço que  $\{v_1, \dots, v_{i+1}\}$ . Para isso, precisamos de uma boa expressão para a projeção de um vetor sobre outro.

Suponha que dados dois vetores  $v, w$ , queremos uma expressão para  $\text{proj}_w(v)$ . Sabemos que ela é um vetor da forma  $\alpha w$  onde  $\langle v - \alpha w, w \rangle = 0$ , que também implica em  $\langle v - \alpha w, \alpha w \rangle = 0$ . Usando a linearidade do produto interno, temos que isso equivale a  $\langle v, \alpha w \rangle - \langle \alpha w, \alpha w \rangle = 0$ , que por sua vez equivale a  $\bar{\alpha} \langle v, w \rangle = \bar{\alpha} \langle \alpha w, w \rangle \rightarrow \langle v, w \rangle = \alpha \langle w, w \rangle \rightarrow \alpha = \frac{\langle v, w \rangle}{\langle w, w \rangle}$ . Agora **temos uma expressão para a projeção de  $v$  sobre  $w$** :  $\text{proj}_w(v) = w \frac{\langle v, w \rangle}{\langle w, w \rangle}$

Veja que  $\{w_1, \dots, w_i, v_{i+1}\}$  gera o mesmo espaço que  $\{v_1, \dots, v_{i+1}\}$ , e isso também infere que  $\left\{ w_1, \dots, w_i, v_{i+1} - \sum_{j=1}^i w_j \frac{\langle v_{i+1}, w_j \rangle}{\langle w_j, w_j \rangle} \right\}$  gera o mesmo espaço que  $\{v_1, \dots, v_{i+1}\}$ , pois o último vetor foi obtido apenas somando múltiplos dos outros vetores já presentes na base, e o mesmo continua valendo ao normalizarmos esse vetor.

Como  $w_{i+1}$  foi normalizado,  $\langle w_{i+1}, w_{i+1} \rangle = 1$ . Além disso,  $\forall k < i + 1$ :

$$\begin{aligned} \langle w_{i+1}, w_k \rangle &= \left\langle v_{i+1} - \sum_{j=1}^i w_j \frac{\langle v_{i+1}, w_j \rangle}{\langle w_j, w_j \rangle}, w_k \right\rangle = \langle v_{i+1}, w_k \rangle - \sum_{j=1}^i \langle w_j \langle v_{i+1}, w_j \rangle, w_k \rangle \\ &= \langle v_{i+1}, w_k \rangle - \langle w_k \langle v_{i+1}, w_k \rangle, w_k \rangle = \langle v_{i+1}, w_k \rangle - \langle v_{i+1}, w_k \rangle \langle w_k, w_k \rangle = 0 \end{aligned}$$

E com isso, concluímos que  $\{w_1, \dots, w_{i+1}\}$  são vetores ortonormais que geram o mesmo espaço que  $\{v_1, \dots, v_{i+1}\}$ . Por indução, vemos que quando  $i = r$ , teremos que esse conjunto será de fato uma base ortonormal para nosso espaço.

### 3.2 Autovalores e Autovetores

Resolver sistemas lineares é o tópico central da Álgebra Linear, e vimos que no geral é um processo complexo, requerendo da ordem de  $O(n^3)$  operações para resolver usando Eliminação Gaussiana. Se a matriz já for **triangular**, podemos usar substituição retroativa ou progressiva diretamente e obter algo melhor: da ordem de  $O(n^2)$ . Mas além desses casos, a matriz ainda pode ser **diagonal**, tornando o processo de resolver o sistema linear da ordem de  $O(n)$ .

Se vamos trabalhar com uma matriz  $A$  e muitos vetores diferentes  $b$ , vimos que uma coisa inteligente a se fazer é realizar a decomposição LU para decompor a matriz em triangulares inferior e superior, tornando o processo de resolver  $Ax = b$  em  $PLUx = b$ , que por ser uma permutação de uma multiplicação de matrizes triangulares, é um processo da ordem de  $O(n^2)$  por vetor que queremos resolver novamente.

Sabendo dos três casos, algo que poderíamos nos perguntar é: como matrizes diagonais são melhores que triangulares, tem como eu fazer algum procedimento para considerar a minha matriz qualquer como uma matriz diagonal? A resposta para essa pergunta é: em alguns contextos, como uma base específica, sim.

Seja  $T : V \rightarrow W$  uma transformação linear. Dizemos que  $T$  é **diagonalizável** se existe uma base  $B = \{b_1, \dots, b_n\}$  e uma matriz  $A_B$  de modo que  $A_B$  seja uma matriz diagonal que representa  $T$  na base  $B$ .

Para interpretar esse conceito melhor, suponha que estamos lidando em um dado momento com uma base  $C = \{c_1, \dots, c_n\}$  qualquer. Também podemos dizer que se existe uma base  $X$ , uma matriz inversível de mudança de base  $R = [r_1, \dots, r_n]$  tal que  $r_i$  corresponde às coordenadas de  $c_i$  na base  $X$  e a transformação linear  $T$  possa ser descrita como  $R^{-1}DR$ , com  $D$  sendo matriz diagonal com entradas  $\lambda_i$ , então  $T$  é diagonalizável.

Veja que a definição acima é equivalente a anterior: apenas estamos lidando com uma base comum  $C$ , encontrando a base desejada  $X$  e aplicando matrizes de mudança de base  $R_{C \rightarrow X}$  e  $R_{X \rightarrow C}^{-1}$  para transitar entre as bases. Podemos definir também  $S = [s_1, \dots, s_n] = R^{-1}$  e escrever  $T$  como  $A = SDS^{-1}$ , pois a matriz  $S$  tem um significado mais relevante em alguns cenários. Por consequência da necessidade da inversa, a diagonalização só existe para matrizes quadradas.

Usando essa última expressão, note que  $Se_i = s_i \forall i \in [1, n]$ , e isso infere que  $S^{-1}Se_i = S^{-1}s_i = e_i$ . Com isso,  $As_i = (SDS^{-1})s_i = SDe_i = S(\lambda_i e_i) = \lambda_i s_i$ . Isso mostra que a matriz  $A$  tem o efeito de escalar por um fator  $\lambda_i$  cada vetor  $s_i$  coluna de  $S$ . Se uma matriz é diagonalizável, os vetores  $s_i$  são chamados de **autovetores** de  $A$ , enquanto os valores  $\lambda_i$  são chamados de **autovalores** de  $A$ . Cada autovetor tem um autovalor correspondente, mas podem ter múltiplos autovetores com mesmo autovalor. Chamamos de **espectro** de  $A$  o conjunto de todos os autovalores  $\Lambda(A)$ .

Mais precisamente: Dizemos que um vetor  $v \neq 0$  é um **autovetor** de  $A$  com autovalor  $\lambda$  se  $Av = \lambda v$ .  $\lambda$  é um autovalor de  $A$  somente se existir um autovetor correspondente. Com base nessa definição, uma matriz é **diagonalizável** se ela possui uma **base de autovetores**.

Uma possível interpretação para isso é que existe um conjunto de direções tais que se você aplicar a transformação linear da matriz  $A$  pelo seu vetor (na base que você está agora), e decompor seu vetor nessas direções, a sua transformação basicamente apenas escala a componente de  $v_i$  em um fator de  $\lambda_i$ . Nos complexos, é relevante saber que escalar por números imaginários é como se fosse uma combinação de rotacionar e escalar, mas os conceitos ainda valem similarmente.

Agora, veremos como encontrar a diagonalização de uma matriz (ou verificar que não existe). Primeiro, seja  $\lambda$  um possível autovalor. Então, queremos encontrar todas as soluções para  $Av = \lambda v$  com  $v \neq 0$ . Podemos escrever isso como  $(A - \lambda I)v = 0$ . Como  $v$  representa uma combinação linear das colunas de  $A - \lambda I$  e  $v \neq 0$ , só existe uma solução se  $A - \lambda I$  não tem posto cheio, isto é,  $\det(A - \lambda I) = 0$ .

Chamamos o valor  $\det(A - \lambda I)$  de **polinômio característico** de  $A$ , onde estamos interessados em suas raízes para a variável  $\lambda$ . Ao calcular as raízes dele, teremos algo como  $\det(A - \lambda I) = (\lambda - \lambda_1)^{\beta_1}(\lambda - \lambda_2)^{\beta_2} \dots (\lambda - \lambda_k)^{\beta_k}$ . Os autovalores são o conjunto  $\{\lambda_1, \dots, \lambda_k\}$ . Os expoentes  $\beta_1, \dots, \beta_k$  são chamados de **multiplicidade algébrica** do autovalor correspondente.

Veja que se  $\lambda_i$  é uma raiz do polinômio característico, significa que existe um vetor  $v$  que representa uma combinação linear das colunas de  $A - \lambda_i I$  que resulta em zero, então, sempre existe ao menos um autovetor para cada autovalor, mas não sabemos imediatamente quantos existem.

Definimos a **multiplicidade geométrica** de um autovalor  $\lambda_i$  como  $\dim(\{v \in \mathbb{F}^n : Av = \lambda_i v\})$ . O conjunto descrito é chamado de **autoespaço**, e a multiplicidade geométrica é a dimensão dele.

No geral, o que acontece é que a multiplicidade algébrica é ao menos tão grande quanto a multiplicidade geométrica, mas nem sempre ambas são iguais. Na realidade, o fato de serem iguais ou não nos diz muito sobre a matriz.

**Lema:** Dada uma matriz  $A$ , a sua multiplicidade algébrica sempre é maior ou igual a sua multiplicidade geométrica.

**Demonstração:** Seja  $V = \{v_1, \dots, v_k\}$  uma base de  $N(A - \lambda I)$ . Complete a base com vetores  $v_{k+1}, \dots, v_n$  para que tenhamos uma base de  $\text{Im}(A - \lambda I)$ . Veja que  $Av_i = \lambda v_i$  com  $i \leq k$ . Ao representarmos a matriz  $A$  na base  $\{v_1, \dots, v_n\}$ , temos  $A = SXS^{-1}$ .

Além disso, veja que  $\det(X - \lambda I) = \det(S^{-1}AS - \lambda I) = \det(S^{-1}AS - S^{-1}(\lambda I)S) = \det(S^{-1}(A - \lambda I)S) = \det(S^{-1})\det(A - \lambda I)\det(S) = \det(A - \lambda I)\det(S)\det(S)^{-1} = \det(A - \lambda I)$ . Agora, podemos calcular  $\det(A - \lambda I)$  usando  $\det(X - \lambda I)$ .

Como as primeiras  $k$  colunas de  $X$  são compostas apenas de elementos da diagonal ou zeros, essa propriedade segue para  $X - \lambda I$ . Sendo  $p_A(x)$  o polinômio característico de  $A$ , usando que ele é calculado com permutações que tenham ao menos um elemento não nulo em cada linha, temos que  $(x - \lambda)^k \mid p_A(x)$ . Com isso, temos que ao menos  $k$  raízes do polinômio são iguais a  $\lambda$ , mostrando que a multiplicidade algébrica é no mínimo igual a multiplicidade geométrica  $k$ .

E com esse lema demonstrado, veja que a nossa única esperança de termos uma base de autovetores é se as multiplicidades algébrica e geométrica forem iguais, e neste caso, por definição a diagonalização existe. Agora, temos um algoritmo pronto para descobrir se uma matriz  $A$  possui diagonalização:

- Encontre os autovalores de  $A$  achando as raízes de  $\det(A - \lambda I)$  e suas multiplicidades algébricas.
- Para cada autovalor  $\lambda$  de multiplicidade algébrica  $k$ , tente resolver o sistema linear  $Av = \lambda v$ , encontrando os autovetores de cada autovalor.
- Se algum sistema linear tiver menos de  $k$  soluções, então não existe diagonalização.
- Se as multiplicidades algébrica e geométrica forem iguais para todo autovalor  $v_i$ , defina  $S = [v_1, \dots, v_n]$  e  $D$  uma matriz diagonal cujas entradas diagonais são, em ordem,  $\{\lambda_1, \dots, \lambda_n\}$ , e então, vale que  $A = SDS^{-1}$ .
- Para encontrar  $S^{-1}$ , uma possível forma de fazer isso é resolver todos os  $n$  sistemas lineares da forma  $Ss_i^{-1} = e_i$ .

Se tivermos total liberdade de escolher a base e trabalhar nas coordenadas daquela base, a diagonalização permite resolver os sistemas lineares em tempo linear. Além disso, a diagonalização é uma forma muito eficiente de calcular potências de matrizes, dado que  $A^k = (SDS^{-1})^k = SD^kS^{-1}$ , onde elevar matrizes diagonais a uma potência consiste de elevar as suas entradas diagonais até aquela potência. Agora, veremos algumas equivalências entre conceitos que podemos comparar dado que vimos sobre autovalores:

**Os seguintes fatos são equivalentes:**

1.  $A$  possui inversa.
2. O posto de  $A$  é  $n$ .
3.  $\text{Im}(A) = \mathbb{F}^n$  e  $N(A) = \{0\}$ .
4.  $0$  não é autovetor de  $A$ .
5.  $\det(A) \neq 0$ .

Para verificar as equivalências acima, faremos uma sequência cíclica de implicações  $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 1$ , que mostra que qualquer uma delas implica todas as outras.

Já verificamos anteriormente que a existência da inversa implica que o posto é cheio, garantindo  $1 \rightarrow 2$ . Se o posto é cheio, também vimos que a dimensão da imagem é  $n$  e a dimensão do núcleo é 0, mostrando que  $Im(A) = \mathbb{F}^n$  e  $N(A) = \{0\}$ , sendo isso  $2 \rightarrow 3$ . Se a dimensão do núcleo é 0, então não existe solução não trivial para  $\det(A - \lambda I) = 0$  com  $\lambda = 0$ , pois isso equivale a  $\det(A) = 0$ , mostrando que  $3 \rightarrow 4$ . Pela mesma razão, veja que como 0 não é autovalor, então  $\det(A) = 0$  não tem solução, não trivial, logo,  $\det(A) \neq 0$ , sendo isso  $4 \rightarrow 5$ . Por fim, também já vimos anteriormente que  $\det(A) \neq 0$  implica que existe a inversa, sendo a última implicação que faltava observarmos.

O mesmo pode ser feito para fatos mais relacionados à matriz  $A - \lambda I$ , dado um  $\lambda$ :

**Os seguintes fatos são equivalentes:**

1.  $A - \lambda I$  não possui inversa.
2.  $Dim(Im(A - \lambda I)) < n$ .
3.  $N(A - \lambda I) \neq \{0\}$ .
4.  $\lambda$  é autovetor de  $A$ .
5.  $\det(A - \lambda I) = 0$ .
6.  $(A - \lambda I)v = 0$  possui solução não trivial.
7.  $\exists v : Av = \lambda v$ .

Novamente, podemos verificar sem muitos problemas os fatos acima da mesma forma que fizemos antes, mas no geral, os fatos são bem intuitivos.

### 3.3 Teorema Espectral

Já investigamos quando é possível encontrar uma base onde uma matriz  $A$  pode ser escrita como uma matriz diagonal. Até então, não fizemos nenhuma exigência sobre a matriz  $A$ , ou seja, bastava obter  $S$  inversível tal que  $S^{-1}AS$  fosse diagonal. Aqui, vamos estudar quando é possível obter uma **base ortonormal** na qual a matriz  $A$  é diagonal, isto é,  $S$  ser uma matriz unitária. Agora, verificaremos algumas propriedades sobre matrizes **unitárias** ( $A^*A = I$ ) e **auto adjuntas** ( $A = A^*$ ).

1. Se  $U$  é unitária, então  $\langle Uv, Uw \rangle = \langle v, w \rangle$  e  $\|Uv\| = \|v\|$ .

Podemos escrever o produto interno como  $(Uv)^*(Uw) = v^*U^*Uw = v^*w$ . Além disso,  $\|Uv\| = \sqrt{\langle Uv, Uv \rangle} = \sqrt{\langle v, v \rangle} = \|v\|$ .

2. Se  $A$  é auto adjunta, então  $\langle Av, w \rangle = \langle v, Aw \rangle$ .

Veja que  $\langle Av, w \rangle = (Av)^*w = v^*A^*w = v^*(A^*w) = \langle v, A^*w \rangle$ . Como  $A$  é auto adjunta, então  $A^* = A$ , portanto,  $\langle Av, w \rangle = \langle v, Aw \rangle$ .

3. Se  $A$  é auto adjunta, então todos os seus autovalores são reais.

Sendo  $\lambda$  autovalor,  $\lambda \langle v, v \rangle = \langle \lambda v, v \rangle = \langle Av, v \rangle = \langle v, Av \rangle = \langle v, \lambda v \rangle = \bar{\lambda} \langle v, v \rangle$ . Portanto,  $\lambda = \bar{\lambda}$ , mostrando que ele deve ser um número real.

4. Se  $A$  é auto adjunta, então autovetores de diferentes autovalores são ortogonais.

Seja  $Av_1 = \lambda_1 v_1$  e  $Av_2 = \lambda_2 v_2$ . Vemos que  $\lambda_1 \langle v_1, v_2 \rangle = \langle Av_1, v_2 \rangle = \langle v_1, Av_2 \rangle = \lambda_2 \langle v_1, v_2 \rangle$ . A única forma disso ser verdade dado que  $\lambda_1 \neq \lambda_2$  é se  $\langle v_1, v_2 \rangle = 0$ .

5. Se  $A$  é auto adjunta e todos os seus autovalores tem multiplicidade algébrica 1, então existe uma forma de diagonalizar  $A$  em  $SDS^{-1}$  com  $S$  sendo uma matriz unitária.

Podemos mostrar isso pois todos os autovalores tem multiplicidade algébrica 1, para todo autovalor existe ao menos um autovetor e autovetores de diferentes autovalores são ortogonais. Portanto,  $S$  existe e seus vetores coluna são ortogonais entre si. Como podemos na diagonalização decidir a norma do autovetor como quisermos, basta escolhê-los com norma 1 e teremos  $S$  unitária.

Essas propriedades no geral já nos dão muitas informações relevantes sobre quando parece existir essa diagonalização ortogonal: quando a matriz  $A$  é auto adjunta. Apesar disso, não provamos que isso vale para todos os casos, embora sabemos que seja verdade caso todos os autovalores tenham multiplicidade algébrica 1.

O **teorema espectral** é o teorema que é capaz de nos garantir isso, e uma das formas de prová-lo é usando o **lema do agito**, mas não será a forma em que demonstraremos nesse momento. Ainda assim, vamos ver o que diz o lema do agito e como demonstrar ele a título de curiosidade, pois sua ideia geral pode nos render frutos futuramente.

**Lema (Lema do Agito):** Seja  $A$  uma matriz  $n \times n$  auto adjunta. É possível obter matrizes  $A_\epsilon$  de modo que:

- $A_\epsilon$  seja auto adjunta para todo  $\epsilon > 0$ ;
- $\lim_{\epsilon \rightarrow 0} A_\epsilon = A$ ;
- Todos os autovalores de  $A_\epsilon$  possuem multiplicidade algébrica 1.

**Demonstração:** Faremos indução em  $n$ . Se  $n = 1$ , então basta tomar  $A_\epsilon = A$ , e as propriedades serão sempre satisfeitas. Agora, suponha que o teorema seja verdade  $\forall k < n$ . Mostraremos que ele também é verdade para  $n$ .

Seja  $v_1$  um autovetor de  $A$  associado a algum  $\lambda$  com norma 1. A matriz  $A' = A + \epsilon(v_1 v_1^*)$  tem um autovetor  $v_1$  associado ao autovalor  $\lambda + \epsilon$ . Agora, tome  $v_2, \dots, v_n$  vetores que formam juntamente a  $v_1$  uma base ortonormal de  $\mathbb{F}^n$  (por exemplo, use a ortogonalização de Gram-Schmidt) e tome  $S = [v_1, \dots, v_n]$ . Veja que  $A' = SXS^*$ , onde  $X$  possui a primeira coluna sendo o vetor  $(\lambda + \epsilon)e_i$ .

Também podemos escrever  $X = S^* A' S$  e verificar que  $X^* = (S^* A' S)^* = S^* A'^* S = S^* A' S$  pois  $A'$  é auto adjunta, e com isso,  $X$  é auto adjunta.

Dito isso, considere como  $B$  a submatriz auto adjunta  $n - 1 \times n - 1$  inferior direita de  $X$ . Veja que por indução,  $\exists B_\epsilon$  com as três propriedades dadas. Assim, ao escolhermos a matriz  $A_\epsilon$  como:

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

Temos que:

1.  $\lim_{\epsilon \rightarrow 0} A_\epsilon = A$ , pois  $B_\epsilon \rightarrow B$  e  $\lambda + \epsilon \rightarrow \lambda$ .



2. Todos os autovalores de  $A_\epsilon$  são distintos, pois basta escolhermos um fator  $\alpha$  para multiplicar  $\epsilon$  de modo que  $\lambda + \alpha\epsilon$  seja diferente dos outros autovalores.
3.  $A_\epsilon$  é auto adjunta, pois  $B_\epsilon$  também é pela indução.

E com isso, concluímos a demonstração.

E por último, mas o **mais** importante, vamos demonstrar o teorema espectral, de uma forma similar a como demonstramos o lema do agito: por construção e indução.

**Teorema (Teorema Espectral):** Seja  $A$  uma matriz  $n \times n$  auto adjunta. Existe uma matriz diagonal  $D$  e uma matriz unitária  $S$  tais que  $A = SDS^*$ , onde  $S$  é a matriz cujas colunas são autovetores de  $A$ .

**Demonstração:** Provaremos por indução na dimensão. Novamente, é trivial ver que o teorema vale para  $n = 1$ , então assumiremos que vale  $\forall k < n$  e provaremos que vale para  $n$ .

Seja  $\lambda_1$  um autovalor com autovetor normalizado  $u_1$  de  $A$  e  $\{u_2, \dots, u_n\}$  o complemento de vetores de modo que  $\{u_1, \dots, u_n\}$  seja uma base ortonormal de  $\mathbb{F}^n$ . Isto é, a matriz  $U = [u_1, \dots, u_n]$  é unitária. Também podemos escrever  $A = UA'U^*$  para  $A'$  sendo a representação de  $A$  na base de  $U$ . Podemos também concluir os seguintes fatos:

- $U^*Au_1 = U^*(UA'U^*)u_1 = A'U^*u_1 = A'e_1$
- $U^*Au_1 = U^*\lambda_1u_1 = \lambda_1(U^*u_1) = \lambda_1e_1$ .

E os dois fatos juntos mostram que  $A'e_1 = \lambda_1e_1$ . Por isso, podemos escrever  $A'$  como:

$$A' = \begin{bmatrix} \lambda_1 & 0 \\ 0 & A_{n-1} \end{bmatrix}$$

Como  $A$  e  $S$  são auto adjuntas, também é  $A'$ . Usando nossa hipótese indutiva, existem  $S_{n-1}$  unitária com autovetores de  $A_{n-1}$  e  $D_{n-1}$  diagonal tais que  $A_{n-1} = S_{n-1}D_{n-1}S_{n-1}^*$ . Agora, considere as seguintes matrizes:

$$S' = \begin{bmatrix} 1 & 0 \\ 0 & S_{n-1} \end{bmatrix} \quad D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & D_{n-1} \end{bmatrix}$$

$S'$  é unitária e  $D$  é diagonal. Também é possível verificar multiplicando a matriz por blocos que  $A' = S'DS'^*$ . Com isso, podemos escrever  $A = UA'U^* = U(S'DS'^*)U^* = (US')D(S'^*U^*) = (US')D(US')^*$ , onde  $D$  é diagonal com os autovalores de  $A$  e  $US'$  é unitária (produto de unitárias é unitária). Veja que tomando  $S = US'$ , conseguimos com sucesso diagonalizar  $A = SDS^*$  da forma que queríamos no início, finalizando a demonstração do teorema espectral.

Veja que como os autovalores são únicos, a decomposição espectral também é **única a menos de sinais e fases**: a única forma de alterar a decomposição espectral é se ao invés de tomarmos algum autovetor  $s_i$ , tomarmos  $\alpha s_i$  com  $|\alpha| = 1$  (como  $\alpha = -1$  ou  $\alpha = i$ ).