CURSO DE

Álgebra Linear Aplicada

Antonio Cândido Faleiros

Centro de Matemática, Computação e Cognição
Universidade Federal do ABC
Santo André, SP

6 de abril de 2009

Sumário

Capítulo 1

Equações lineares

1.1 Equação algébrica linear

Uma equação algébrica linear típica nas variáveis x_1, x_2 e x_3 é

$$x_1 + 2x_2 - 3x_3 = 5.$$

Resolvê-la significa determinar todos os valores reais para x_1, x_2 e x_3 que tornam verdadeira a igualdade. Neste caso, explicitando x_1 em relação a x_2 e x_3 na equação, obtemos $x_1 = 5 - 2x_2 + 3x_3$. Para qualquer x_2 e x_3 reais, basta tomar $x_1 = 5 - 2x_2 + 3x_3$ para obter uma solução. Neste exemplo, temos uma infinidade de soluções, onde podemos variar livremente x_2 e x_3.

De modo geral, dados os números reais $a_1, \ldots, a_n$ e b, uma equação da forma

$$a_1x_1 + \cdots + a_nx_n = b \tag{1.1}$$

é chamada de **equação algébrica linear** nas **variáveis** $x_1, x_2, \ldots, x_n$. As variáveis também são chamadas de **incógnitas** por serem os valores a serem determinados para valer a igualdade. Os números reais a_i são chamados de **coeficientes** e b é a **constante** da equação. A primeira incógnita com coeficiente não nulo é chamada de **variável principal** ou **incógnita principal** e as demais são chamadas de **variáveis livres**.

Uma matriz coluna real $v = [v_1, \ldots, v_n]^T$ é **solução** desta equação quando

$$a_1v_1 + \cdots + a_nv_n = b.$$

Diz-se ainda que a ênupla de números reais $(v_1, \ldots, v_n)$ **satisfaz** a equação.

Uma equação

$$0x_1 + \cdots + 0x_n = b,$$

em que todos os coeficientes são nulos é **degenerada**. Se b for igual a zero, então toda matriz coluna $[x_1, \ldots, x_n]^T$ é solução. Se b for diferente de zero, a equação degenerada não possui solução.

As equações não degeneradas com duas ou mais variáveis possui infinitas soluções. Uma equação não degenerado com uma única variável possui uma única solução.

Exemplo 1.1 *Para todo s real, a matriz coluna $[7+3s,\ 2s]^T$ é solução de $2x_1 - 3x_2 = 8$ que, portanto, possui infinitas soluções. A variável s que aparece neste exemplo é chamado de **parâmetro**.*

O conjunto de todas as soluções de uma equação é chamado **conjunto solução** ou **solução geral**. Cada elemento deste conjunto é, evidentemente, uma solução e, quando for conveniente, será chamado de **solução particular**.

Para determinar a solução geral de uma equação não degenerada $a_1x_1 + \cdots + a_nx_n = b$ basta explicitar a incógnita principal em função das variáveis livres.

Exemplo 1.2 *Para obter a solução geral de $x_1 - 7x_2 + x_3 = 1$, basta explicitar x_1 para obter $x_1 = 1 + 7x_2 - x_3$. A solução geral é o conjunto de matrizes coluna*

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

A equação

$$a_1x_1 + \cdots + a_nx_n = 0$$

é denominada de **equação homogênea**. Ela está associada à equação não homogênea (1.1) e, por esse motivo, é chamada de **equação homogênea associada** à equação não homogênea

$$a_1x_1 + \cdots + a_nx_n = b.$$

O uso de matrizes pode simplificar a notação. Sendo $a = [a_1,\ \ldots,\ a_n]^T$ a **matriz dos coeficientes** e $x = [x_1,\ \ldots,\ x_n]^T$ a **matriz das variáveis**, a equação acima pode ser colocada na forma

$$a^Tx = b.$$

Exemplo 1.3 *Consideremos novamente a equação do exemplo anterior $x_1 - 7x_2 + x_3 = 1$, cuja solução geral é*

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

É interessante observar que $[1,\ 0,\ 0]^T$ é solução da equação e que tanto $[7,\ 1,\ 0]^T$ quanto $[-1,\ 0,\ 1]^T$ são soluções da equação homogênea associada.

Este exemplo apresenta um fato geral.

Se $v_1, \ldots, v_p$ forem soluções da equação homogênea $a^Tx = 0$, então

$$c_1v_1 + \cdots + c_pv_p$$

continua sendo solução, para qualquer escolha dos números reais $c_1,\ \ldots,\ c_n$. Esta soma é chamada de **combinação linear** das matrizes $v_1, \ldots, v_p$.

Se um conjunto $\{v_1,\ \ldots,\ v_p\}$ de soluções da equação homogênea for tal que toda solução da equação homogênea é uma combinação linear dos seus elementos, diremos que ele é um **conjunto gerador** das soluções da equação homogênea.

Exemplo 1.4 *Explicitando x_1 na equação $x_1 - 3x_2 + x_3 = 0$, obtemos $x_1 = 3x_2 - x_3$ para daí obter todas as soluções desta equação*

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

Portanto, $[3,\ 1,\ 0]^T$ e $[-1,\ 0,\ 1]^T$ formam um conjunto gerador de soluções para a equação dada.

Se w_0 for uma solução da equação não homogênea $a^T x = b$ e v for uma solução da equação homogênea $Ax = 0$, então $w_0 + v$ é solução da equação não homogênea. Além disso, se w_1 for outra solução de $Ax = b$, então existe uma solução u de $Ax = 0$ tal que $w_1 = w_0 + u$. Esta solução u é exatamente $w_1 - w_0$.

Do parágrafo acima tiramos uma lição muito interessante. Conhecendo todas as soluções da homogênea e uma única solução da não homogênea, conheceremos todas as soluções da não homogênea.

1.2 Produto escalar

O produto matricial $a^T x$ é denominado de **produto escalar** das matrizes coluna a e x, sendo denotado por $\langle a, x \rangle$, isto é,

$$\langle a, x \rangle = a^T x.$$

Este conceito de produto escalar é importante e voltaremos a ele posteriormente.

Propriedades do produto escalar

Se x, y, z forem vetores coluna e k um número real,

1. $\langle x, x \rangle \geq 0$ e $\langle x, x \rangle = 0$ se e só se $x = 0$.
2. $\langle x, y \rangle = \langle y, x \rangle$
3. $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$
4. $\langle x, ky \rangle = k \langle x, y \rangle$

Usando o produto escalar, a equação (1.1) assume a forma

$$\langle a, x \rangle = b.$$

1.3 Sistemas de equações algébricas lineares

Um sistema de equações como

$$\begin{aligned} 3x_1 - 2x_2 &= 6 \\ x_1 + x_2 &= 7 \end{aligned}$$

é um sistema de equações algébricas lineares. Nos problemas onde estes sistemas ocorrem, o interesse se volta para a determinação dos valores de x_1 e x_2 que tornam verdadeiras as duas igualdades. Neste exemplo, para determiná-los, pode-se, por exemplo explicitar x_1 na segunda equação $x_1 = 7 - x_2$, substituir esta expressão no lugar de x_1 na primeira equação $3(7 - x_2) - 2x_2 = 6$ e expliciar x_2 obtendo $x_2 = 3$. Substituindo este valor na expressão de x_1 em função de x_2 obtemos $x_1 = 7 - x_2 = 7 - 3 = 4$. Portanto os valores de x_1 e x_2 que tornam verdadeiras as duas igualdades do sistema são $x_1 = 4$ e $x_2 = 3$.

Dados os números reais a_{ij} e b_i, com $i = 1, \ldots, m$ e $j = 1, \ldots, n$, o sistema de equações

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ \cdots &= \cdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

é chamado de **sistema de equações algébricas lineares** com m equações e n incógnitas. Os números a_{ij} são denominados **coeficientes** do sistema, b_i são os **termos constantes** e x_j são as **incógnitas** ou **variáveis** do sistema. Esta forma de apresentar o sistema é denominada de **forma padrão**.

Podemos simplificar a notação usando matrizes. Em

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \text{e} \quad b = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix},$$

denominamos A de **matriz dos coeficientes**, x de **matriz das incógnitas** e b de **matriz dos termos constantes** do sistema. Na forma matricial, o sistema se reduz a

$$Ax = b.$$

A matriz $[A \mid b]$ obtida acrescentando-se à matriz A uma coluna final com os elementos de b, é chamada de **matriz aumentada** do sistema linear.

Um vetor coluna real w tal que $Aw = b$ é chamado de **solução** do sistema $Ax = b$. Isto significa que w é solução de cada equação do sistema. Um sistema como este pode ter ou não uma solução.

Exemplo 1.5 *O sistema*

$$\begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

não possui solução pois não existem x_1 e x_2 que tornam verdadeira a segunda equação. A segunda equação do sistema é degenerada e seu segundo membro é diferente de zero.

O sistema

$$\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$$

possui uma única solução $x_1 = 2$ e $x_2 = 1$. Para obtê-la, basta observar que, da segunda equação $x_2 = 1$ e, da primeira, $x_1 + 2x_2 = 4$. Como $x_2 = 1$, devemos ter $x_1 = 2$.

O sistema

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

possui infinitas soluções. De fato, explicitano x_1 na primira equação segue $x_1 = 3 - 2x_2$. Substituindo esta expressão na segunda vem $2(3 - 2x_2) + 4x_2 = 6$ que se simplifica em $6 = 6$, ou seja, é sempre satisfeita. Logo, qualquer matrix coluna $[x_1,\, x_2]^T = [3 - 2x_2,\, x_2]^T$ é uma solução do sistema. A variável x_2 pode variar livremente nos reais.

O conjunto de todas as soluções do sistema é chamado de **conjunto solução** ou **solução geral** do sistema. Este conjunto pode ser vazio, ter um único elemento ou possuir infinitos elementos. O sistema de equações que não possui solução é chamado **incompatível**. Quando possui uma única solução é **compatível determinado** e, quando possui infinitas soluções, é chamado de **compatível indeterminado.**

O sistema de equações $Ax = 0$ é chamado de **homogêneo**. Quando $b \neq 0$, o sistema de equações $Ax = b$ é chamado de **não homogêneo**. Um sistema está intimamente ligado ao outro e, por esta razão, $Ax = 0$ é chamado de sistema homogêneo de equações **associado** ao sistema $Ax = b$.

A equação homogênea $Ax = 0$ possui sempre a solução **trivial** $x = 0$. Entretanto, quando o sistema homogêneo $Ax = 0$ possui uma solução v não trivial, ela possuirá infinitas soluções pois cv será solução para qualquer número real c.

Podemos ir um pouco além. Se $v_1, \ldots, v_p$ forem soluções do sistema homogêneo $Ax = 0$, então

$$c_1 v_1 + \cdots + c_p v_p$$

ainda será uma solução do sistema homogêneo para qualquer escolha dos números reais $c_1, \ldots, c_n$. A soma acima é chamada de **combinação linear** dos vetores $\{v_1, \ldots, v_p\}$. Se toda solução de $Ax = 0$ for uma combinação linear dos elementos deste conjunto, ele será chamado de **conjunto gerador** das soluções do sistema homogêneo $Ax = 0$.

Se v for uma solução de $Ax = 0$ e w_0 for uma solução de $Ax = b$, então $w_0 + v$ é solução de $Ax = b$. Se w_1 for outra solução de $Ax = b$, diferente de w_0, então $u = w_1 - w_0$ é solução de $Ax = 0$. Logo, qualquer solução w_1 do sistema $Ax = b$ é da forma $w_1 = w_0 + u$ onde u é solução da equação homogênea $Ax = 0$. Em outras palavras, conhecida uma solução w_0 de $Ax = b$, outra solução w_1 deste sistema é da forma $w_1 = w_0 + u$, onde u é solução do sistema homogêneo $Ax = 0$.

Ao conhecer uma única solução do sistema não homogêneo $Ax = b$ e a solução geral do sistema homogêneo $Ax = 0$, se conhece a solução geral do sistema não homogêneo.

O sistema não homogêneo $Ax = b$ pode ter uma solução ou não. Se a única solução do sistema homogêneo $Ax = 0$ for a trivial e $Ax = b$ tiver uma solução, ela será única. Quando $Ax = 0$ possuir solução não trivial e $Ax = b$ possuir uma solução, então possuirá infinitas outras.

Exemplo 1.6 *Considere o sistema*

$$\begin{bmatrix} 1 & -2 & 5 \\ 0 & 1 & -6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 7 \\ 3 \end{bmatrix}.$$

Explicitando x_2 na segunda equação, $x_2 = 3 + 6x_3$. Usando esta expressão de x_2 na primeira equação e explicitando x_1, segue $x_1 = 13 + 7x_3$. Logo, toda solução deste sistema é da forma

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 13 \\ 3 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 7 \\ 6 \\ 1 \end{bmatrix}$$

Observe que $[13,\ 3,\ 0]^T$ é uma solução particular do sistema e $[7,\ 6,\ 1]^T$ é solução do sistema homogêneo associado. O valor de x_3 poder variar livremente no conjunto dos números reais.

No exemplo anterior, as variáveis x_1 e x_2 foram expressas em termos de x_3. neste caso, chamamos x_1 e x_2 de **variáveis principais** e x_3 é a **variável livre**.

1.4 Sistema escalonado

Uma matriz **escalonada** é aquela em que

1. Todas as linhas nulas estão abaixo das linhas não nulas.

2. Numa linha não nula, o primeiro elemento não nulo é igual a 1. Este elemento é chamado de **pivô** ou **líder** da linha.

3. O pivô de uma linha está à direita do pivô da linha de cima.

Exemplo 1.7 *A matriz*

$$\begin{pmatrix} 1 & 0 & 3 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

é escalonada.

Um sistema $Ax = b$ é **escalonado** quando a matriz A for escalonada. As variáveis que multiplicam os pivôs são denominadas de **variáveis principais** e as demais de **variáveis livres** ou **parâmetros**.

Achar as soluções de um sistema escalonado é bastante simples. Podem aparecer equações degeneradas na parte inferior do sistema. Se uma dessas equações degeneradas possuir segundo membro não nulo, o sistema não possui solução. Se todos os segundos membros das equações degeneradas forem nulas, o sistema tem solução. Para obtê-las, podemos desconsiderar as equações degeneradas.

Eliminadas as equações degeneradas, explicitamos as variáveis principais de cada linha em função das demais, começando na última linha e retornando até a primeira. A partir da penúltima equação use as variáveis principais já explicitadas para colocar a variável principal daquela equação em termos das variáveis livres. Com este processo obtém-se todas as variáveis principais em termos das variáveis livres. Esta técnica de solução é denominada de **substituição reversa**.

Exemplo 1.8 *O sistema*

$$\begin{pmatrix} 1 & 0 & 2 & -1 \\ 0 & 1 & 3 & 5 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -3 \\ 0 \\ 8 \\ 0 \end{pmatrix}$$

é escalonado. As variáveis x_1, x_2 e x_4 são as variáveis prinicipais e x_3 é a variável livre. A última equação é degenerada mas compatível pois o segundo membro também é nulo. O sistema possui solução e esta última equação pode ser desconsidereda uma vez que qualquer matriz coluna real $[x_1,\ x_2,\ x_3,\ x_3]^T$ é uma solução. Eliminada esta equação, a terceira passa a ser a última, onde explicitamos $x_4 = 8$. Da segunda, explicitamos $x_2 = -3x_3\ -5x_4$. Usando o valor de x_4 determinado na etapa anterior, obtemos $x_2 = -3x_3\ -40$. Na primeira, explicitamos $x_1 = -3\ -2x_3\ +x_4$. Usando o valor de x_4 determinado anteriormente, obtemos $x_1 = -3\ -2x_3\ +8 = 5\ -2x_3$. Colocamos as três variáveis principais x_1, x_2 e x_4 em função da variável livre x_3. A solução geral do sistema será

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 5 - 2x_3 \\ -40 - 3x_3 \\ x_3 \\ 8 \end{bmatrix} = \begin{bmatrix} 5 \\ -40 \\ 0 \\ 8 \end{bmatrix} + x_3 \begin{bmatrix} -2 \\ -3 \\ 1 \\ 0 \end{bmatrix}$$

onde a variável livre x_3 pode assumir qualquer valor real. É interessante observar que $[-2,\ -3,\ 1,\ 0]^T$ é solução do sistema homogêneo associado $Ax = 0$.

Uma matriz A de tamanho $m \times n$ é **escalonada reduzida** se for escalonada e cada pivô é o único elemento não nulo em sua coluna. Neste caso, o sistema $Ax = b$ é denominado de **sistema escalonado reduzido**.

Exemplo 1.9 *O sistema*

$$\begin{pmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -3 \\ 0 \\ 0 \end{pmatrix}$$

é escalonado reduzido. As variáveis x_1 e x_3 são principais e x_2 e x_4 são livres. A última equação é degenerada mas compatível. O método da substituição reversa nos fornece $x_3 = -x_4$ e $x_1 = -3\ -2x_2\ -3x_4$, onde as variáveis principais estão em função das variáveis livres.

Algoritmo da substituição reversa

Este algoritmo resolve o sistema $Rx = b$ pelo método da substituição reversa, onde R é quadrada, inversível e triangular superior. Isto significa que

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ & r_{22} & \cdots & r_{2m} \\ & & \ddots & \vdots \\ & & & r_{mm} \end{bmatrix}$$

com $r_{ii} \neq 0$, para $i = 1, \ldots, m$. Para resolver o sistema $Rx = b$, iniciamos explicitando x_m na última equação e, retornando até a primeira, explicitando as variáveis principais de cada equação em função das variáveis determinadas nas etapas anteriores. Assim,

$$\begin{aligned} x_m &= b_m / r_{mm} \\ x_{m-1} &= \left(b_{m-1} - r_{m-1,m} x_m\right) / r_{m-1,m-1} \\ x_{m-2} &= \left(b_{m-2} - r_{m-2,m-1} x_{m-1} - r_{m-2,m} x_m\right) / r_{m-2,m-2} \end{aligned}$$

e assim por diante. O caso geral, em que $j = m-1,\ m-2,\ \ldots,\ 1$, assume a forma

$$x_j = \left(b_j - \sum_{k=j+1}^{m} r_{jk} x_k \right) \Bigg/ r_{m-j,m-j}$$

===================================

Entrada: Matriz R de tamanho $m \times m$ e matriz b de tamanho $m \times 1$.
Saída: Matriz x de tamanho $m \times 1$.

===================================

```
x = b ;
x(m) = b(m) / R(m,m);
for j = m-1:-1:1
  x(j) = ( b(j) - R(j, j+1:m) * x(j+1:m) ) / R(j,j);
end
```

===================================

1.5 Sistema inferiormente escalonado

Um procedimento semelhante pode ser adotado para matrizes $m \times n$ **inferiormente escalonadas**, que são aquelas com as seguintes características:

1. Se existirem linhas nulas, elas se localizam na parte inferior da matriz.
2. O último elemento não nulo de uma linha é igual a 1, sendo denominado de **pivô** ou **lider** da linha.
3. O pivô de uma linha se encontra à direita do pivô da linha anterior.

Quando A for escalonada inferiormente, o sistema $Ax = b$ é chamado de **sistema inferiormente escalonado**. As variáveis que multiplicam os pivôs são denominadas de **principais** e as demais são denominadas **livres**. Se as equações degeneradas deste sistema forem compatíveis, o sistema possui solução que pode ser obtida pelo processo de **substituição direta**. Primeiro, descartam-se as equações degeneradas. Em seguida, a partir da primeira equação, explicita-se a variável principal em função das variáveis livres. A partir da segunda, prossiga até a última, explicitando a variável principal daquela equação em função das demais, usando as expressões das variáveis principais obtidas anteriormente para explicitar a variável principal em função das variáveis livres apenas.

Uma matriz A de tamanho $m \times n$ é **inferiormente escalonada reduzida** quando for inferiormente escalonada e cada pivô for o único elemento não nulo em sua coluna. Neste caso, o sistema $Ax = b$ é denominado de **sistema inferiormente escalonado reduzido**. Tais sistemas, quando compatíveis, são facilmente resolvidos pelo processo de substituição direta.

Algoritmo da substituição direta

Este algoritmo resolve o sistema $Rx = b$ pelo método da substituição reversa, onde R é quadrada, inversível e triangular inferior. Isto significa que

$$R = \begin{bmatrix} r_{11} & & & \\ r_{21} & r_{22} & & \\ \vdots & & \ddots & \\ r_{m1} & r_{m2} & \cdots & r_{mm} \end{bmatrix}$$

com $r_{ii} \neq 0$, para $i = 1, \ldots, m$. Para resolver o sistema $Rx = b$, iniciamos explicitando x_1 na primeira equação e, prosseguindo até a última, vamos explicitando as variáveis principais de cada equação em função das variáveis determinadas nas etapas anteriores. Assim,

$$\begin{aligned} x_1 &= b_1/r_{11} \\ x_2 &= (b_2 - r_{21}x_1)/r_{22} \\ x_3 &= (b_3 - r_{31}x_1 - r_{32}x_2)/r_{3,3} \end{aligned}$$

e assim por diante. O caso geral, em que $j = 2, 3, \ldots, m$, assume a forma

$$x_j = \left(b_j - \sum_{k=1}^{j-1} r_{jk} x_k \right) \bigg/ r_{jj}$$

Algoritmo da substituição direta

Este algoritmo resolve pelo método da substituição direta um sistema $Rx = b$, onde R é uma matriz quadrada $m \times m$, triangular inferior, inversível e b é uma matriz coluna $m \times 1$.

==================================

Entrada: Matrizes R e b.

Saída: Matriz x, solução dos sistema $Rx = b$.

```
==================================
x = b ;
x(1) = b(1) / R(1,1);
for j = 2:m
  x(j) = ( b(j) - R(j, 1:j-1) * x(1:j-1) ) / R(j,j);
end
==================================
```

1.6 Sistemas equivalentes

Uma técnica muito usada para resolver sistemas de equações lineares consiste em realizar transformações sobre o sistema original até se chegar a um sistema escalonado cuja solução é simples. Para que esta técnica se torne efetiva, as transformações não podem alterar o conjunto solução do sistema.

Definição 1.10 *Dois sistemas $Ax = b$ e $Bx = c$ são* ***equivalentes*** *quando ambos possuem o mesmo conjunto solução.*

Existem operações, denominadas **elementares** que, ao serem aplicadas a um sistema, preserva suas soluções, transformando-o em outro sistema equivalente. As operações elementares são:

1. Permutar a ordem de duas equações.

2. Multiplicar uma equação por uma constante não nula.

3. Adicionar a uma equação um múltiplo de outra.

Num sistema de equações, podemos enumerá-las: equação 1, 2, ..., m. Sejam i e j números inteiros entre 1 e n.

O operação que permuta as equações i e j será denotada por $O(l_i \leftrightarrow l_j)$, a operação que multiplica a equação i por um número r não nulo será denotada por $O(rl_i)$ e a operação que consiste em adicionar à equação i um múltiplo r de outra equação j será denotada por $O(l_i + rl_j)$.

As operações elementares são reversíveis. A operação $O(l_i \leftrightarrow l_j)$ pode ser revertida aplicando novamente esta mesma operação. A operação $O(rl_i)$ pode ser revertida aplicando a operação $O(r^{-1}l_i)$ e a operação $O(l_i+rl_j)$ pode ser revertida aplicando a operação $O(l_i - rl_j)$.

Vamos mostrar que essas transformações levam o sistema original em outro equivalente. Façamos a prova para um caso particular que representa o caso geral.

Se $[x_1,\, x_2,\, x_3]^T$ for uma solução do sistema

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= b_3 \end{aligned} \tag{1.2}$$

e r for um número real, então vale ainda a igualdade

$$\begin{aligned} (a_{11} + ra_{21})x_1 + (a_{12} + ra_{22})x_2 + (a_{13} + ra_{23})x_1 &= \\ a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + r(a_{21}x_1 + a_{22}x_2 + a_{23}x_3) &= b_1 + rb_2 \end{aligned}$$

mostrando que $[x_1,\, x_2,\, x_3]^T$ é solução do sistema

$$\begin{aligned} (a_{11} + ra_{21})x_1 + (a_{12} + ra_{22})x_2 + (a_{13} + ra_{23})x_1 &= b_1 + rb_2 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= b_3 \end{aligned} \tag{1.3}$$

Isto significa que as soluções do sistema (1.2) são soluções do sistema (1.3) que foi obtido do original a partir da transformação elementar $O(l_1 + rl_2)$. Logo, as soluções de (1.3) são soluções de (1.2) pois esta pode ser obtida daquela pela operação $O(l_1 - rl_2)$. Concluímos que os sistemas original e o transformado são equivalentes.

De modo semelhante se pode provar que as outras operações elementares transformam um sistema em outro equivalente.

1.7 O método da eliminação de Gauss

O método de Gauss consiste em realisar operações elementares sobre linhas no sistema $Ax = b$, transformando-o num sistema escalonado equivalente e resolvendo-o por **substituição reversa**.

Como a matriz A dos coeficientes e a matriz b das constantes contêm todas as informações necessárias para montar o sistema, vamos considerar a **matriz completa** do sistema, obtida ao acrescentar a coluna b à direita de A. Esta matriz será denotada por $[A\ b]$. A realização de operações elementares sobre as equações é equivalente à realização de operações elementares sobre as linhas da matriz completa.

Vamos escreve $A \to R$ quando for possível levar A em R efetuando operações elementares sobre as linhas de A. Se R for escalonada, diremos que ela é a **forma escalonada** de A. Se R for escalonada reduzida, diremos que ela é a **forma escalonada reduzida** de A. Pode-se provar que a forma escalonada reduzida de uma matriz é única.

O processo de Gauss para resolver um sistema $Ax = b$ é descrito pelo algoritmo abaixo, realizado sobre a matriz completa $[A\ b]$.

Passo 1. Se $A = 0$, encerre o algoritmo. O sistema já é escalonado.

Passo 2. Percorra as colunas da matriz completa $[A\ b]$ da esquerda para a direita, localizando a primeira não nula.

Passo 3. Percorra esta coluna de cima para baixo, localizando seu primeiro elemento não nulo. Seja p o valor deste elemento.

Passo 4. Permute esta linha com a primeira.

Passo 5. Multiplique a atual primeira linha por p^{-1}, fazendo com que o primeiro elemento não nulo da primeira linha fique igual a 1. Este será o **pivô** da primeira linha.

A partir deste ponto, a primeira linha não sofrerá outras modificações.

Passo 6. Passe à segunda linha, tranformando-a na primeira da próxima etapa.

Passo 7. Repita os passos de 1 a 6 com todas as linhas restantes.

Com este algoritmo, partimos da matriz $[A\ b]$ e chegamos à matriz $[R\ c]$, onde R é a forma escalonada de A. O sistema $Rx = c$ é equivalente ao original.

Se existirem equações degeneradas incompatíveis no sistema $Rx = c$, então o sistema $Ax = b$ não tem solução.

Se todas as equações degeneradas de $Rx = c$ forem compatíveis, o sistema $Ax = b$ tem solução. Exclua as equações degeneradas e use a substituição reversa para obter as soluções do sistema euqivalente $Rx = c$. Estas soluções possuirão a forma

$$x = w_0 + c_1 v_1 + \cdots + c_r v_r$$

onde w_0 é uma solução de $Rx = c$ e $v_1, \ldots, v_r$ são soluções do sistema homogêneo associado $Rx = 0$. Os números reais c_i são arbitrários e relacionados com as variáveis livres. Os números reais $c_1, \ldots, c_r$ são denominados de **parâmetros**.

O número de pivôs de R é igual ao número de linhas não nulas de R. Se existirem k pivôs, este será o número de variáveis principais do sistema. Se o número de incógnitas do sistema for n, o número de variáveis livres será $n - k$.

Se R for escalonada e A pode ser levada em R por transformações elementares, o número de pivôs de R é chamado de **posto** da matriz A.

Exemplo 1.11 *Considere o sistema*

$$\begin{aligned} x + y - 2z &= 0 \\ 2x + 2y - 3z &= 2 \\ 3x - y + 2z &= 12 \end{aligned}$$

cuja matriz aumentada é

$$\begin{pmatrix} 1 & 1 & -2 & 0 \\ 2 & 2 & -3 & 2 \\ 3 & -1 & 2 & 12 \end{pmatrix}$$

Realizando as operações $O(l_2 = l_2 - 2l_1)$ *e* $O(l_3 = l_3 - 3l_1)$ *sobre a matriz chegamos em*

$$\begin{pmatrix} 1 & 1 & -2 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & -4 & 8 & 12 \end{pmatrix}.$$

Realizando a operação $O(l_2 \leftrightarrow l_3)$ *segue*

$$\begin{pmatrix} 1 & 1 & -2 & 0 \\ 0 & -4 & 8 & 12 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

que é uma matriz diagonal superior. Encerramos a primeira etapa do método de eliminação de Gauss.

Para completar o método, caminhando de baixo para cima e da esquerda para a direita, anulamos os elementos nas colunas acima da diagonal principal. Com as operações $O(l_2 = l_2 - 8l_3)$ *e* $O(l_1 = l_1 + 2l_3)$, *obtemos*

$$\begin{pmatrix} 1 & 1 & 0 & 4 \\ 0 & -4 & 0 & -4 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

Com as operações $O(l_2 = -(1/4)l_2)$ *seguida de* $O(l_1 = l_1 - l_2)$ *chegamos à matriz*

$$\begin{pmatrix} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

A matriz A foi transformada até se tornar uma matriz identidade. Agora, obter a solução do problema é trivial $x = 3$, $y = 1$ *e* $z = 2$.

1.8 Matrizes inversas

Uma matriz quadrada A de tamanho $m \times m$ é inversível quando existir uma matriz quadrada B de tamanho $m \times m$ tal que $AB = BA = I_m$ onde I_m é a matriz identidade de ordem m. A matriz B é chamada de **inversa** de A e é denotada por A^{-1}.

Pela própria definição, a matriz B também é inversível e sua inversa é A. Assim, $B^{-1} = A$ e

$$\left(A^{-1}\right)^{-1} = A.$$

Se A e B forem inversíveis, então AB são inversíveis e $(AB)^{-1} = B^{-1}\ A^{-1}$. Se uma matriz for triangular inferior, sua inversa também será triangular inferior e, quando ela for triangular superior, sua inversa também será triangular superior.

Teorema 1.12 *Seja A uma matriz quadrada. São equivalentes as afirmações:*

1. *A é inversível.*

2. *O sistema homogêneo $Ax = 0$ possui apenas a solução trivial $x = 0$.*

3. *A forma escalonada reduzida de A é a matriz identidade.*

4. *O sistema $Ax = b$ possui uma única solução para cada matriz coluna b.*

5. *Existe uma matriz quadrada B tal que $AB = I$.*

Prova. (1) $\Longrightarrow$ (2) pois, se A é inversível e $Ax = 0$, então $A^{-1}Ax = 0$ o que implica em $x = 0$.

(2) $\Longrightarrow$ (3) pois, se a forma escalonada reduzida R de A não for a matriz identidade, uma de suas linhas é nula pois R é quadrada. Portanto, $Rx = 0$ tem soluções não nulas. Se este fosse o caso, o sistema $Ax = 0$ teria soluções não nulas, contrariando (2).

(3) $\Longrightarrow$ (4) pois, se $A \to I$ então o sistema $Ax = b$ é equivalente ao sistema $Ix = c$, para alguma matriz coluna c, cuja solução é $x = c$, mostrando que o sistema $Ax = b$ tem sempre uma única solução para cada b.

(4) $\Longrightarrow$ (5) pois, sendo e_j a coluna j da matriz identidade I, o sistema $Ax = e_j$ tem uma única solução $x = b_j$ para $j = 1, 2, \ldots, n$. Sendo $B = [b_1, \ldots, b_n]$, obtemos $AB = I$.

(5) $\Longrightarrow$ (1) pois, se $AB = I$ e $Bx = 0$, então $ABx = 0$ ou $Ix = 0$ o que implica em $x = 0$. Logo, a condição (2) vale para B no lugar de A e, consequentemente, valem (3) e (4) com B no lugar de A. Logo, pela parte (5), existe uma matriz C tal que $BC = I$. Como $C = IC = (AB)C = A(BC) = A$, obtemos $BA = I$. Como $AB = I$ por hipótese, provamos que A é inversível. $\square$

Corolário 1.13 *Sejam A e B matrizes quadradas $m \times m$. Se $AB = I$, então A e B são inversíveis e uma é a inversa da outra.*

Prova. Se $AB = I$, provamos que A é inversível e que B é a inversa de A. Logo, B é inversível e sua inversa é A. $\square$

Este corolário garante que $AB = I$ é o bastante para garantir que A e B são inversíveis, sendo uma a inversa da outra.

Corolário 1.14 *Se $A = BC$ for inversível, então B e C são inversíveis.*

Prova. Sendo $A = BC$ inversível, $(A^{-1}B)\,C = A^{-1}(BC) = A^{-1}A = I$ e assim C é inversível. Por outro lado, $B(CA^{-1}) = (BC)A^{-1} = AA^{-1} = I$ e B é inversível. $\square$

1.9 Matrizes elementares

As **matrizes elementares** são aquelas obtidas a partir da identidade mediante uma única operação elementar. Vamos denotar por $E(l_i \longleftrightarrow l_j)$ a matriz elementar obtida a partir da identidade pela permuta das linhas i e j. A matriz $E(l_i + rl_j)$ denotará a matriz elementar obtida da identidade adicionando à linha i um múltiplo r da linha j. Se r é um número não nulo, $E(rl_i)$ denotará a matriz elementar obtida da identidade multiplicando sua linha i por r.

Exemplo 1.15 *As matrizes abaixo são elementares*

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 7 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix}$$

sendo, respectivamente, as matrizes $E(l_1 \leftrightarrow l_3)$, $E(7l_1)$ *e* $E(l_3 + 3l_1)$.

Os produtos

$$E(l_i \longleftrightarrow l_j)A \ , \quad E(rl_i)A \ , \quad E(l_i + rl_j)A$$

realizam sobre A as operações elementares $O(l_i \longleftrightarrow l_j)$, $O(rl_i)$, e $O(l_i + rl_j)$, respectivamente.

Exemplo 1.16 *Os produtos abaixo ilustram as afirmações acima. Seja*

$$A = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}.$$

O produto

$$E(l_1 \leftrightarrow l_3)A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} = \begin{bmatrix} c_1 & c_2 & c_3 \\ b_1 & b_2 & b_3 \\ a_1 & a_2 & a_3 \end{bmatrix}$$

permuta a primeira com a terceira linha de A. O produto

$$E(7l_1)A = \begin{bmatrix} 7 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} = \begin{bmatrix} 7a_1 & 7a_2 & 7a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}$$

multiplica a primeira linha de A por 7. O produto

$$E(l_3 + 5l_1)A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 5 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ 5a_1 + c_1 & 5a_2 + c_2 & 5a_3 + c_3 \end{bmatrix}$$

adiciona à terceira linha de A o quíntuplo de sua primeira linha.

As matrizes elementares são inversíveis. A inversa de $E(l_i + rl_j)$ é $E(l_i - rl_j)$, a inversa de $E(l_i \leftrightarrow l_j)$ é ela mesma e, para $r \neq 0$, a inversa de $E(rl_i)$ é $E((1/r)l_i)$.

Há um teorema muito interessante relacionando matrizes inversíveis com matrizes elementares.

Teorema 1.17 *Uma matriz quadrada é inversível se e só se for igual a um produto de matrizes elementares.*

Prova. Se uma matriz quadrada for o produto de matrizes elementares, ela é inversível pois cada matriz elementar é inversível.

Se A for inversível, então o teorema 1.12 garante que a forma escalonada reduzida de A é a matriz identidade. Em consequência,, existem matrizes elementares E_1, E_2, $\ldots$, E_k tais que $E_k \cdots E_2\ E_1\ A = I$. Neste caso, $A = E_1^{-1}\ E_2^{-1} \cdots E_k^{-1}$. Como as inversas de matrizes elementares são elementares, segue que A é o produto de matrizes elementares. $\square$

Em termos de matrizes elementares, o método da eliminação de Gauss usado para resolver o sistema $Ax = b$ pode ser descrito nos seguintes termos: multiplicamos os dois lados do sistema sucessivamente por matrizes elementares E_1, E_2, $\ldots$, E_k

$$E_k \cdots E_2 E_1 A x = E_k \cdots E_2 E_1 b$$

executando operações elementares sobre as linhas de A, até obter a matriz escalonada

$$U = E_k \cdots E_2 E_1 A.$$

Sendo $E = E_k \cdots\ E_2\ E_1$, o sistema original se transforma no sistema equivalente escalonado

$$Ux = Eb$$

que terá solução se linhas nulas em U corresponderem a linhas nulas em Eb. Quando este for o caso, o sistema é resolvido por substituição reversa.

Se existirem linhas nulas em U e as linhas correspondentes de Eb não forem nulas, o sistema não tem solução, é incompatível.

Exemplo 1.18 *Vamos usar transformações elementares para obter a forma escalonada de*

$$A = \begin{bmatrix} 5 & 1 & 3 \\ 10 & 5 & 8 \\ 15 & 6 & 16 \end{bmatrix}.$$

Em lugar de executar uma operação elementar por vez, vamos executar as operações lineares necessárias para anular todos os elementos de cada coluna abaixo da diagonal principal. Efetuando o produto

$$E_1 A = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 5 & 1 & 3 \\ 10 & 5 & 8 \\ 15 & 6 & 16 \end{bmatrix} = \begin{bmatrix} 5 & 1 & 3 \\ 0 & 3 & 2 \\ 0 & 3 & 7 \end{bmatrix}$$

obtemos uma matriz onde os elementos da primeira linha abaixo da diagonal principal são nulos. A matriz E_1 não é elementar mas é o produto de duas matrizes elementares

$$E_1 = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}.$$

Efetuando o produto de E_1A pela matriz elementar E_3 definida abaixo, obtemos

$$E_2E_1A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 5 & 1 & 3 \\ 0 & 3 & 2 \\ 0 & 3 & 7 \end{bmatrix} = \begin{bmatrix} 5 & 1 & 3 \\ 0 & 3 & 2 \\ 0 & 0 & 5 \end{bmatrix}$$

que é triangular superior. Denotemos por U esta matriz, de modo que E_2 E_1 $A = U$. A matriz

$$E_2E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix}$$

é triangular inferior, os elementos de sua diagonal principal é unitária e sua inversa é

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

Um fato notável desta inversa reside no fato de ser exatamente igual ao produto E_2 E_1, onde os elementos abaixo da diagonal principal aparecem com os sinais trocados. Assim,

$$E_2E_1A = U = \begin{bmatrix} 5 & 1 & 3 \\ 0 & 3 & 2 \\ 0 & 0 & 5 \end{bmatrix}.$$

1.10 Cálculo da inversa

Podemos completar o processo iniciado no exemplo da seção anterior até obter a inversa de A.

Exemplo 1.19 *No exemplo anterior, multiplicamos*

$$A = \begin{bmatrix} 5 & 1 & 3 \\ 10 & 5 & 8 \\ 15 & 6 & 16 \end{bmatrix} \quad e \quad E_2E_1 = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix},$$

para obter

$$U = E_2E_1A = \begin{bmatrix} 5 & 1 & 3 \\ 0 & 3 & 2 \\ 0 & 0 & 5 \end{bmatrix}.$$

Podemos multiplicar U por matrizes elementares até obter a inversa de A. Efetuando o produto

$$E_3U = \begin{bmatrix} 1 & 0 & -3/5 \\ 0 & 1 & -2/5 \\ 0 & 0 & 1/5 \end{bmatrix} \begin{bmatrix} 5 & 1 & 3 \\ 0 & 3 & 2 \\ 0 & 0 & 5 \end{bmatrix} = \begin{bmatrix} 5 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

anulamos os elementos da terceira coluna acima da diagonal principal. A matriz E_3 é o produto de três matrizes elementares

$$E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/5 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -2/5 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -3/5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Em seguida, efetuamos o seguinte produto

$$E_4E_3U = \begin{bmatrix} 1 & -1/3 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 5 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

onde E_4 é o produto de duas matrizes elementares

$$E_4 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1/3 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Finalmente, multiplicando E_4E_3U pela matriz elementar

$$E_5 = \begin{bmatrix} 1/5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

obtemos

$$E_5E_4E_3U = E_5E_4E_3E_2E_1A = I$$

onde I é a matriz identidade. O produto $E_5E_4E_3E_2E_1$ é a inversa procurada

$$A^{-1} = \begin{bmatrix} \frac{32}{75} & \frac{2}{75} & -\frac{7}{75} \\ -\frac{8}{15} & \frac{7}{15} & -\frac{2}{15} \\ -\frac{1}{5} & -\frac{1}{5} & \frac{1}{5} \end{bmatrix}.$$

Este exemplo é típico do **método de Gauss-Jordan** para determinar a inversa de uma matriz A. Se E for o produto de matrizes elementares para as quais $EA = I$, então $A^{-1} = E$. Esta observação nos permite construir o seguinte algoritmo: tome a matriz aumentada $[A\ I]$ e realize operações elementares sobre ela obtendo a matriz $[EA\ EI]$. No momento que EA for igual á identidade, $EI = E$ será a inversa A^{-1} de A.

Se nalgum ponto deste processo chegarmos a uma matriz aumentada $[EA\ EI]$ com linha nula, concluímos que A não tem inversa.

Exemplo 1.20 *Vamos usar o método de Gauss-Jordan para obter a inversa de*

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 3 & 4 \\ 2 & 7 & 12 \end{bmatrix}.$$

Inicialmente formamos a matriz aumentada

$$[A \quad I] = \begin{bmatrix} 1 & 2 & 1 & 1 & 0 & 0 \\ 1 & 3 & 4 & 0 & 1 & 0 \\ 2 & 7 & 12 & 0 & 0 & 1 \end{bmatrix}$$

e realizamos operações elementares sobre linha até chegar a $(I \mid A^{-1})$. *Em lugar de aplicar uma transformação elementar por vez, vamos aplicar um produto de transformações lineares que agirão sobre toda uma coluna.*

$$\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 & 1 & 0 & 0 \\ 1 & 3 & 4 & 0 & 1 & 0 \\ 2 & 7 & 12 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 1 & 3 & -1 & 1 & 0 \\ 0 & 3 & 10 & -2 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 1 & 3 & -1 & 1 & 0 \\ 0 & 3 & 10 & -2 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 1 & 3 & -1 & 1 & 0 \\ 0 & 0 & 1 & 1 & -3 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -3 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 1 & 3 & -1 & 1 & 0 \\ 0 & 0 & 1 & 1 & -3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 & 0 & 3 & -1 \\ 0 & 1 & 0 & -4 & 10 & -3 \\ 0 & 0 & 1 & 1 & -3 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 & 0 & 3 & -1 \\ 0 & 1 & 0 & -4 & 10 & -3 \\ 0 & 0 & 1 & 1 & -3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 8 & -17 & 5 \\ 0 & 1 & 0 & -4 & 10 & -3 \\ 0 & 0 & 1 & 1 & -3 & 1 \end{bmatrix}$$

Logo, a inversa de A *é*

$$\begin{bmatrix} 8 & -17 & 5 \\ -4 & 10 & -3 \\ 1 & -3 & 1 \end{bmatrix}.$$

1.11 Fatoração LU

Multiplicando uma matriz A por matrizes elementares, podemos chegar a uma matriz U triangular superior.

Para descrever o processo, vamos ampliar um pouco nosso conceito de matriz elementar e também denominar de elementar aquelas matrizes obtidas a partir da identidade permitindo que os elementos de uma única coluna abaixo da diagonal principal sejam diferentes de zero.

Exemplo 1.21 *Neste conceito ampliado, a matriz*

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

é elementar. Ela é o produto de duas matrizes elementares

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

Podemos usar essas matrizes elementares para zerar todos os elementos abaixo da diagonal principal de uma coluna de uma matriz A.

Exemplo 1.22 *Este exemplo ilustra o anulamento de todos os elementos de abaixo da diagonal da primeira coluna de uma matriz mediante o uso de uma matriz elementar*

$$\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -6 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 6 & -1 \\ 4 & 3 & 8 \\ 12 & 7 & 9 \end{bmatrix} = \begin{bmatrix} 2 & 6 & -1 \\ 0 & -9 & 10 \\ 0 & -29 & 15 \end{bmatrix}.$$

Seja A uma matriz $m \times m$. Sejam $E_1, \ldots, E_{m-1}$ matrizes elementares que, multiplicadas à esquerda de A a levam numa matriz triangular superior U. Os elementos abaixo da diagonal principal da primeira coluna de $E_1 A$ são nulos. Os elementos abaixo da diagonal principal da primeira e segunda colunas de $E_2 E_1 A$ são nulos e assim por diante. Este procedimento resulta em

$$E_k \cdots E_1 A = U$$

onde U é triangular superior. A matriz

$$E = E_k \cdots E_1$$

é triangular inferior e os elementos da diagonal principal são todos iguais a 1. Sua inversa L também é triangular inferior e os elementos da diagonal principal são todos iguais a 1. Com isto, obtemos a fatoração

$$A = LU$$

onde U é uma matriz triangular superior e L é uma matriz triangular inferior inversível, cujos elementos da diagonal principal são iguais a 1.

Exemplo 1.23 *Vamos obter a decomposição LU da matriz*

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & 5 \\ 4 & -1 & 13 \end{bmatrix}.$$

Efetuando o produto

$$E_1A = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & 5 \\ 4 & -1 & 13 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 \\ 0 & -3 & 5 \\ 0 & -9 & 13 \end{bmatrix}$$

obtemos uma matriz cujos elementos abaixo da diagonal principal da primeira coluna são iguais a zero. Agora, efetuando o produto

$$E_2E_1A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 \\ 0 & -3 & 5 \\ 0 & -9 & 13 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 \\ 0 & -3 & 5 \\ 0 & 0 & -2 \end{bmatrix}$$

obtemos a forma escalonada de A. Para chegar à decomposição LU, basta calcular a inversa $L = E_1^{-1}\ E_2^{-2}$. As matrizes E_1 e E_2 nas multiplicações acima são elementares

$$E_1 = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix} \qquad e \qquad E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{bmatrix}$$

e pode-se verificar que

$$E_1^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 4 & 0 & 1 \end{bmatrix} \qquad e \qquad E_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 1 \end{bmatrix}$$

Oberve um fato interessante: para obter as inversas de E_1 e E_2, basta trocar os sinais dos elementos não nulos abaixo da diagonal principal. Em seguida, efetuando o produto

$$L = E_1^{-1}E_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 4 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 4 & 3 & 1 \end{bmatrix}$$

percebemos outro fato fantástico: para obter o produto L, basta colocar na matriz identidade os elementos não nulos de E_1^{-1} e E_2^{-1} nos seus devidos lugares. Agora tem-se a decomposição LU de A

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 4 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 \\ 0 & -3 & 5 \\ 0 & 0 & -2 \end{bmatrix}.$$

O fato ocorrido no cálculo de L do exemplo anterior não é fortuito e sim um resultado geral.

Para provar esta afirmação baseando-nos no livro de Trefethen e Bau.

As matrizes L e U da decomposição LU de A podem ser obtidas pelo método de eliminação de Gauss. Como é raro aplicar este método a matrizes retangulares, vamos descrevê-lo para matrizes quadradas A pertencentes a $\mathbf{C}^{m\times m}$. Considere a matriz

$$A = \begin{bmatrix} x & x & x & x \\ x & x & x & x \\ x & x & x & x \\ x & x & x & x \end{bmatrix}$$

onde o x indica números quaisquer. Façamos a primeira transformação

$$L_1 A = \begin{bmatrix} x & x & x & x \\ \mathbf{0} & \mathbf{x} & \mathbf{x} & \mathbf{x} \\ \mathbf{0} & \mathbf{x} & \mathbf{x} & \mathbf{x} \\ \mathbf{0} & \mathbf{x} & \mathbf{x} & \mathbf{x} \end{bmatrix}$$

zerando os elementos abaixo da diagonal principal da primeira coluna. Façamos a segunda transformação

$$L_2 L_1 A = \begin{bmatrix} x & x & x & x \\ 0 & x & x & x \\ 0 & \mathbf{0} & \mathbf{x} & \mathbf{x} \\ 0 & \mathbf{0} & \mathbf{x} & \mathbf{x} \end{bmatrix}$$

zerando os elementos abaixo da diagonal principal da segunda coluna. Finalmente, com a terceira transformação,

$$L_3 L_2 L_1 A = \begin{bmatrix} x & x & x & x \\ 0 & x & x & x \\ 0 & 0 & x & x \\ 0 & 0 & \mathbf{0} & \mathbf{x} \end{bmatrix} = U$$

zeramos o elemento abaixo da diagonal principal da terceira coluna, obtendo assim a matriz U. O negrito indica os elementos que foram modificados na transformação. Vejamos um caso concreto.

Exemplo 1.24 *A decomposição LU de*

$$A = \begin{bmatrix} 1 & 3 & 2 & 0 \\ 2 & 7 & 5 & 1 \\ 1 & 5 & 5 & 4 \\ 0 & 3 & 4 & 6 \end{bmatrix}$$

é

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

que foi obtida multiplicando A pela esquerda por

$$L_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \; L_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -2 & 1 & 0 \\ 0 & -3 & 0 & 1 \end{bmatrix}, \; L_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

As inversas de L_1, L_2 e L_3 são

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \; \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 3 & 0 & 1 \end{bmatrix}, \; \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix},$$

e podem ser obtidas de L_1, L_2 e L_3 trocando o sinal dos elementos não nulos abaixo da diagonal principal. Ainda

$$L = L_1^{-1} L_2^{-1} L_3^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 3 & 1 & 1 \end{bmatrix}$$

é obtido a partir de L_1^{-1}, L_2^{-1} e L_3^{-1} simplesmente colocando na matriz identidade os termos não nulos dessas três matrizes em suas respectivos posições.

Fórmulas gerais e dois golpes de sorte

Seja A uma matriz $m \times m$ e denote por X a matriz obtida depois de $k-1$ passo de eliminação. Denote por x_k a coluna k de X no início do passo k. A transformação L_k deve ser escolhida de modo que

$$x_k = \begin{bmatrix} x_{1k} \\ \vdots \\ x_{kk} \\ x_{k+1,k} \\ \vdots \\ x_{m,k} \end{bmatrix} \rightarrow L_k x_k = \begin{bmatrix} x_{1k} \\ \vdots \\ x_{kk} \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Para obter este efeito, para $j = k+1, \ldots, m$, subtraímos

$$\lambda_{jk} = \frac{x_{jk}}{x_{kk}}$$

vezes a linha k da linha j. A forma da matriz L_k é

$$L_k = \begin{matrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & -\lambda_{k+1,k} & 1 & & \\ & & \vdots & & \ddots & \\ & & -\lambda_{m,k} & & & 1 \end{matrix} .$$

Nos exemplos anteriores observamos dois golpes da sorte:

1. A inversa de L_k é obtida trocando os sinais dos elementos abaixo da diagonal.
2. A matriz $L = L_1^{-1}L_2^{-1}\cdots L_{m-1}^{-1}$ pode ser formada coletando as entradas de λ_{jk} nos locais apropriados.

Podemos reunir esses pedaços de boa fortuna como segue. Seja

$$\lambda_k = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \lambda_{k+1,k} \\ \vdots \\ \lambda_{m,k} \end{bmatrix} .$$

Então $L_k = I - \lambda_k e_k^*$, onde e_k é a coluna k da matriz identidade $m \times m$. Das definições de λ_k e e_k obtemos

$$e_k^* \lambda_k = 0$$

e

$$(I - \lambda_k e_k^*)(I + \lambda_k e_k^*) = I - \lambda_k (e_k^* \lambda_k) e_k^* = I,$$

mostrando que a inversa de L_k é

$$L_k^{-1} = I + \lambda_k e_k^* .$$

Para o segundo golpe de sorte, argumentamos como segue. Considere, por exemplo, o produto $L_k^{-1}L_{k+1}^{-1}$. Como $e_k^* \lambda_{k+1} = 0$, segue

$$L_k^{-1}L_{k+1}^{-1} = (I + \lambda_k e_k^*)(I + \lambda_{k+1} e_{k+1}^*) = I + \lambda_k e_k^* + \lambda_{k+1} e_{k+1}^*$$

que escrita por extenso é

$$L_k^{-1}L_{k+1}^{-1} = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & \lambda_{k+1,k} & 1 & & & \\ & & \lambda_{k+2,k} & \lambda_{k+2,k+1} & 1 & & \\ & & \vdots & \vdots & & \ddots & \\ & & \lambda_{m,k} & \lambda_{m,k+1} & & & 1 \end{bmatrix}$$

Esta matriz é triangular inferior sendo obtida a partir da matriz identidade substituindo os elementos abaixo da diagonal principal das coluna k e $k+1$ pelos elementos de L_k^{-1} e L_{k+1}^{-1} inseridas em seus lugares usuais abaixo da diagonal. Quando tomamos o produto de todas estas matrizes para formar L, obtemos

$$L = L_1^{-1} L_2^{-1} \cdots L_{m-1}^{-1} = \begin{bmatrix} 1 & & & & \\ \lambda_{21} & 1 & & & \\ \lambda_{31} & \lambda_{32} & 1 & & \\ \vdots & \vdots & \ddots & \ddots & \\ \lambda_{m1} & \lambda_{m2} & \cdots & \lambda_{m,m-1} & 1 \end{bmatrix}$$

onde

$$\lambda_{jk} = \frac{x_{jk}}{x_{kk}}$$

são os multiplicadores necessários para anular os elementos abaixo da diagonal da matriz $X = E_{k-1} \cdots E_1 A$.

Tais fatos geram o seguinte algoritmo

======================================

Algoritmo da eliminação gaussiana sem pivotamento

======================================

Entrada: A
Saída: U e L.

======================================

```
U = A e L = I.
for k = 1:m-1
  for j = k+1:m
    L(j,k) = U(j,k)/U(k,k);
    U(j,k:m) = U(j,k:m) - L(j,k) * U(k,k:m);
  end
end
```

======================================

Neste algoritmo podemos usar uma única matriz para armazenar L e U se abrirmos mão de gravar a diagonal de L cujos elementos são unitário. Se a matriz A não for mais necessária, podemos usá-la para gravar L e U.

Solução $Ax = b$ por fatoração LU

Dada a decomposição $A = LU$, o sistema $Ax = b$ é equivalente a $LUx = b$. Defina $y = Ux$, resolva $Ly = b$ e, em seguida, $Ux = y$ para obter a solução do sistema original. A primeira etapa, para a fatoração $A = LU$, exige $\sim \frac{2}{3} m^3$ flops. O segundo e o terceiro, para resolver os sistemas triangulares $Ly = b$ e $Ux = y$, exigem $\sim m^2$ flops. Dessa forma, a resolução pelo método de Gauss, exige $\sim \frac{2}{3}m^3$ flops.

A resolução usando refletores de Householder, que veremos posteriormente, usa $\sim \frac{4}{3}m^3$ flops. Qual seria a vantagem da fatoração QR sobre a fatoração LU?

1.12 Decomposição PLU

Nem sempre uma matriz A possui uma decomposição LU e um exemplo clássico é $\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$. Entretanto, a matriz $B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ obtida de A pela permutação das linhas possui uma decomposição LU

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

Sempre é possível permutar as linhas de uma matriz A de modo que a matriz assim obtida possui uma decomposição LU.

Uma **matriz de permutação** é aquela obtida a partir da matriz identidade mediante uma permutação qualquer de suas linhas. A matriz elementar $E(l_i \longleftrightarrow l_j)$ obtida da identidade pela permutação das linhas i e j pertence a esta classe. Toda matriz de permutação é o produto de matrizes elementares deste tipo. O produto de duas matrizes de permutação é uma matriz de permutação e a inversa de uma matriz de permutação é uma matriz de permutação. Como tivemos a oportunidade de destacar, a inversa de $E(l_i \longleftrightarrow l_j)$ é ela mesma.

Exemplo 1.25 *São matrizes de permutação*

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

A segunda permutação é a transformação elementar $E(l_2 \longleftrightarrow l_4)$.

Seja P uma matriz de permutação obtida da identidade permutando as linha i e j. Seja E a matriz elementar que, a não ser pelo fato de a coluna k possuir elementos não nulos abaixo da diagonal principal, é a identidade. Se $k < i < j$, então

$$\tilde{E} = PEP$$

é uma matriz com os elementos das linhas i e j da coluna k permutados de seus lugares.

Exemplo 1.26 *Sejam*

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad e \quad E = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 0 & 1 & 0 \\ 4 & 0 & 0 & 1 \end{bmatrix}$$

onde P é a matriz de permutação obtida da identidade pela troca das linhas 2 e 4 e E é a matriz elementar com elementos não nulos fora da diagonal principal da primeira coluna. Neste caso, $k = 1$, $i = 2$ e $j = 4$. O produto PAP é igual a

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 4 & 1 & 0 & 0 \\ 3 & 0 & 1 & 0 \\ 2 & 0 & 0 & 1 \end{bmatrix}.$$

Esta matriz pode ser obtida de A permutando os elementos das linhas 2 e 4 da coluna 1. Fato interessantíssimo.

Vamos descrever a decomposição PLU, obtida pelo método da eliminação de Gauss com pivotamento. Seguiremos o tratamento de Trefethen e Bau.

Seja X a matriz obtida no processo de eliminação Gaussiana depois de zerados os elementos abaixo da diagonal principal das $k-1$ primeiras colunas. No passo seguinte, múltiplos da linha k são subtraídas das linhas $k+1, \ldots, m$ da matriz X com a qual se está trabalhando, para introduzir zeros nas entradas k dessas linhas. A entrada x_{kk} da matriz X é chamado de **pivô** da coluna k. Prosseguindo o processo de eliminação, aplica-se a X uma transformação elementar para zerar os elementos da coluna k abaixo da diagonal principal.

$$\begin{bmatrix} x_{kk} & \mathbf{x} & \mathbf{x} & \mathbf{x} \\ x & x & x & x \\ x & x & x & x \\ x & x & x & x \end{bmatrix} \to \begin{bmatrix} x_{kk} & x & x & x \\ \mathbf{0} & \mathbf{x} & \mathbf{x} & \mathbf{x} \\ \mathbf{0} & \mathbf{x} & \mathbf{x} & \mathbf{x} \\ \mathbf{0} & \mathbf{x} & \mathbf{x} & \mathbf{x} \end{bmatrix}$$

Entretanto, x_{kk} pode ser nulo ou muito pequeno quando comparado aos demais elementos daquela coluna abaixo da diagonal. Neste caso, para evitar instabilidade numérica, procura-se naquela coluna, dentre os elementos abaixo da diagonal principal, o de maior módulo. Troca-se esta linha com a linha k e este elemento de maior módulo passa a ser o pivô da linha k. Esta troca de linhas é denominada de **pivotamento parcial**. Há um processo conhecido por **pivotamento** onde se toma por pivô o elemento de maior módulo na submatriz $X_{k:m,k:m}$ e o coloca na posição (k,k) mediante a troca de linhas e colunas. Devido à dificuldade de gerenciar a troca de colunas e ao trabalho computacional para se encontrar o pivô, prefere-se o pivotamento parcial.

$$\begin{bmatrix} x_{kk} & * & * & * \\ * & * & * & * \\ x_{jk} & * & * & * \\ * & * & * & * \end{bmatrix} \overset{P_1}{\to} \begin{bmatrix} x_{jk} & * & * & * \\ * & * & * & * \\ x_{kk} & * & * & * \\ * & * & * & * \end{bmatrix} \overset{L_1}{\to} \begin{bmatrix} x_{jk} & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \end{bmatrix}$$

Este algoritmo pode ser expresso como um produto de matrizes. No pivotamento parcial, em cada etapa, realiza-se uma permutação para posicionar o pivô da coluna no local correto para, em seguida, aplicar uma matriz elementar para zerar os elementos abaixo da diagonal principal da coluna k. Este processo pode ser repetido coluna a coluna, até transformar A numa matriz U triangular superior

$$L_{m-1}P_{m-1}\cdots L_2P_2L_1P_1A = U.$$

Exemplo 1.27 *Considere a matriz (exemplo copiado)*

$$A = \begin{bmatrix} 2 & 1 & 1 & 0 \\ 4 & 3 & 3 & 1 \\ 8 & 7 & 9 & 5 \\ 6 & 7 & 9 & 8 \end{bmatrix}.$$

Para o pivotamento parcial, permutamos a primeira com a terceira coluna calculando $P_1A =$

$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 & 0 \\ 4 & 3 & 3 & 1 \\ 8 & 7 & 9 & 5 \\ 6 & 7 & 9 & 8 \end{bmatrix} = \begin{bmatrix} 8 & 7 & 9 & 5 \\ 4 & 3 & 3 & 1 \\ 2 & 1 & 1 & 0 \\ 6 & 7 & 9 & 8 \end{bmatrix}.$$

Agora efetuamos o primeiro passo de eliminação: $L_1P_1A =$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 & 0 \\ -\frac{1}{4} & 0 & 1 & 0 \\ -\frac{3}{4} & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 8 & 7 & 9 & 5 \\ 4 & 3 & 3 & 1 \\ 2 & 1 & 1 & 0 \\ 6 & 7 & 9 & 8 \end{bmatrix} = \begin{bmatrix} 8 & 7 & 9 & 5 \\ 0 & -\frac{1}{2} & -\frac{3}{2} & -\frac{3}{2} \\ 0 & -\frac{3}{4} & -\frac{5}{4} & -\frac{5}{4} \\ 0 & \frac{7}{4} & \frac{9}{4} & \frac{17}{4} \end{bmatrix}.$$

Em seguida, trocamos a segunda com a quarta linha: $P_2L_1P_1A =$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 8 & 7 & 9 & 5 \\ 0 & -\frac{1}{2} & -\frac{3}{2} & -\frac{3}{2} \\ 0 & -\frac{3}{4} & -\frac{5}{4} & -\frac{5}{4} \\ 0 & \frac{7}{4} & \frac{9}{4} & \frac{17}{4} \end{bmatrix} = \begin{bmatrix} 8 & 7 & 9 & 5 \\ 0 & \frac{7}{4} & \frac{9}{4} & \frac{17}{4} \\ 0 & -\frac{3}{4} & -\frac{5}{4} & -\frac{5}{4} \\ 0 & -\frac{1}{2} & -\frac{3}{2} & -\frac{3}{2} \end{bmatrix}.$$

Efetuamos a segunda eliminação: $L_2P_2L_1P_1A =$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & \frac{3}{7} & 1 & 0 \\ 0 & \frac{2}{7} & 0 & 1 \end{bmatrix} \begin{bmatrix} 8 & 7 & 9 & 5 \\ 0 & \frac{7}{4} & \frac{9}{4} & \frac{17}{4} \\ 0 & -\frac{3}{4} & -\frac{5}{4} & -\frac{5}{4} \\ 0 & -\frac{1}{2} & -\frac{3}{2} & -\frac{3}{2} \end{bmatrix} = \begin{bmatrix} 8 & 7 & 9 & 5 \\ 0 & \frac{7}{4} & \frac{9}{4} & \frac{17}{4} \\ 0 & 0 & -\frac{2}{7} & \frac{4}{7} \\ 0 & 0 & -\frac{6}{7} & -\frac{2}{7} \end{bmatrix}.$$

Agora permutamos a terceira linha com a quarta: $P_3L_2P_2L_1P_1A =$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 8 & 7 & 9 & 5 \\ 0 & \frac{7}{4} & \frac{9}{4} & \frac{17}{4} \\ 0 & 0 & -\frac{2}{7} & \frac{4}{7} \\ 0 & 0 & -\frac{6}{7} & -\frac{2}{7} \end{bmatrix} = \begin{bmatrix} 8 & 7 & 9 & 5 \\ 0 & \frac{7}{4} & \frac{9}{4} & \frac{17}{4} \\ 0 & 0 & -\frac{6}{7} & -\frac{2}{7} \\ 0 & 0 & -\frac{2}{7} & \frac{4}{7} \end{bmatrix}.$$

Finalmente, efetuamos a última eliminação: $L_3P_3L_2P_2L_1P_1A =$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{3} & 1 \end{bmatrix} \begin{bmatrix} 8 & 7 & 9 & 5 \\ 0 & \frac{7}{4} & \frac{9}{4} & \frac{17}{4} \\ 0 & 0 & -\frac{6}{7} & -\frac{2}{7} \\ 0 & 0 & -\frac{2}{7} & \frac{4}{7} \end{bmatrix} = \begin{bmatrix} 8 & 7 & 9 & 5 \\ 0 & \frac{7}{4} & \frac{9}{4} & \frac{17}{4} \\ 0 & 0 & -\frac{6}{7} & -\frac{2}{7} \\ 0 & 0 & 0 & \frac{2}{3} \end{bmatrix}.$$

Um terceiro golpe de sorte na fatoração PLU

Todos os elementos de L abaixo da diagonal principal são menores ou iguais a 1 pois o pivô de cada linha é escolhido de modo a tornar $|x_{kk}| = \{|x_{jk}| : k \leq j \leq m\}$

Analisemos a decomposição de uma matriz A de tamanho 4×4 que toma a forma

$$L_3 P_3 L_2 P_2 L_1 P_1 A = U$$

As matrizes P_1, P_2 e P_3 são suas próprias inversas. Assim podemos escrever

$$L_3 P_3 L_2 P_2 L_1 P_1 = L_3 P_3 L_2 (P_3 P_3) P_2 L_1 (P_2 P_3 P_3 P_2) P_1$$

onde acrescentamos algumas matrizes ao produto e foram colocadas entre parêntesis. Note que elas são iguais à matriz identidade. Podemos associar este produto

$$L_3 P_3 L_2 P_2 L_1 P_1 = L_3 (P_3 L_2 P_3)(P_3 P_2 L_1 P_2 P_3)(P_3 P_2 P_1) = (\tilde{L}_1 \tilde{L}_2 \tilde{L}_3)(P_3 P_2 P_1)$$

onde

$$\tilde{L}_3 = L_3, \quad \tilde{L} = P_3 L_2 P_3, \quad \tilde{L} = P_3 P_2 L_1 P_2 P_3$$

são matrizes elementares obtidas de L_3, L_2, L_1 permutando elementos abaixo da diagonal principal.

Em geral, para uma matriz $m \times m$, a fatoração fornecida pela eliminação Gaussiana com pivotamento parcial pode ser escrita na forma

$$(\tilde{L}_{m-1} \cdots \tilde{L}_2 \tilde{L}_1)(P_{m-1} \cdots P_2 P_1) A = U,$$

onde

$$\tilde{L}_k = P_{m-1} \cdots P_{k+1} L_k P_{k+1}^{-1} \cdots P_{m-1}^{-1}.$$

O produto das matrizes $\tilde{L}_k$ é triangular inferior com elementos unitários na diagonal principal e facilmente invertível. Basta trocar o sinal das entradas abaixo da diagonal, como na eliminação Gaussiana sem pivotamento. Escrevendo

$$L = (\tilde{L}_{m-1} \cdots \tilde{L}_2 \tilde{L}_1)^{-1} \qquad \text{e} \qquad P = P_{m-1} \cdots P_2 P_1,$$

temos

$$PA = LU.$$

Qualquer matriz quadrada A, singular ou não, possui uma fatoração deste tipo, onde P é uma matriz de permutação, L é uma matriz triangular inferior com elementos unitários na diagonal principal e U é triangular superior. Esta fatoração é conhecida por fatoração PLU de A.

Para obter a fatoração PLU de A, multiplique a matriz A por uma matriz de permutação P e calcule a decomposição LU de A. Na prática, não é assim que se procede pois não se conhece P a priori.

Vamos descrever um procedimento que justifica o algoritmo que vamos descrever abaixo. Seja A uma matriz $m \times m$ e $X = E_k P_k \cdots \; E_1 P_1 \; A = \tilde{E}_k \cdots \tilde{E}_1 \; P_k \cdots P_1 \; A$, onde P_i são matrizes de permutação que posicionam o pivô no lugar correto e $\tilde{E}_i$ são matrizes elementares que zeram as entradas abaixo da diagonal da coluna i de $P_k \cdots P_1$ A. Vamos escrever $X = \tilde{E}\ \tilde{A}$ onde $\tilde{E} = \tilde{E}_k \cdots \tilde{E}_1$ e $\tilde{A} = P_k \cdots P_1\ A$. Se $k < m - 1$, o processo não terminou e X, em geral, não é triangular superior. A próxima etapa consiste em aplicar uma permutação P que trocará uma linha i de X com sua linha $k + 1$ para posicionar o pivô da coluna no local correto. Neste caso, $i > k + 1$. A inversa de P é P e assim $PP = I$. Podemos usar este fato para escrever

$$PX = P\tilde{E}\tilde{A} = P\tilde{E}(PP)\tilde{A} = (P\tilde{E}P)(P\tilde{A}).$$

Lembramos que $P\tilde{E}P$ é triangular inferior e é a matriz $\tilde{E}$ onde se permutou a parte não nula das linhas i e $k + 1$, situadas abaixo da diagonal ficam permutadas. Desta forma, sempre que se aplica uma permutação à matriz $\tilde{A}$ se deve efetuar uma permutação correspondente na matriz $\tilde{E}$.

Este comentáriio justifica o algoritmo da eliminação Gaussiana com pivotamento parcial descrito abaixo.

Algoritmo da eliminação Gaussiana com pivotamento parcial

```
=============================
U = A, L = I, P = I
for k = 1:m-1
  Selecione na coluna k a linha i na qual |u(i,k)| eh maximo
  Permute as linhas U(k,k:m) e U(i,k:m)
  Permute as linhas L(k,1:k-1) e L(i,1:k-1)
  Permute as linhas P(k,:) e P(i,:)
  for j = k+1:m
    L(j,k) = U(j,k) / U(k,k)
    U(j,k:m) = U(j,k:m) - L(j,k)*U(k,k:m)
  end
end
=============================
```

1.13 Decomposição de Cholesky

Se a matriz A for simétrica e inversível, uma permutação PA dessa matriz tem uma decomposição PLU. Vamos, num primeiro momento, nos esquecer da permutação P e escrever esta decomposição na forma $A = LU$ de modo que

$$LU = A = A^T = U^T L^T.$$

Como L e U são inversíveis,

$$U \left(L^T\right)^{-1} = L^{-1}U^T = D$$

é diagonal pois $U\left(L^T\right)^{-1}$ é triangular superior e $L^{-1}U^T$ é triangular inferior. Assim, $U = DL^T$ e obtemos a decomposição

$$A = LDL^T$$

onde L é triangular inferior cujos elementos diagonais são iguais a 1 e $D = L^{-1}U^T$ é diagonal.

Como os elementos da diagonal principal de L são iguais a 1, $D = diag(U)$ onde $diag(U)$ é uma matriz diagonal, cujos elementos da diagonal principal são iguais aos elementos da diagonal de U.

Exemplo 1.28 *Considere a decomposição $A = LU$ abaixo*

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -1/2 & 1 & 0 \\ 0 & -2/3 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 & 0 \\ 0 & 3/2 & -1 \\ 0 & 0 & 4/3 \end{bmatrix}$$

Sendo $D = L^{-1}U^T = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3/2 & 0 \\ 0 & 0 & 4/3 \end{bmatrix}$ obtemos a decomposição LDL^T

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -1/2 & 1 & 0 \\ 0 & -2/3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3/2 & 0 \\ 0 & 0 & 4/3 \end{bmatrix} \begin{bmatrix} 1 & -1/2 & 0 \\ 0 & 1 & -2/3 \\ 0 & 0 & 1 \end{bmatrix}.$$

Definição 1.29 *Uma matriz simétrica A é **positiva definida** se os elementos diagonais de D na decomposição $A = LDL^T$ forem todos maiores do que zero. Neste caso, podemos calcular a matriz $\sqrt{D}$ e definir $M = L\sqrt{D}$, para assim obter a decomposição $A = MM^T$ denominada de **decomposição de Cholesky** da matriz A.*

No exemplo acima,

$$M = \begin{bmatrix} 1 & 0 & 0 \\ -1/2 & 1 & 0 \\ 0 & -2/3 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & \sqrt{3/2} & 0 \\ 0 & 0 & \sqrt{4/3} \end{bmatrix} = \begin{bmatrix} \sqrt{2} & 0 & 0 \\ -\sqrt{1/2} & \sqrt{3/2} & 0 \\ 0 & -\sqrt{2/3} & \sqrt{4/3} \end{bmatrix}.$$

A decomposição de Cholesky de A é

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} = \begin{bmatrix} \sqrt{2} & 0 & 0 \\ -\frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{6} & 0 \\ 0 & -\frac{1}{3}\sqrt{6} & \frac{2}{3}\sqrt{3} \end{bmatrix} \begin{bmatrix} \sqrt{2} & -\frac{1}{2}\sqrt{2} & 0 \\ 0 & \frac{1}{2}\sqrt{6} & -\frac{1}{3}\sqrt{6} \\ 0 & 0 & \frac{2}{3}\sqrt{3} \end{bmatrix}.$$

Capítulo 2

Espaço vetorial

2.1 Conceito de espaço vetorial

Seja K um corpo e V um conjunto não vazio, onde definimos duas operações, sendo uma a **adição de vetores** e a outra a **multiplicação** de um elemento do corpo K por um elemento de V. Sejam v e w dois elementos de V e k um elemento do corpo K. Denotaremos a adição de v e w por $v + w$ e a multiplicação de k e v por kv. O conjunto V, com essas operações é denominado de **espaço vetorial sobre o corpo** K se, para todo u, v, w de V e todo α, β de K, se verificarem as propriedades

1. Comutativa: $v + w = w + v$.
2. Associativa: $(u + v) + w = u + (v + w)$.
3. Elemento neutro: Existe um elemento de V denotado por 0 tal que $0+v = v+0 = v$.
4. Elemento oposto: Dado v em V existe um elemento denotado por $-v$ e tal que $v + (-v) = (-v) + v = 0$.
5. Associatividade: $(\alpha\beta)v = \alpha(\beta v)$.
6. Distributividade: $(\alpha + \beta)v = \alpha v + \beta v$.
7. Distributividade: $\alpha(v + w) = \alpha v + \alpha w$.
8. Elemento unitário: A multiplicação do elemento unitário 1 de K pelo elemento v de V é igual a v, isto é, $1v = v$.

Os elementos de V são chamados **vetores** e os elementos de K de **escalares**. O elemento $v + w$ é o **vetor soma** de v com w e o elemento αv é o **produto** de α por v ou ainda que αv é um múltiplo de v. O vetor $-v$ é denominado **oposto** de v e 0 é o **vetor nulo** ou **vetor zero**. Definimos a **diferença** $v - w$ (leia-se v menos w) entre os vetores v e w por $v + (-w)$.

Em nosso curso, o corpo K será o corpo $\mathbf{R}$ dos números reais ou o corpo $\mathbf{C}$ dos números complexos. Quando V for um espaço vetorial sobre o corpo dos números reais, diremos que V é um **espaço vetorial real**. Quando V for um espaço vetorial sobre o corpo dos números complexos, diremos que V é um **espaço vetorial complexo**.

Quando se diz que V é um espaço vetorial sobre o corpo K entenda-se que está implícito a existência das operações de adição de vetores e multiplicação de um escalar por um vetor. Quando o contexto permitir, omite-se a referência ao corpo K e se diz apenas que V é um espaço vetorial. O espaço vetorial $\{0\}$ que contém apenas o vetor nulo é denominado de **espaço vetorial trivial**.

Exemplo 2.1 *Seja $\mathbf{R}^n$ o conjunto de todas as ênuplas ordenadas $(x_1,\ x_2,\ \ldots,\ x_n)$ de números reais. Duas ênuplas ordenadas $(x_1,\ x_2,\ \ldots,\ x_n)$ e $(y_1,\ y_2,\ \ldots,\ y_n)$ são iguais se $x_1 = y_1,\ x_2 = y_2,\ \ldots,\ x_n = y_n$. Define-se a operação de adição em $\mathbf{R}^n$ por*

$$(x_1,\ x_2,\ \ldots,\ x_n) + (y_1,\ y_2,\ \ldots,\ y_n) = (x_1 + y_1,\ x_2 + y_2,\ \ldots,\ x_n + y_n)$$

e a multiplicação de um número real por uma ênupla ordenada é definida por

$$\alpha(x_1,\ x_2,\ \ldots,\ x_n) = (\alpha x_1,\ \alpha x_2,\ \ldots,\ \alpha x_n).$$

O $\mathbf{R}^n$ com as operações de adição de duas ênuplas ordenadas e multiplicação de um escalar por uma ênupla é um espaço vetorial sobre os reais.

Exemplo 2.2 *O conjunto $\mathbf{R}^{m\times n}$ das matrizes $m \times n$ com elementos reais munido com as operações de adição de matrizes e multiplicação de um número complexo por uma matriz é um espaço vetorial sobre o corpo dos números reais. O zero deste espaço vetorial é a matriz nula e o elemento oposto (inverso aditivo) de $A = [a_{ij}]$ é $-A = [-a_{ij}]$.*

Exemplo 2.3 *O conjunto das matrizes m por n com elementos complexos, que denotaremos por $\mathbf{C}^{m\times n}$, munido com as operações de adição de matrizes e multiplicação de um número complexo por uma matriz é um espaço vetorial sobre o corpo dos números complexos. O zero deste espaço vetorial é a matriz nula e o elemento oposto (inverso aditivo) de $A = [a_{ij}]$ é $-A = [-a_{ij}]$.*

Exemplo 2.4 *O conjunto de todos os polinômios de grau menor ou igual a n, com coeficientes reais, munido com as operações de adição de polinômios e multiplicação de um número real por um polinômio, é um espaço vetorial sobre o corpo dos reais. O conjunto dos polinômios de grau menor ou igual a n com coeficientes complexos com as operações acima é um espaço vetorial sobre o corpo dos números complexos.*

Exemplo 2.5 *O conjunto de todos os polinômios com coeficientes reais, munido com as operações de adição de polinômios e multiplicação de um número real por um polinômio, é um espaço vetorial sobre o corpo dos reais. O conjunto de todos os polinômios com coeficientes complexos com as operações acima é um espaço vetorial sobre o corpo dos números complexos.*

Exemplo 2.6 *O conjunto $C[a,\ b] = \{f : [a,\ b] \to \mathbf{R} : f$ é contínua$\}$ com as operações de adição de funções e multiplicação de um número real por uma função é um espaço vetorial sobre $\mathbf{R}$.*

2.2 Dependência linear

Todo elemento (x, y) do $\mathbf{R}^2$ pode ser decomposto na seguinte soma

$$(x, y) = x(1, 0) + y(0, 1).$$

Esta maneira de decompor um vetor é muito utilizada em Álgebra Linear.

Sejam $v_1, \ldots, v_n$ vetores do espaço vetorial V e escalares $\alpha_1, \ldots, \alpha_n$. O vetor $\alpha_1 v_1 + \cdots + \alpha_n v_n$ é uma **combinação linear** dos vetores $v_1, \ldots, v_n$.

Exemplo 2.7 *O vetor* $(2, 3)$ *do* $\mathbf{R}^2$ *é uma combinação linear dos vetores* $(1, 0)$ *e* $(0, 1)$ *pois* $(2, 3) = 2(1, 0) + 3(0, 1)$.

Seja $\{v_1, \ldots, v_n\}$ um subconjunto finito de V. Este conjunto é **linearmente dependente** se existirem escalares $\alpha_1, \ldots, \alpha_n$, nem todos nulos tais que

$$\alpha_1 v_1 + \cdots + \alpha_n v_n = 0.$$

Também se diz que os vetores $v_1, \ldots, v_n$ são linearmente dependentes. Notem que a igualdade acima se verifica para $\alpha_1 = \cdots = \alpha_n = 0$. Se a ênupla $(\alpha_1, \ldots, \alpha_n) = (0, \ldots, 0)$ for a única para a qual

$$\alpha_1 v_1 + \cdots + \alpha_n v_n = 0,$$

diremos que o conjunto $\{v_1, \ldots, v_n\}$ é **linearmente independente** ou que os vetores $v_1, \ldots, v_n$ são linearmente independentes.

Exemplo 2.8 *O conjunto* $S = \{\ (5, 7),\ (1, 0),\ (0, 1)\ \}$ *de vetores do* $\mathbf{R}^2$ *é linearmente dependente pois*

$$1(5, 7) - 5(1, 0) - 7(0, 1) = (0, 0).$$

O conjunto $\{\ (1, 2, 3), (0, 1, 1), (0, 0, 2)\ \}$ *de vetores do* $\mathbf{R}^3$ *é linearmente independente. De fato, se* α_1, α_2 *e* α_3 *forem escalares tais que*

$$\alpha_1(1, 2, 3) + \alpha_2(0, 1, 1) + \alpha_3(0, 0, 2) = (0, 0, 0)$$

então

$$\begin{aligned} \alpha_1 + 0\alpha_2 + 0\alpha_3 &= 0 \\ 2\alpha_1 + \alpha_2 + 0\alpha_3 &= 0 \\ 3\alpha_1 + \alpha_2 + 2\alpha_3 &= 0 \end{aligned}$$

cuja única solução é $\alpha_1 = \alpha_2 = \alpha_3 = 0$.

Todo conjunto $\{0, v_1, \ldots v_p\}$ que contém o vetor nulo é linearmente dependente pois

$$1 \cdot 0 + 0v_1 + \cdots + 0v_p = 0.$$

Observe que, a dependência linear do conjunto $S = \{\ (5,7),\ (1,0),\ (0,1)\ \}$ de vetores do $\mathbf{R}^2$ que se expressa por

$$1(5,7) - 5(1,0) - 7(0,1) = (0,0).$$

implica na possibilidade de escrever (5, 7) como uma combinação linear dos vetores (1, 0) e (0, 1)

$$(5,7) = 5(1,0) + 7(0,1).$$

Esta igualdade também implica na dependência linear de $S = \{\ (5,7),\ (1,0),\ (0,1)\ \}$. Tal fato é enunciado de modo geral no próximo teorema.

Proposição 2.9 *Um conjunto $\{v_1, \ldots, v_n\}$ de vetores de um espaço vetorial V é linearmente dependente se e só se um dos seus elementos for combinação linear dos demais.*

Prova. Se $\{v_1, \ldots, v_n\}$ for linearmente dependente, existem escalares $\alpha_1, \ldots, \alpha_n$, nem todos nulos, tais que $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$. Supondo $\alpha_1 \neq 0$ (se $\alpha_1 = 0$, basta permutar os vetores do conjunto para trazer o coeficiente não nulo para a primeira posição) podemos escrever v_1 como combinação linear de $v_2, \ldots, v_n$

$$v_1 = \left(-\alpha_1^{-1}\alpha_2\right) v_2 - \cdots \left(-\alpha_1^{-1}\alpha_n\right) v_n.$$

Se v_1 for uma combinação linear de $v_2, \ldots, v_n$, então existem escalares $\beta_2, \ldots, \beta_n$ tais que

$$v_1 = \beta_2 v_2 + \cdots + \beta_n v_n$$

e

$$v_1 + (-\beta_2) v_2 + \cdots + (-\beta_n) v_n = 0,$$

mostrando que $\{v_1, \ldots, v_n\}$ é linearmente dependente. □

Todo conjunto que contém um subconjunto linearmente dependente é linearmente dependente. Todo subconjunto de um conjunto de vetores linearmente independente é linearmente independente.

Proposição 2.10 *Seja S um conjunto finito de vetores.*

1. *Se S for linearmente dependente, qualquer conjunto finito de vetores que o contém também será linearmente dependente.*

2. *Se S for linearmente independente, qualquer subconjunto de S será linearmente independente.*

Prova. Seja $S = \{v_1, \ldots, v_n\}$.

1. Se S for linearmente dependente, existem escalares $\alpha_1, \ldots, \alpha_n$ nem todos nulos tais que $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$. Seja S' um conjunto finito que contém S. Se $w_1, \ldots, w_m$ forem os elementos de S' que não pertencem a S, então
$$\alpha_1 v_1 + \cdots + \alpha_n v_n + 0 w_1 + \cdots + 0 w_m = 0$$
provando que S' é linearmente dependente.

2. Se S for linearmente independente, seja S' um subconjunto de S. Se S' fosse linearmente dependente, S também o seria pela primeira parte. Logo S' é linearmente independente.

$\square$

2.3 Base e dimensão

Seja $\mathcal{B} = \{v_1, \ldots, v_n\}$ um conjunto finito de vetores em V. Se todo elemento de V for uma combinação linear dos elementos de $\mathcal{B}$, diremos que $\mathcal{B}$ gera V.

Exemplo 2.11 *O conjunto $\mathcal{B} = \{(1,2),\ (1,0),\ (0,1)\}$ gera o $\mathbf{R}^2$. Qualquer par ordenado $(x,\ y)$ pode ser decomposto nas combinações lineares*
$$(x,y) = 0(1,2) + x(1,0) + y(0,1)$$
ou
$$(x,y) = x(1,2) + 0(1,0) + (y-2x)(0,1).$$
Neste exemplo, o modo de escrever (x,y) como combinação linear dos elementos de $\mathcal{B}$ não é única.

Exemplo 2.12 *O conjunto $\mathcal{B} = \{(2,1),\ (1,\ 0)\ \}$ gera o $\mathbf{R}^2$ pois podemos escrever um par ordenado $(x,\ y)$ qualquer como combinação linear desses dois vetores*
$$(x,y) = x(2,1) + (y-x)(1,0).$$
Neste exemplo, o modo de escrever (x,y) como combinação linear dos elementos de $\mathcal{B}$ é única.

Que diferença existe entre os conjuntos geradores dos exemplos acima? O primeiro é linearmente dependente e o segundo é linearmente dependente.

Definição 2.13 *Um conjunto finito de vetores linearmente independente e que gera V é uma **base** de V.*

Uma base $\mathcal{B} = \{v_1, \ldots, v_n\}$ gera V. Assim, para cada vetor v em V existem escalares $\alpha_1, \ldots, \alpha_n$ tais que

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

Os vetores $\alpha_1 v_1, \ldots, \alpha_n v_n$ são denominados de **componentes** do vetor v na base $\mathcal{B}$, os escalares $\alpha_1, \ldots, \alpha_n$ são as **coordenadas** de v na base $\mathcal{B}$ e a matriz coluna

$$[v]_{\mathcal{B}} = [\alpha_1 \;\ldots\; \alpha_n]^T$$

é a **matriz das coordenadas** de v na base $\mathcal{B}$.

Uma **base ordenada** $\mathcal{B} = \{v_1, v_2, \ldots, v_n\}$ é aquela em que se estabelece que v_1 é o seu primeiro elemento, que v_2 é o seu segundo elemento, e assim por diante. A ordem em que seus elementos são escritos é relevante.

Proposição 2.14 *A matriz das coordenadas de um vetor numa base ordenada é única.*

Prova. Seja $\mathcal{B} = \{v_1, \ldots, v_n\}$ uma base ordenada de um espaço vetorial V. Se $v = x_1 v_1 + \cdots + x_n v_n$ e $v = y_1 v_1 + \cdots + y_n v_n$ forem duas decomposições de v nos elementos da base $\mathcal{B}$, então

$$0 = v - v = (x_1 - y_1)v_1 + \cdots + (x_n - y_n)v_n$$

e, da independência linear dos vetores da base, $x_i = y_i$ para $i = 1, \ldots, n$. $\square$

De ora em diante, uma base ordenada será chamada simplesmente de base. O contexto indicará a necessidade de ser a base ordenada ou não.

Exemplo 2.15 *Considere as ênuplas $e_1 = (1, 0, \ldots, 0)$, $e_2 = (0, 1, \ldots, 0)$, $e_n = (0, 0, \ldots, 1)$, onde e_k é a linha k da matriz identidade $n \times n$. O conjunto de vetores $\{e_1, e_2, \ldots, e_n\}$ é uma base tanto do $\mathbf{R}^n$ quanto do $\mathbf{C}^n$ e é chamada de **base canônica**. Se $x = (x_1, \ldots, x_n)$, então $x = x_1 e_1 + \cdots + x_n e_n$. Isto significa que as coordenadas de x na base canônica são exatamente os elementos da ênupla x.*

Exemplo 2.16 *O conjunto $\{1, x, x^2\}$ é uma base do espaço vetorial dos polinômios de grau menor ou igual a dois com coeficientes complexos.*

Nem todo espaço vetorial possui uma base tal como se definiu acima. O espaço vetorial de todos os polinômios com coeficientes complexos não possui base no sentido definido neste texto. Não existe conjunto finito de polinômios que gera todos os demais. Todo conjunto finito de polinômios tem um polinômio de grau máximo, que não seria capaz de gerar os polinômios de grau superior ao polinômio de grau máximo do conjunto.

Todas as bases de um espaço vetorial possuem o mesmo número de elementos, como provaremos em seguida. Precederemos o teorema principal por três lemas.

Lema 2.17 *Seja $\{v_1, \ldots, v_n\}$ uma base de V e $w = \alpha_1 v_1 + \cdots + \alpha_n v_n$. Se $\alpha_i \neq 0$, então*

$$\{v_1, \ldots, v_{i-1}, w, v_{i+1}, , \ldots, v_n\}$$

também é base.

Prova. Para simplificar, provaremos o teorema supondo $\alpha_1 \neq 0$. Se $\alpha_1 = 0$, podemos reordenar os elementos da base para trazer para a primeira posição uma componente de w diferente de zero. Sendo $\alpha_1 \neq 0$, podemos explicitar v_1 na igualdade $w = \alpha_1 v_1 + \cdots + \alpha_n v_n$ para obter

$$v_1 = \frac{1}{\alpha_1} w - \frac{\alpha_2}{\alpha_1} v_2 - \cdots - \frac{\alpha_n}{\alpha_1} v_n = \beta_1 w + \beta_2 v_2 + \cdots + \beta_n v_n.$$

Vamos provar que $\{w,\ v_2,\ \ldots, v_n\}$ gera V. Sendo v um vetor qualquer de V, existem escalares $x_1,\ x_2,\ \ldots,\ x_n$ tais que

$$\begin{aligned} v &= x_1 v_1 + x_2 v_2 + \cdots + x_n v_n \\ &= x_1(\beta_1 w + \beta_2 v_2 + \cdots + \beta_n v_n) + x_2 v_2 + \cdots + x_n v_n \\ &= (x_1 \beta_1) w + (x_1 \beta_2 + x_2) v_2 + \cdots + (x_1 \beta_n + x_n) v_n, \end{aligned}$$

provando que o conjunto $\{w,\ v_2,\ \ldots, v_n\}$ gera V.

Vamos provar que $\{w,\ v_2,\ \ldots, v_n\}$ é linearmente independente. Sejam $k_1,\ k_2,\ \ldots,\ k_n$ escalares tais que $k_1 w + k_2 v_2 + \cdots + k_n v_n = 0$. Se $k_1 \neq 0$, então

$$k_1(\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n) + k_2 v_2 + \cdots + k_n v_n = 0$$

ou

$$k_1 \alpha_1 v_1 + (k_1 \alpha_2 + k_2) v_2 + \cdots + (k_1 \alpha_n + k_n) v_n = 0$$

com $k_1 \alpha_1 \neq 0$, o que contraria o fato de $\{v_1,\ v_2,\ \ldots,\ v_n\}$ ser base de V. Logo, $k_1 = 0$ e a combinação linear $k_1 w + k_2 v_2 + \cdots + k_n v_n = 0$ se reduz a $k_2 v_2 + \cdots + k_n v_n = 0$. Da independência linear do conjunto $\{v_2,\ \ldots,\ v_n\}$, obtemos $k_2 = \cdots = k_n = 0$, provando a independência linear de $\{w,\ v_2,\ \ldots,\ v_n\}$ que, portanto, é base de V. $\Box$

Lema 2.18 *Seja $\{v_1,\ \ldots,\ v_n\}$ uma base com n elementos do espaço vetorial V. Todo conjunto linearmente independente com n elementos é base de V.*

Prova. Seja $\{w_1,\ \ldots,\ w_n\}$ um conjunto linearmente independente com n vetores de V. Pode-se decompor w_1 na base $\{v_1,\ \ldots,\ v_n\}$ e escrever

$$w_1 = c_{11} v_1 + \cdots + c_{n1} v_1.$$

Como $w_1 \neq 0$, pelo menos um dos coeficientes desta combinação linear é diferente de zero. Podemos supor que $c_{11} \neq 0$ (se o c_{11} fosse nulo, bastaria reordenar a base $\{v_1,\ v_2,\ \ldots,\ v_n\}$ de modo que, nesta nova ordem, $c_{11} \neq 0$).

Pelo lema anterior, $\{w_1,\ v_2,\ \ldots,\ v_n\}$ é base e podemos escrever

$$w_2 = c_{12} w_1 + c_{22} v_2 + \cdots + c_{n2} v_n.$$

Os coeficientes $c_{22}, \ldots, c_{n2}$ não podem ser todos nulos. De fato, se todos eles fossem nulos, então $w_2 = c_{12} w_1$, o que contraria a hipótese de o conjunto $\{w_1,\ \ldots,\ w_n\}$ ser linearmente independente. Assim, pelo menos um dos coeficientes $c_{22}, \ldots, c_{n2}$ não é nulo. Como antes, podemos supor, sem perda de generalidade, que $c_{22} \neq 0$.

Pelo lema anterior, $\{w_1, w_2, v_3, \ldots, v_n\}$ é base de V.

Prosseguindo com este raciocínio, substituímos todos os elementos da base $\{v_1, \ldots, v_n\}$ por $w_1, w_2, \ldots, w_n$, provando que $\{w_1, w_2, \ldots, w_n\}$ é base. $\square$

Lema 2.19 *Se um espaço vetorial V possuir uma base com n elementos, então todo conjunto de vetores em V com mais de n elementos é linearmente dependente.*

Prova. De fato, se houvesse um conjunto linearmente independente com mais do que n elementos, qualquer subconjunto dele com n elementos seria base e os vetores restantes seriam combinações lineares desses n selecionados, contrariando a hipótese de independência linear do conjunto. Logo, não existe conjunto de vetores linearmente independente com mais do que n elementos. $\square$

Estes lemas nos permitem enunciar o

Teorema 2.20 *Se um espaço vetorial V possuir uma base com n elementos, todas as outras bases deste espaço vetorial têm o mesmo número de elementos.*

Prova. De fato, como todo conjunto com mais do que n elementos é linearmente dependente, não há base com mais do que n elementos.

Seja $\mathcal{B}_1$ a base com n elementos. Se existisse alguma base $\mathcal{B}_2$ com k elementos e $k < n$, pelo lema anterior, a base $\mathcal{B}_1$ seria linearmente dependente, possibilidade que se exclui pela definição de base. Logo não existe base com menos do que n elementos. $\square$

Este teorema garante que todas as bases de um espaço vetorial possui o mesmo número de elementos o que justifica a definição que segue.

Definição 2.21 *Se um espaço vetorial possui uma base, diremos que ele possui* ***dimensão finita*** *e que o número de elementos das bases é a sua* ***dimensão****. Por definição, a dimensão do espaço vetorial trivial, aquele que contém apenas o vetor nulo, é zero.*

2.4 Matriz de mudança de base

Seja V um espaço vetorial complexo de dimensão finita $n > 0$. Sejam $\mathcal{B}_1 = \{u_1, \ldots, u_n\}$ e $\mathcal{B}_2 = \{v_1, \ldots, v_n\}$ duas bases de V. Podemos decompor cada elemento de $\mathcal{B}_2$ numa combinação linear dos elementos de $\mathcal{B}_1$

$$\begin{aligned}
v_1 &= p_{11}u_1 + p_{21}u_2 + \cdots + p_{n1}u_n \\
v_2 &= p_{12}u_1 + p_{22}u_2 + \cdots + p_{n2}u_n \\
&\cdots \\
v_n &= p_{1n}u_1 + p_{2n}u_2 + \cdots + p_{nn}u_n
\end{aligned}$$

A matriz
$$M_{12} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$
é chamada de **matriz de mudança de base**, mais especificamente, matriz de mudança da base $\mathcal{B}_1$ para a base $\mathcal{B}_2$. Observe que as coordenadas do desenvolvimento de v_1 na base $\mathcal{B}_1$ formam a primeira coluna, as coordenadas do desenvolvimento de v_2 na base $\mathcal{B}_1$ formam a primeira coluna,

Sendo $\mathcal{B}_3 = \{w_1, \ldots, w_n\}$ uma terceira base de V, podemos escrever os vetores de $\mathcal{B}_3$ como combinações lineares dos elementos da base $\mathcal{B}_2$. Usando o símbolo de somatório,
$$w_j = \sum_{i=1}^{n} q_{ij} v_i$$
e agora, $M_{23} = [q_{ij}]$ é a matriz de mudança da base $\mathcal{B}_2$ para a base $\mathcal{B}_3$.

Das duas decomposições acima segue
$$\begin{aligned} w_j &= \sum_k q_{kj} v_k = \sum_k q_{kj} \sum_i p_{ik} u_i \\ &= \sum_i \left(\sum_k p_{ik} q_{kj} \right) u_i = \sum_i r_{ij} u_i \end{aligned}$$
onde $M_{13} = [r_{ij}] = [\sum_k p_{ik} q_{kj}]$ é a matriz de mudança da base $\mathcal{B}_1$ para a base $\mathcal{B}_3$. Como
$$M_{13} = [r_{ij}] = \left[\sum_k p_{ik} q_{kj} \right] = [p_{ik}][q_{kj}] = M_{12} M_{23},$$
provamos a identidade
$$M_{13} = M_{12} M_{23}.$$
Quando $\mathcal{B}_3 = \mathcal{B}_1$, a matriz M_{13} é a identidade I e $M_{23} = M_{21}$. Da igualdade acima segue
$$M_{12} M_{21} = I,$$
mostrando que as matrizes de mudança de base são inversíveis e que a inversa de M_{12} é M_{21}.

Sejam i e j inteiros do conjunto $\{1, 2, \ldots, n\}$. O **delta de Kronecker** δ_{ij}, é um conjunto de n^2 números definidos do seguinte modo: $\delta_{ij} = 1$ quando $i = j$ e $\delta_{ij} = 0$ quando $i \neq j$.

Observe que o elemento da linha i coluna j da matriz identidade I de ordem $n \times n$ é exatamente δ_{ij} e podemos usar o delta de Kronecker para escrever $I = [\delta_{ij}]$.

As igualdades matriciais
$$M_{12} M_{21} = I \quad \text{e} \quad M_{21} M_{12} = I$$

quando escritas componente a componente, fornece

$$\sum_k p_{ik} q_{kj} = \delta_{ij} \qquad \text{e} \qquad \sum_k q_{ik} p_{kj} = \delta_{ij}$$

para i e j percorrendo os valores $1, \ldots, n$.

Mudança de coordenadas

Teorema 2.22 *Sejam $\mathcal{B}_1$ e $\mathcal{B}_2$ duas bases do espaço vetorial V. Sejam*
$[u]_1$ a matriz das coordenadas de u na base $\mathcal{B}_1$,
$[u]_2$ a matriz das coordenadas de u na base $\mathcal{B}_2$ e
M_{12} a matriz de mudança da base $\mathcal{B}_1$ para a base $\mathcal{B}_2$.
Então

$$[u]_1 = M_{12}[u]_2$$

Prova. Sejam $\mathcal{B}_1 = \{v_1, \ldots, v_n\}$ e $\mathcal{B}_2 = \{w_1, \ldots, w_n\}$ as bases em questão. Se $[u]_1 = [x_1, \ldots, x_n]^T$ for a matriz das coordenadas de u na base $\mathcal{B}_1$, se $[u]_2 = [y_1, \ldots, y_n]^T$ for a matriz das coordenadas de u na base $\mathcal{B}_2$ e se $M_{12} = [p_{ij}]$ for a matriz de mudança da base $\mathcal{B}_1$ para a base $\mathcal{B}_2$, segue

$$u = \sum_i x_i v_i = \sum_j y_j w_j$$

e

$$w_j = \sum_{i=1}^n p_{ij} v_i.$$

Portanto,

$$\begin{aligned} u &= \sum_j y_j w_j = \sum_j y_j \sum_i p_{ij} v_i \\ &= \sum_i \left(\sum_j p_{ij} y_j \right) v_i. \end{aligned}$$

Como $u = \sum_i x_i v_i$, segue da unicidade da decomposição de um vetor nos elementos da base que

$$x_i = \sum_j p_{ij} y_j$$

que corresponde à igualdade matricial

$$[u]_1 = M_{12}[u]_2.$$

□

2.5 Subespaço vetorial

Seja V um espaço vetorial e W um subconjunto não vazio de V. Diremos que W é um **subespaço vetorial** de V se, para todo v e w em W e todo escalar λ, os vetores λw e $v + w$ pertencerem a W. Em outras palavras, o subespaço vetorial é aquele subconjunto fechado em relação à adição e à multiplicação por um escalar.

As operações de adição de vetores e multiplicação por uma escalar definidas em V, também se aplicam aos vetores de W, que está contido em V. Certamente, essas operações em W gozam das mesmas propriedades que em V. Deste argumento se conclui que todo subespaço vetorial é, ele próprio, um espaço vetorial.

O próprio V é um subespaço vetorial dele mesmo. O subespaço $\{0\}$ é denominado de **subespaço trivial** de V. Os subespaços distintos de V são denominados de **subespaços próprios** de V.

Exemplo 2.23 *O conjunto* $W = \{\ (x, y, 0) : x, y \in \mathbf{R}\ \}$ *é um subespaço próprio de* $\mathbf{R}^3$.

Um subespaço vetorial sempre contém o vetor nulo. De fato, sendo 0 o escalar nulo, para todo vetor v do subespaço, $0v$ é o vetor nulo e, por definição, pertence ao subespaço.

Sejam W_1 e W_2 subespaços vetoriais de um espaço vetorial V. A **soma dos subespaços** $W_1 + W_2$, definida por

$$W_1 + W_2 = \{w_1 + w_2 : w_1 \in W_1 \quad \text{e} \quad w_2 \in W_2\}$$

e a **interseção dos subespaços** $W_1 \cap W_2$, definida por

$$W_1 \cap W_2 = \{w : w \in W_1 \quad \text{e} \quad w \in W_2\}$$

são subespaços vetoriais de V.

Nota 2.24 *Nem sempre a união*

$$W_1 \cup W_2 = \{w \in V : w \in W_1 \quad ou \quad w \in W_2\}$$

de dois subespaços W_1 *e* W_2 *de* V *é um subespaço vetorial de* V. *Se* u *e* v *pertencerem à união,* $u + v$ *pode não pertencer. Quando* W_1 *estiver contido em* W_2, *então* $W_1 \cup W_2 = W_2$ *e daí a união será um subespaço vetorial de* V.

Seja W um subepaço vetorial de V, um espaço vetorial com dimensão finita. Então $\dim(W) = \dim(V)$ se e só se $W = V$ e $\dim(W) < \dim(V)$ se e só se W for subespaço próprio de V.

Os subespaços de dimensão $n - 1$ de um espaço vetorial de dimensão n são chamados de **hiperplanos**.

Exemplo 2.25 *O subespaço vetorial* $W = \{\ x(1, 2, 0) + y(0, 3, 1) : x \text{ e } y \in \mathbf{R}\ \}$ *do* $\mathbf{R}^3$ *é um hiperplano.*

2.6 Subespaço gerado

Seja S um subconjunto de V. O conjunto de todas as combinações lineares finitas de elementos de S é um subespaço vetorial de V, chamado de **subespaço gerado** por S e é denotado por $\langle S \rangle$. Diz-se ainda que S **gera** $\langle S \rangle$ ou que $\langle S \rangle$ é **gerado** por S. O subespaço gerado por S é um subespaço vetorial de V.

Sendo $S = \{ w_1, \ldots, w_k \}$ finito, então

$$\langle S \rangle = \{\alpha_1 w_1 + \cdots + \alpha_k w_k : \alpha_1, \ldots, \alpha_k \in \mathbf{R}\}.$$

Exemplo 2.26 *Seja $S = \{e_1, e_3\}$ um subconjunto do $\mathbf{R}^3$ onde $e_1 = (1, 0, 0,)$ e $e_3 = (0, 0, 1)$. O subespaço gerado por S é $\langle S \rangle = \{ (x, 0, y) : x, y \in \mathbf{R}^3 \}$. Se considerarmos S como subconjunto de $\mathbf{C}^3$ então $\langle S \rangle = \{ (x, 0, y) : x, y \in \mathbf{C}\}$.*

Base do subespaço gerado

Seja $S = \{w_1, \ldots, w_k\}$ um conjunto de vetores de $\mathbf{C}^n$, de modo que

$$\begin{aligned} w_1 &= (w_{11}, w_{12}, \ldots, w_{1n}) \\ w_2 &= (w_{21}, w_{22}, \ldots, w_{2n}) \\ &\cdots \\ w_k &= (w_{k1}, w_{k2}, \ldots, w_{kn}) \end{aligned}$$

Vamos descrever um processo para determinar uma base para o espaço gerado por S no qual lançamos mão de alguns fatos para nos auxiliar nesta tarefa. Vamos enumerá-los abaixo.

1. Retirar os vetores nulos de S não altera o espaço gerado por S.

2. Permutar a ordem dos vetores de S não altera o espaço gerado por S.

3. Se multiplicarmos um ou mais vetores de S por escalares não nulos, o espaço gerado por S não se altera.

4. Se substituirmos em S o vetor w_i pelo vetor $w_i + cw_j$, onde c é um escalar, o espaço gerado por S permanece inalterado.

5. Se nenhum vetor de S for nulo e a matriz

$$\begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{k1} & w_{k2} & \cdots & w_{kn} \end{bmatrix}$$

   for escalonada, então S é linearmente independente e, portanto, é uma base do espaço gerado por S.

Os fatos enumerados acima nos permitem usar o método da eliminação de Gauss para determinar uma base para o espaço gerado por S : Construa a matriz cujas linhas são os elementos de $w_1,\ w_2,\ \ldots,\ w_k$, como acima e obtenha sua forma escalonada usando o método da eliminação de Gauss. As linhas não nulas da forma escalonada desta matriz

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ 0 & 0 & \cdots & r_{3n} \\ \vdots & \vdots & \ddots & \vdots \end{bmatrix}$$

formarão a base de $\langle S \rangle$.

Este procedimento pode ser usado para determinar o subespaço gerado por $S = \{\ w_1,\ \ldots,\ w_k\}$ mesmo quando S for um conjunto de vetores num espaço vetorial de dimensão finita V qualquer. Basta tomar uma base $\mathcal{B} = \{\ v_1,\ v_2,\ \ldots,\ v_n\}$ de V e decompor cada elementos de S numa combinação linear de elementos de $\mathcal{B}$

$$\begin{aligned} w_1 &= \beta_{11} v_1 + \beta_{12} v_2 + \cdots + \beta_{1n} v_n \\ w_2 &= \beta_{21} v_1 + \beta_{22} v_2 + \cdots + \beta_{2n} v_n \\ &\cdots \\ w_k &= \beta_{k1} v_1 + \beta_{k2} v_2 + \cdots + \beta_{kn} v_n \end{aligned}$$

formar a matriz

$$\begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1n} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{k1} & \beta_{k2} & \cdots & \beta_{kn} \end{bmatrix}$$

e proceder como no caso em que o espaço vetorial é o $\mathbf{C}^n$, obtendo, obter sua forma escalonada

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ 0 & 0 & \cdots & r_{3n} \\ \vdots & \vdots & \ddots & \vdots \end{bmatrix}$$

Os vetores não nulos obtidos na forma escalonada

$$\begin{gathered} r_{11} v_1 + r_{12} v_2 + \cdots + r_{1n} v_n \\ r_{22} v_2 + r_{23} v_3 + \cdots + r_{2n} v_n \\ r_{33} v_3 + r_{34} v_4 + \cdots + r_{3n} v_n \\ \vdots \end{gathered}$$

formarão uma base para o espaço gerado por V.

Exemplo 2.27 *Vamos determinar uma base do subespaço vetorial do* $\mathbf{R}^5$ *gerado por*

$$\begin{aligned} w_1 &= (1,2,2,-3,-4), \\ w_2 &= (3,8,0,2,8), \\ w_3 &= (1,2,2,-1,0), \\ w_4 &= (-1,-2,8,8,8), \\ w_5 &= (2,6,3,5,9). \end{aligned}$$

Construímos a matriz

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

e a escalonamos

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -3 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ -2 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 2 & -3 & -4 \\ 3 & 8 & 0 & 2 & 8 \\ 1 & 2 & 2 & -1 & 0 \\ -1 & -2 & 8 & 8 & 8 \\ 2 & 6 & 3 & 5 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 2 & -3 & -4 \\ 0 & 2 & -6 & 11 & 20 \\ 0 & 0 & 0 & 2 & 4 \\ 0 & 0 & 10 & 5 & 4 \\ 0 & 2 & -1 & 11 & 17 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 2 & -3 & -4 \\ 0 & 2 & -6 & 11 & 20 \\ 0 & 0 & 0 & 2 & 4 \\ 0 & 0 & 10 & 5 & 4 \\ 0 & 2 & -1 & 11 & 17 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 2 & -3 & -4 \\ 0 & 2 & -6 & 11 & 20 \\ 0 & 0 & 0 & 2 & 4 \\ 0 & 0 & 10 & 5 & 4 \\ 0 & 0 & 5 & 0 & -3 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 2 & -3 & -4 \\ 0 & 2 & -6 & 11 & 20 \\ 0 & 0 & 0 & 2 & 4 \\ 0 & 0 & 10 & 5 & 4 \\ 0 & 0 & 5 & 0 & -3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 2 & -3 & -4 \\ 0 & 2 & -6 & 11 & 20 \\ 0 & 0 & 5 & 0 & -3 \\ 0 & 0 & 10 & 5 & 4 \\ 0 & 0 & 0 & 2 & 4 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 2 & -3 & -4 \\ 0 & 2 & -6 & 11 & 20 \\ 0 & 0 & 5 & 0 & -3 \\ 0 & 0 & 10 & 5 & 4 \\ 0 & 0 & 0 & 2 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 2 & -3 & -4 \\ 0 & 2 & -6 & 11 & 20 \\ 0 & 0 & 5 & 0 & -3 \\ 0 & 0 & 0 & 5 & 10 \\ 0 & 0 & 0 & 2 & 4 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1/5 & 0 \\ 0 & 0 & 0 & -2/5 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 2 & -3 & -4 \\ 0 & 2 & -6 & 11 & 20 \\ 0 & 0 & 5 & 0 & -3 \\ 0 & 0 & 0 & 5 & 10 \\ 0 & 0 & 0 & 2 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 2 & -3 & -4 \\ 0 & 2 & -6 & 11 & 20 \\ 0 & 0 & 5 & 0 & -3 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

e assim, uma base do espaço gerado por w_1, w_2, w_3, w_4, w_5 *é formada pelos vetores*

$$\begin{aligned} z_1 &= (1, 2, 2, -3, -4), \\ z_2 &= (0, 2, -6, 11, 20), \\ z_3 &= (0, 0, 5, 0, -3), \\ z_4 &= (0, 0, 0, 1, 2). \end{aligned}$$

Capítulo 3

Transformação linear

Neste capítulo consideraremos que os espaços vetoriais estão definidos em um mesmo corpo K. Nos exemplos, K será o corpo dos números reais ou o corpo dos números complexos. Sejam V e W dois espaços vetoriais sobre um mesmo corpo K. Uma função $L : V \to W$ é uma **transformação linear** se, para todo par de vetores v, w em V e todo escalar α do corpo K,

$$\begin{aligned} L(v+w) &= L(v) + L(w), \\ L(\alpha v) &= \alpha L(v). \end{aligned}$$

A notação Lv também é usada para indicar $L(v)$.

Podemos unir as duas igualdades acima dizendo que L é linear quando a igualdade

$$L(\alpha v + \beta w) = \alpha L(v) + \beta L(w)$$

se verificar para todo α e β escalares e para todo v e w em V.

Toda transformação linear leva o zero de V no zero de W. De fato, $L(0) = L(0+0) = L(0) + L(0) = 2L(0)$ o que implica em $L(0) = 0$.

Exemplo 3.1 *A transformação $L_1 : \mathbf{R}^2 \to \mathbf{R}$ definida por $L_1(x, y) = 3x + 2y$ é linear.*

A transformação $L_2 : \mathbf{R}^2 \to \mathbf{R}^2$ definida por $L_2(x, y) = (x - y, 0)$ é linear.

A transformação $T : \mathbf{R}^2 \to \mathbf{R}$ definida por $T(x, y) = x + y + 2$ não é linear pois $T(2x, 2y) = 2x + 2y + 2$ é diferente de $2T(x, y) = 2x + 2y + 4$.

Uma transformação linear $L : V \to V$ de um espaço V sobre ele mesmo recebe o nome de **operador linear**. Se V for um espaço vetorial sobre um corpo K, uma transformação linear $L : V \to K$ recebe o nome de **funcional linear**.

Sendo $L : V \to W$ e $T : W \to U$, definimos a composta $T \circ L : V \to U$ por

$$T \circ L(v) = T(L(v)).$$

Também se denota $T \circ L$ por TL. Assim, $TL(v) = T(L(v))$.

Pode-se provar por indução que, se $L : V \to W$ for linear, se $v_1, \ldots, v_n$ forem vetores de V e se $\alpha_1, \ldots, \alpha_n$ forem escalares, então

$$L(\alpha_1 v_1 + \cdots + \alpha_n v_n) = \alpha_1 L(v_1) + \cdots + \alpha_n L(v_n).$$

A partir desta fórmula podemos afirmar que quando L for linear e $\{v_1, \ldots, v_n\}$ for base de V, o conhecimento dos vetores $w_1 = L(v_1), \ldots, w_n = L(v_n)$ é suficiente para calcular o valor de L em qualquer vetor v. Basta decompor v em uma combinação linear dos vetores da base

$$v = x_1 v_1 + \cdots + x_n v_n$$

e calcular

$$L(v) = L(x_1 v_1 + \cdots + x_n v_n) = x_1 L(v_1) + \cdots + x_n L(v_n) = x_1 w_1 + \cdots + x_n w_n.$$

Exemplo 3.2 *Seja $L : \mathbf{R}^3 \to \mathbf{R}$ uma transformação linear e $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$ e $e_3 = (0, 0, 1)$ os elementos da base canônica do $\mathbf{R}^3$. Se $Le_1 = 5$, $Le_2 = 7$, $Le_3 = 11$, para qualquer (x_1, x_2, x_3) em $\mathbf{R}^3$, teremos*

$$\begin{aligned} L(x_1, x_2, x_3) &= L(x_1 e_1 + x_2 e_2 + x_3 e_3) \\ &= x_1 L(e_1) + x_2 L(e_2) + x_3 L(e_3) \\ &= 5x_1 + 7x_2 + 11x_3. \end{aligned}$$

Generalizando este exemplo, se $Le_1 = a_1$, $Le_2 = a_2$ e $Le_3 = a_3$, então

$$\begin{aligned} L(x_1, x_2, x_3) &= L(x_1 e_1 + x_2 e_2 + x_3 e_3) \\ &= x_1 L(e_1) + x_2 L(e_2) + x_3 L(e_3) \\ &= a_1 x_1 + a_2 x_2 + a_3 x_3. \end{aligned}$$

Este exemplo nos dá uma indicação da forma geral de um funcional linear L de $\mathbf{R}^n$ em $\mathbf{R}$. Vamos determiná-la. Seja $\{e_1, e_2, \ldots, e_n\}$ a base canônica do $\mathbf{R}^n$. Se $L(e_i) = a_i$ para $i = 1, \ldots, n$, então, para todo $(x_1, \ldots, x_n)$ vale

$$L(x_1, \ldots, x_n) = L(x_1 e_1 + \cdots + x_n e_n) = x_1 L(e_1) + \cdots + x_n L(e_n)$$

ou

$$L(x_1, \ldots, x_n) = a_1 x_1 + \cdots + a_n x_n.$$

Esta é a forma geral de um funcional linear do $\mathbf{R}^n$ em $\mathbf{R}$.

Exemplo 3.3 *Utilizando a forma geral, vemos que $L_1(x, y) = 5x - 4y$ e $L_2(x, y) = 3x$ são transformações lineares de $\mathbf{R}^2$ em $\mathbf{R}$. Todavia, $T(x, y) = x + 2$ não é linear pois não possui o formato estabelecido acima e observe que T não leva o zero de $\mathbf{R}^2$ no zero de $\mathbf{R}$.*

Vamos determinar agora a forma geral de uma uma transformação linear L de $\mathbf{R}^n$ em $\mathbf{R}^m$. Iniciemos com um exemplo ilustrativo com a transformação

$$L(x_1, x_2) = (x_1 - 3x_2, 2x_1, -x_1 + 4x_2)$$

de $\mathbf{R}^2$ em $\mathbf{R}^3$. Se definirmos

$$L_1(x_1, x_2) = x_1 - 3x_2, \quad L_2(x_1, x_2) = 3x_2 \quad \text{e} \quad L_3(x_1, x_2) = -x_1 + 4x_2$$

então

$$L(x_1, x_2) = (\ L_1(x_1, x_2),\ L_2(x_1, x_2),\ L_3(x_1, x_2)\ ).$$

Baseados neste exemplo, vemos que, se L é uma transformação linear do $\mathbf{R}^n$ em $\mathbf{R}^m$, para qualquer x no $\mathbf{R}^n$, tem-se

$$L(x) = (\ L_1(x),\ \ldots,\ L_m(x)\ ),$$

onde $L_1(x)$, ..., $L_m(x)$ são números reais, dependentes de x. Vamos mostrar que L_1, ..., L_m são funcionais lineares de $\mathbf{R}^n$ em $\mathbf{R}$. De fato, sendo α e β escalares e x, y ênuplas ordenadas, então

$$L(\alpha x + \beta y) = \alpha Lx + \beta Ly$$

e assim,

$$(\ L_1(\alpha x + \beta y),\ \ldots,\ L_m(\alpha x + \beta y)\ ) = (\ \alpha L_1 x + \beta L_1 y,\ \ldots,\ \alpha L_m x + \beta L_m y\ )$$

e, da igualdade desses elementos de $\mathbf{R}^m$, obtemos

$$\begin{aligned} L_1(\alpha x + \beta y) &= \alpha L_1 x + \beta L_1 y \\ &\cdots \\ L_m(\alpha x + \beta y) &= \alpha L_m x + \beta L_m y \end{aligned}$$

mostrando que L_1, ..., L_m são funcionais lineares de $\mathbf{R}^n$ em $\mathbf{R}$. A partir daí, concluímos que toda transformação linear L de $\mathbf{R}^n$ em $\mathbf{R}^m$ é da forma

$$L(x_1, \ldots, x_n) = (\ a_{11}x_1 + \cdots + a_{1n}x_n,\ \ldots,\ a_{m1}x_1 + \cdots + a_{m,n}x_n\ )$$

onde a_{ij}, para $i = 1$, ..., m e $j = 1$, ..., n, são números reais.

Exemplo 3.4 *A transformação*

$$L(x, y) = (2x - y, x + y, y)$$

de $\mathbf{R}^2$ *em* $\mathbf{R}^3$ *é linear. A transformação*

$$T(x, y) = (x, 0, x + 2y, -3x + y, 4y)$$

de $\mathbf{R}^2$ *em* $\mathbf{R}^5$ *é linear.*

Seja $L : V \to W$ uma transformação linear. O conjunto

$$\ker L = \{v \in V : Lv = 0\}$$

é chamado de **núcleo** (kernel em inglês) de L e o conjunto

$$\operatorname{Im} L = \{Lv : v \in V\}$$

é a **imagem** de L. Tanto o núcleo de L quanto a sua imagem, são subespaços vetoriais de V. A dimensão do núcleo de L é denominada de **nulidade** de L.

Exemplo 3.5 *Seja L a transformação linear de $\mathbf{R}^3$ em $\mathbf{R}^3$ definida por*

$$L(x, y, z) = (x + z, 2x + y + z, x + 2y - z).$$

Para determinar o núcleo de L escreva

$$L(x, y, z) = (x + z, 2x + y + z, x + 2y - z) = (0, 0, 0)$$

e resolva o sistema correspondente à igualdade acima

$$\begin{aligned} x + z &= 0 \\ 2x + y + z &= 0 \\ x + 2y - z &= 0 \end{aligned}$$

cuja solução $x = -z$ e $y = z$ pode ser obtida pelo método da eliminação de Gauss. Assim,

$$\ker L = \{(-z, z, z) : z \in \mathbf{R}\}.$$

O $\ker L$ *é gerado por* $(-1,\ 1,\ 1)$ *e, portanto, tem dimensão* 1.

Para determinar a imagem de L escrevemos

$$\begin{aligned} L(x, y, z) &= (x + z, 2x + y + z, x + 2y - z) \\ &= x(1, 2, 1) + y(0, 1, 2) + z(1, 1, -1) \end{aligned}$$

mostrando que todo elemento da imagem de L é uma combinação linear dos vetores (1, 2, 1), (0, 1, 2) *e* (1, 1, −1). *Para determinar uma base do espaço gerado usamos o processo de escalonamento. Construímos a matriz*

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

cujas linhas são os elementos dos vetores que geram o subespaço. Usando operações elementares sobre as linhas chegamos a

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 2 \\ 1 & 1 & -1 \end{bmatrix} \overset{l_3 = l_3 - l_1}{\longrightarrow} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 2 \\ 0 & -1 & -2 \end{bmatrix} \overset{l_3 = l_3 - l_2}{\longrightarrow} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

e chegamos a uma base da imagem de L que é formada pelos vetores

$$(1,2,1)\ ,\quad (0,1,2)$$

e concluímos que a imagem de L tem dimensão 2.

Adicionando a dimensão do núcleo com a dimensão da imagem obtemos 3 *que é a dimensão do domínio de L.*

O resultado do exemplo anterior em que a soma das dimensões do núcleo e da imagem de L é igual à dimensão do domínio de L é um resultado geral, como enuncia o próximo teorema.

Teorema 3.6 *Sejam V e W espaços vetoriais e $L : V \to W$ linear. Se a dimensão de V for finita, então*

$$\dim V = \dim \mathrm{Im}\,(L) + \dim \ker(L).$$

Prova. Quando L é a transformação linear nula, que leva todos os vetores de V no zero a $\mathrm{Im}\,(L) = \{0\}$ e nada resta a provar, uma vez que $\ker(L) = V$. Assim, como $\dim \mathrm{Im}\,(L) = 0$ e $\dim \ker(L) = \dim(V)$.

Se L não for a transformação linear nula, $\mathrm{Im}\,(L) \neq \{0\}$. Seja $\{v_1, \ldots, v_p\}$ uma base do $\ker(L)$. Podemos acrescentar vetores a este conjunto até obter uma base $\mathcal{B} = \{v_1, \ldots, v_p, v_{p+1}, \ldots, v_n\}$ de V. Se provarmos que $\{L(v_{p+1}), \ldots, L(v_n)\}$ é base da $\mathrm{Im}\,(L)$, o teorema estará provado pois

$$\dim \ker(L) + \dim \mathrm{Im}\,(L) = p + (n-p) = n = \dim(V).$$

Inicialmente, observamos que nenhum dos vetores $L(v_{p+1}), \ldots, L(v_n)$ é nulo. Se fosse, o vetor correspondente pertenceria ao núcleo de L e $\mathcal{B}$ seria linearmente dependente, o que não é o caso pois é base de V.

Provemos agora que $\{L(v_{p+1}), \ldots, L(v_n)\}$ é base da $\mathrm{Im}\,(L)$.

1. O conjunto $\{L(v_{p+1}), \ldots, L(v_n)\}$ gera $\mathrm{Im}\,(L)$.

   De fato, se w pertence à $\mathrm{Im}\,(L)$, então existe v em V tal que $w = Lv$. Podemos escrever v como uma combinação linear dos elementos da base $\mathcal{B}$,

   $$v = x_1 v_1 + \cdots + x_p v_p + x_{p+1} v_{p+1} + \cdots + x_n v_n$$

   de onde segue

   $$\begin{aligned} w &= Lv = L(x_1 v_1 + \cdots + x_p v_p + x_{p+1} v_{p+1} + \cdots + x_n v_n) \\ &= x_1 L v_1 + \cdots + x_p L v_p + x_{p+1} L v_{p+1} + \cdots + x_n L v_n \\ &= x_{p+1} L v_{p+1} + \cdots + x_n L v_n, \end{aligned}$$

   mostrando que $\{L v_{p+1}, \ldots, L v_n\}$ gera a $\mathrm{Im}\,(L)$.

2. O conjunto $\{Lv_{p+1}, \ldots, Lv_n\}$ é linearmente independente.

Se fosse linearmente dependente, existiriam escalares $k_{p+1}, \ldots, k_n$, nem todos nulos, tais que

$$k_{p+1}L(v_{p+1}) + \cdots + k_n L(v_n) = 0$$

o que implica em

$$L(k_{p+1}v_{p+1} + \cdots + k_n v_n) = 0$$

indicando que o vetor $k_{p+1}v_{p+1} + \cdots + k_n v_n$ pertenceria ao $\ker(L)$, sendo igual a uma combinação linear dos vetores $v_1, \ldots, v_p$ que formam uma base do núcleo de L. Portanto, existem escalares $k_1, \ldots, k_p$ tais que

$$k_{p+1}v_{p+1} + \cdots + k_n v_n = k_1 v_1 + \cdots + k_p v_p$$

ou

$$k_1 v_1 + \cdots + k_p v_p - k_{p+1}v_{p+1} - \cdots - k_n v_n = 0$$

onde pelo menos um dos k_i, com $i = 1, \ldots, p$, diferente de zero, o que vai contraria a hipótese de $\mathcal{B}$ ser base de V.

Das partes 1 e 2 concluímos que $\{Lv_{p+1}, \ldots, Lv_n\}$ é base da $\mathrm{Im}\,(L)$.

Como $\{v_1, \ldots, v_p\}$ é base do $\ker(L)$ e $\{Lv_{p+1}, \ldots, Lv_n\}$ é base da $\mathrm{Im}\,(L)$, então

$$\dim \ker L + \dim \mathrm{Im}\, L = p + (n - p) = n = \dim V$$

e o teorema está provado. □

3.1 Matriz de uma transformação linear

Seja $\mathcal{B}_1 = \{v_1, \ldots, v_n\}$ uma base de um espaço vetorial V, $\mathcal{B}_2 = \{w_1, \ldots, w_m\}$ uma base de um espaço vetorial W e $L : V \to W$ uma transformação linear. Podemos decompor cada vetor Lv_j, com $j = 1, \ldots, n$, numa combinação linear dos elementos da base $\mathcal{B}_2$

$$\begin{aligned}
Lv_1 &= a_{11}w_1 + a_{21}w_2 + \cdots + a_{m1}w_m \\
Lv_2 &= a_{12}w_1 + a_{22}w_2 + \cdots + a_{m2}w_m \\
&\cdots \\
Lv_n &= a_{1n}w_1 + a_{2n}w_2 + \cdots + a_{mn}w_m
\end{aligned}$$

Estas expressões podem ser escritas de modo taquigráfico usando somatório: para $j = 1, 2, \ldots, n$

$$Lv_j = \sum_{i=1}^{m} a_{ij} w_i.$$

A matriz m por n

$$[L]_{12} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

é denominada de **matriz** de L nas bases $\mathcal{B}_1$ e $\mathcal{B}_2$. Ainda se diz que $[L]_{12}$ é a **representação matricial** de L nas bases $\mathcal{B}_1$ e $\mathcal{B}_2$.

Quando $W = V$ e $\mathcal{B}_2 = \mathcal{B}_1$, a matriz $[L]_{12}$ é denotada por $[L]_1$ e denominada de matriz de L na base $\mathcal{B}_1$. Também se diz que $[L]_1$ é a representação matricial de L na base $\mathcal{B}_1$. Para simplificar a notação, podemos escrever $[L]$ em lugar de $[L]_{12}$ ou $[L]_1$, conforme o caso, sempre que o contexto for claro quanto às bases envolvidas.

Teorema 3.7 *Sejam $\mathcal{B}_1$ e $\mathcal{B}_2$ bases dos espaços vetoriais V e W, respectivamente. Seja $L : V \to W$ uma transformação linear e v um vetor de V. Se $[v]_1$ for a matriz de v na base $\mathcal{B}_1$, $[Lv]_2$ a matriz de Lv na base $\mathcal{B}_2$ e $[L]_{12}$ for ma matriz de Lv nas bases $\mathcal{B}_1$ e $\mathcal{B}_2$ então*

$$[Lv]_2 = [L]_{12}[v]_1.$$

Prova. Sejam $\mathcal{B}_1 = \{v_1, \ldots, v_n\}$ e $\mathcal{B}_2 = \{w_1, \ldots, w_m\}$ as bases de V e W. Se

$$v = \sum_{j=1}^{n} x_j v_j, \qquad Lv = \sum_{i=1}^{m} y_i w_i, \qquad Lv_j = \sum_{i=1}^{m} a_{ij} w_i,$$

então $[v]_1 = [x_1, \ldots, x_n]^T$, $[Lv]_2 = [y_1, \ldots, y_m]^T$ são matrizes coluna e $[L]_{12} = [a_{ij}]$ é uma matriz retangular $m \times n$.

Por um lado,

$$\begin{aligned} Lv &= L\left(\sum_j x_j v_j\right) = \sum_j x_j L(v_j) \\ &= \sum_j x_j \sum_i a_{ij} w_i = \sum_i \left(\sum_j a_{ij} x_j\right) w_i \end{aligned}$$

e por outro,

$$Lv = \sum_i y_i w_i.$$

Da unicidade da decomposição de um vetor nos elementos da base,

$$y_i = \sum_{j=1}^{n} a_{ij} x_j$$

para $i = 1, \ldots, m$ que equivale à igualdade matricial

$$[Lv]_2 = [L]_{12}[v]_1.$$

$\square$

Teorema 3.8 *Sejam $\mathcal{B}_1$, $\mathcal{B}_2$ e $\mathcal{B}_3$, respectivamente, bases dos espaços vetoriais V, W e U. Sejam $L_1 : V \to W$ e $L_2 : W \to U$ lineares. Então*

$$[L_2 \circ L_1]_{13} = [L_2]_{23}[L_1]_{12}.$$

Prova. Sejam $\mathcal{B}_1 = \{v_1, \ldots, v_n\}$, $\mathcal{B}_2 = \{w_1, \ldots, w_m\}$ e $\mathcal{B}_3 = \{u_1, \ldots, u_p\}$ as bases em questão. Se

$$\begin{aligned} L_1 v_j &= \sum_{i=1}^{m} a_{ij} w_i, \\ L_2 w_i &= \sum_{k=1}^{p} b_{ki} u_k, \\ L_2 L_1 v_j &= \sum_{k=1}^{p} c_{kj} u_k, \end{aligned}$$

então $[L_1]_{12} = [a_{ij}]$, $[L_2]_{23} = [b_{ki}]$ e $[L_2 L_1]_{13} = [c_{kj}]$. Como

$$\begin{aligned} L_2 L_1 v_j &= L_2 \left(\sum_{i=1}^{m} a_{ij} w_i \right) = \sum_{i=1}^{m} a_{ij} L_2 (w_i) \\ &= \sum_{i=1}^{m} a_{ij} \sum_{k=1}^{p} b_{ki} u_k = \sum_{k=1}^{p} \left(\sum_{i=1}^{m} b_{ki} a_{ij} \right) u_k \end{aligned}$$

segue

$$c_{kj} = \sum_{i=1}^{m} b_{ki} a_{ij}$$

para $k = 1, \ldots, p$ e $j = 1, \ldots, n$, que resulta na igualdade matricial

$$[L_2 L_1]_{13} = [L_2]_{23}[L]_{12}.$$

□

Teorema 3.9 *Sejam $\mathcal{B}_1$ e $\mathcal{B}_2$ duas bases do espaço vetorial V e $L : V \to V$ um operador linear. Seja M_{12} a matriz de mudança da base $\mathcal{B}_1$ para a base $\mathcal{B}_2$. Então*

$$M_{12}\ [L]_2 = [L]_1\ M_{12} \qquad ou \qquad [L]_2 = M_{12}^{-1}\ [L]_1\ M_{12}.$$

Prova. Sejam $\mathcal{B}_1 = \{v_1, \ldots, v_n\}$ e $\mathcal{B}_2 = \{w_1, \ldots, w_n\}$ as bases em questão. Seja $M_{12} = [m_{ij}]$ a matriz de mudança da base $\mathcal{B}_1$ para a base $\mathcal{B}_2$, de modo que

$$w_j = \sum_{i=1}^{n} m_{ij} v_i.$$

Sejam $[L]_1 = [a_{ij}]$ e $[L]_2 = [b_{ij}]$ as matrizes de L nas bases $\mathcal{B}_1$ e $\mathcal{B}_2$, respectivamente. Podemos escrever, para $j = 1, \ldots, n$,

$$Lv_j = \sum_{i=1}^{n} a_{ij} v_i, \qquad \text{e} \qquad Lw_j = \sum_{i=1}^{n} b_{ij} w_i.$$

Por um lado,

$$Lw_j = \sum_k b_{kj} w_k = \sum_k b_{kj} \sum_i m_{ik} v_i = \sum_i \left(\sum_k m_{ik} b_{kj} \right) v_i$$

e por outro,

$$\begin{aligned} Lw_j &= L\left(\sum_k m_{kj} v_k \right) = \sum_k m_{kj} L\left(v_k\right) = \sum_k m_{kj} \sum_i a_{ik} v_i \\ &= \sum_i \left(\sum_k a_{ik} m_{kj} \right) v_i. \end{aligned}$$

Pela unicidade de decomposição de vetores numa base, segue

$$\sum_k m_{ik} b_{kj} = \sum_k a_{ik} m_{kj},$$

igualdade válida para $i = 1, \ldots, n$ e $j = 1, \ldots, m$. Usando a notação matricial, conclui-se a prova do teorema:

$$M_{12}\ [L]_2 = [L]_1\ M_{12}.$$

$\square$

Definição 3.10 *Duas matrizes A e B são* ***semelhantes*** *se existir uma matriz inversível P tal que*

$$B = P^{-1} A P.$$

De acordo com o teorema anterior, as representações matriciais de uma transformação linear $L : V \to V$ são semelhantes.

3.2 Isomorfismo

Sejam V e W dois espaços vetoriais sobre o mesmo corpo. Uma tranformação $L : V \to W$ é **injetora** se, para todo $v_1 \neq v_2$ em V, então $L(v_1) \neq L(v_2)$. De forma equivalente, L é injetora se para todo v_1 e v_2 em V com $L(v_1) = L(v_2)$ tem-se $v_1 = v_2$.

Teorema 3.11 *Sejam V e W espaços vetoriais sobre o mesmo corpo. Seja $L : V \to W$ linear. L é injetora se e só se* $\ker(L) = \{0\}$. *Em outras palavras, L é injetora se e só se o zero é o único vetor levado por L em zero.*

Uma transformação $L : V \to W$ é **sobrejetora** se, para todo w em W, existe pelo menos um v em V tal que $Lv = w$. Uma transformação **bijetora** é aquela que é ao mesmo tempo injetora e sobrejetora. As transformações bijetoras $L : V \to W$ possuem inversa $L^{-1} : W \to V$ que, por sua vez, é bijetora e sua inversa é L.

Denominamos **isomorfismo** à transformação $L : V \to W$ que é linear e bijetora. Neste caso sua inversa $L^{-1} : W \to V$ é linear e, portanto, um isomorfismo. Dois espaços vetoriais V e W são **isomorfos** quando houver um isomorfismo de V em W.

Teorema 3.12 *Dois espaços vetoriais V e W sobre um mesmo corpo e de dimensão finita são isomorfos se e só se* $\dim V = \dim W$.

Prova. Sejam V e W isomorfos e $L : V \to W$ um isomorfismo entre eles. Dada uma base $\{v_1, \ldots, v_n\}$ de V, vamos mostrar que $\{Lv_1, \ldots, Lv_n\}$ é base de W.

1. Provemos que $\{Lv_1, \ldots, Lv_n\}$ gera W. Sendo L bijetora, para qualquer w em W, existe v em V tal que $w = Lv$. Decompondo v na base $\{v_1, \ldots, v_n\}$, obtemos $v = x_1v_1 + \cdots + x_nv_n$ e $w = Lv = x_1Lv_1 + \cdots + x_nLv_n$, provando que $\{Lv_1, \ldots, Lv_n\}$ gera W.

2. Provemos que $\{Lv_1, \ldots, Lv_n\}$ é linearmente independente. Sejam $k_1, \ldots, k_n$ escalares tais que $k_1Lv_1 + \cdots + k_nLv_n = 0$. Então $L(k_1v_1 + \cdots + k_nv_n) = 0$ e, como $L(0) = 0$ e L é bijetora, segue $k_1v_1 + \cdots + k_nv_n = 0$. Pelo fato de $\{v_1, \ldots, v_n\}$ ser base de V, concluímos que $k_1 = \cdots = k_n = 0$, provando a independência linear do conjunto $\{Lv_1, \ldots, Lv_n\}$.

As partes 1 e 2 provam que $\dim V = \dim W$.

Tomando $\dim V = \dim W = n$ como hipótese, provemos que V e W são isomorfos. Seja $\mathcal{B}_1 = \{v_1, \ldots, v_n\}$ base de V e $\mathcal{B}_2 = \{w_1, \ldots, w_n\}$ base de W. Seja L a transformação linear de V em W que leva v_i em w_i, para $i = 1, \ldots, n$, ou seja $Lv_i = w_i$. Vamos provar que L é um isomorfismo entre V e W.

3. Provemos que L é sobrejetor. Dado qualquer s em W, podemos escrever
$$\begin{aligned} s &= s_1w_1 + \cdots + s_nw_n = s_1Lv_1 + \cdots + s_nLv_n \\ &= L(s_1v_1 + \cdots + s_nv_n), \end{aligned}$$
provando que s está na imagem de L, provando a sobrejetividade de L.

4. Provemos que L é injetor. Sejam $x = x_1v_1 + \cdots + x_nv_n$ e $y = y_1v_1 + \cdots + y_nv_n$ dois vetores de V tais que $Lx = Ly$. Logo,
$$L(x_1v_1 + \cdots + x_nv_n) = L(y_1v_1 + \cdots + y_nv_n)$$

ou seja,

$$x_1w_1 + \cdots + x_nw_n = y_1w_1 + \cdots + y_nw_n.$$

Da independência linear de $\mathcal{B}_2$, concluímos que $x_1 = y_1, \ldots, x_n = y_n$, de onde resulta a igualdade $x = y$, provando que L é injetora.

De 3 e 4 concluímos que L é um isomorfismo. □

Teorema 3.13 *Sejam V e W espaços vetoriais sobre o mesmo corpo, ambos com a mesma dimensão e $L : V \to W$ linear. São equivalentes:*

1. *L é um isomorfismo.*
2. *L é sobrejetora.*
3. *L é injetora.*
4. *Se $L(v) = 0$, então $v = 0$. Em outras palavras,* $\ker(L) = \{0\}$.

Prova. Provemos que (2) implica em (3). Seja L sobrejetora e $\{w_1, \ldots, w_n\}$ uma base de W. Da sobrejetividade de L, existe um conjunto de vetores $\mathcal{B} = \{v_1, \ldots, v_n\}$ em V tais que $Lv_i = w_i$, para $i = 1, \ldots, n$. O conjunto $\mathcal{B}$ é base de V. Sejam $x = x_1v_1 + \cdots + x_nv_n$ e $y = y_1v_1 + \cdots + y_nv_n$ dois vetores de V tais que $Lx = Ly$. Desta igualdade segue $x_1w_1 + \cdots + x_nw_n = y_1w_1 + \cdots + y_nw_n$. A independência linear de $\{w_1, \ldots, w_n\}$ implica em $x_1 = y_1, \ldots, x_n = y_n$, ou $x = y$, provando a injetividade de L.

Provemos que (3) implica em (4). Se L é injetora e $L(v) = 0$, como $L(0) = 0$, segue que $v = 0$, provando que (3) implica em (4).

Provemos que (4) implica em (2). Sendo $\ker(L) = \{0\}$ e $\{v_1, \ldots, v_n\}$ uma base de V, então $\{Lv_1, \ldots, Lv_n\}$ é uma base de W. De fato, se $k_1, \ldots, k_n$ forem escalares tais que $k_1Lv_1 + \cdots + k_nLv_n = 0$, então $L(k_1v_1 + \cdots + k_nv_n) = 0$ e, como o núcleo de L contém apenas o zero, $k_1v_1 + \cdots + k_nv_n = 0$. A independência linear dos v_i acarreta em $k_1 = \cdots = k_n = 0$, provando que conjunto $\{Lv_1, \ldots, Lv_n\}$ é linearmente independente e, portanto, base de W. Logo, L é sobrejetor. □

Teorema 3.14 *Sejam $\mathcal{B}_1$ e $\mathcal{B}_2$ bases dos espaços V e W respectivamente e $L : V \to W$ um isomorfismo. Se A for a representação matricial de L nas bases $\mathcal{B}_1$ e $\mathcal{B}_2$, então A^{-1} será a representação matricial de L^{-1} nas bases $\mathcal{B}_2$ e $\mathcal{B}_1$.*

Prova. Seja $A = [a_{ij}]$ a representação matricial de L nas bases $\mathcal{B}_1$ e $\mathcal{B}_2$. Seja $B = [b_{ij}]$ a representação matricial de L^{-1} nas bases $\mathcal{B}_2$ e $\mathcal{B}_1$. Como $L^{-1}L$ o operador identidade, BA é a matriz da transformação identidade na base $\mathcal{B}_1$ e esta é a matriz identidade. Ainda temos que LL^{-1} é o operador identidade e, desta maneira, AB é a matriz da transformação identidade na base $\mathcal{B}_2$ e esta é a matriz identidade o que prova ser $B = A^{-1}$. □

Teorema 3.15 *Todo um espaço vetorial V de dimensão n sobre um corpo K é isomorfo a K^n.*

Prova. Se $\{v_1, \ldots, v_n\}$ for uma base de V, então definimos a transformação linear $L : V \to K^n$ por $Lv_i = e_i$, onde $\{e_1, \ldots, e_n\}$ é a base canônica do K^n, isto é, $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ onde o único elemento não nulo é o i-ésimo. Como L leva uma base de V numa base de K^n ela é injetora e, portanto, um isomorfismo. □

Podemos afirmar que os isomorfismos são os mensageiros que trazem e levam as propriedades de um espaço vetorial a outro. Se dois espaços vetoriais vetoriais forem isomorfos, todas as propriedades de um podem ser levados ao outro pelo isomorfismo. Isto significa que, ao estudar as propriedades de um deles, teremos estudado as propriedades do outro.

Por esta razão, ao estudar um espaço vetorial real ou complexo V de dimensão n, os protótipos são o $\mathbf{R}^n$ e o $\mathbf{C}^n$, respectivamente. Se soubermos como proceder com um dos dois, saberemos como proceder com V. Se$\{v_1, \ldots, v_n\}$ for uma base de V, basta usar a correspondência

$$x_1v_1 + \cdots + x_nv_n \leftrightarrow (x_1, \ldots, x_n)$$

definida pelo isomorfismo estabelecido no teorema anterior.

3.3 Transformações lineares em $\mathbf{C}^{m\times 1}$

Seja A uma matriz complexa m por n. Então $L : \mathbf{C}^{n\times 1} \to \mathbf{C}^{m\times 1}$ definida por $L(x) = Ax$ é uma transformação linear.

Reciprocamente, vamos mostrar que toda transformação linear $L : \mathbf{C}^{n\times 1} \to \mathbf{C}^{m\times 1}$ é da forma $L(x) = Ax$, onde A é uma matriz m por n.

Se $\{e_1, \ldots, e_n\}$ for a base canônica do $\mathbf{C}^{n\times 1}$ então $a_j = L(e_j)$ são vetores coluna em $\mathbf{C}^{m\times 1}$. Dado o vetor coluna $x = [x_1, \ldots, x_n]^T = x_1e_1 + \cdots + x_ne_n$ do $\mathbf{C}^n$, então

$$\begin{aligned} L(x) &= L\left(\sum_{j=1}^{n} x_je_j\right) = \sum_{j=1}^{n} x_jL(e_j) \\ &= \sum_{j=1}^{n} x_ja_j = Ax \end{aligned}$$

onde $A = [a_1, \ldots, a_n]$ é a matriz complexa m por n, cujas colunas são $a_1, \ldots, a_n$.

Em síntese, toda transformação $L : \mathbf{C}^{n\times 1} \to \mathbf{C}^{m\times 1}$ é do tipo $L(x) = Ax$, onde A é uma matriz complexa de ordem m por n. Por esta razão, podemos dizer que a matriz A é uma transformação linear e escrever $A : \mathbf{C}^{n\times 1} \to \mathbf{C}^{m\times 1}$.

É intessante observar que, se $\mathcal{B}_1 = \{e_1, \ldots, e_n\}$ for a base canônica do $\mathbf{C}^{n\times 1}$, se $\mathcal{B}_2 = \{f_1, \ldots, f_m\}$ for a base canônica do $\mathbf{C}^{m\times 1}$ e se A for uma matriz complexa m por n, então a representação matricial da transformação linear $A : \mathbf{C}^{n\times 1} \to \mathbf{C}^{m\times 1}$ nas bases $\mathcal{B}_1$ e $\mathcal{B}_2$ é, exatamente, a matriz A.

Capítulo 4

Produto interno e norma

Neste capítulo trabalharemos apenas com espaços vetoriais sobre o corpo dos números reais ou sobre o corpo dos números complexos.

4.1 Produto interno em espaços vetoriais reais

Seja V um espaço vetorial sobre o corpo $\mathbf{R}$ dos números reais. Um **produto interno** em V é uma operação

$$\langle\ ,\ \rangle : V \times V \to \mathbf{R}$$

que possui as propriedades abaixo, válidas para todo v, w e z em V e todo a e b em $\mathbf{R}$:

1. O produto interno é positivo definido

$$\langle v, v\rangle \geq 0 \quad \text{e} \quad \langle v, v\rangle = 0 \text{ se e só se } v = 0.$$

2. O produto interno é simétrico

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

3. O produto interno é linear na segunda variável

$$\langle v, aw + bz\rangle = a\,\langle v, w\rangle + b\,\langle v, z\rangle\,.$$

Das propriedades (2) e (3) se conclui que o produto interno é liner na primeira variável

$$\langle av + bw, z\rangle = a\,\langle v, z\rangle + b\,\langle w, z\rangle\,.$$

Essas propriedades se extrai a linearidade do produto interno em relação à primeira e à segunda variável. Tanto é que, se $v_1, \ldots, v_p$ e $w_1, \ldots, w_q$ forem vetores de V e a_i, b_j forem números reais, então

$$\left\langle \sum_{i=1}^{p} a_i v_i, \sum_{j=1}^{q} b_j w_j \right\rangle = \sum_{i=1}^{p}\sum_{j=1}^{q} a_i b_j\,\langle v_i, w_j\rangle\,.$$

Exemplo 4.1 *Se x e y forem matrizes coluna $\mathbf{R}^{n\times1}$, o produto matricial x^Ty, onde x^T é a matriz transposta de x, é denominado **produto escalar** das matrizes x e y. Um produto interno em $\mathbf{R}^{n\times1}$ é proveniente do produto escalar $\langle x,y\rangle = x^Ty$.*

Exemplo 4.2 *Seja $\mathcal{P}_2(\mathbf{R})$ o conjunto dos polinômios com coeficientes reais e grau menor ou igual a 2. Neste espaço vetorial,*

$$\left\langle a_0 + a_1x + a_2x^2, b_0 + b_1x + b_2x^2\right\rangle = a_0b_0 + a_1b_1 + a_2b_2$$

é um produto interno e

$$\langle f(x), g(x)\rangle = \int_{-1}^{1} f(x)g(x)dx$$

é outro.

Exemplo 4.3 *Seja $\mathcal{C}\,[a,b]$ o conjunto das funções reais de variável real, definidas e contínuas no intervalo $[a,b]$. Com as operações de adição de funções e a multiplicação de um número real por uma função, $\mathcal{C}\,[a,b]$ é um espaço vetorial sobre o corpo de números reais. Nele*

$$\langle f,g\rangle = \int_a^b f(t)g(t)\ dt$$

é um produto interno.

4.2 Produto interno em espaços vetoriais complexos

Vamos agora definir produto interno em um espaço vetorial sobre o corpo $\mathbf{C}$ dos números complexos. O corpo dos números complexos é formado pelo conjunto de pares ordenados (a,b) de números reais, no qual definimos duas operações, uma de adição e outra de multiplicação. Os elementos desse corpo são denominados **números complexos**. Dois números complexos (a,b) e (c,d) são iguais quando $a = c$ e $b = d$ e se escreve $(a,b) = (c,d)$ para expressar esta igualdade. O a é a **parte real** e o b é a **parte imaginária** do número complexo (a,b). Definimos as operações de **adição** e de **multiplicação** de dois números complexos por

$$(a,b) + (c,d) = (a+c, b+d)$$

e

$$(a,b) \times (c,d) = (ac - bd, ad + bc).$$

O sinal de multiplicação pode ser omitido e tanto $(a,b)\times\,(c,d)$ quanto $(a,b)\ (c,d)$ possuem o mesmo significado. O número complexo $(0,1)$ é denotado por i e denominado **unidade imaginária**. Se denotarmos o número complexo $(a,0)$ simplesmente por a e vemos que todo número complexo (a,b) pode ser escrito na forma $a+bi$. De fato,

$$(a,b) = (a,0) + (b,0)(0,1) = a + bi$$

e, a partir daí, obtemos

$$(a+bi)+(c+di)=(a+c)+(b+d)i$$

e

$$(a+bi)\times(c+di)=(ac-bd)+(ad+bc)i.$$

O sinal de multiplicação pode ser omitido e tanto $(a+bi)\times\ (c+di)$ quanto $(a+bi)\ (c+di)$ possuem o mesmo significado. Com a notação introduzida que identifica o par (a,b) com $a+bi$, os números complexos $a+bi$ e $c+di$ são iguais se $a=c$ e $b=d$ e se escreve $a+bi=c+di$ para expressar esta igualdade. O número complexo z_1+z_2 é a **soma** de z_1 e z_2. O número complexo $z_1\times z_2$ é o **produto** de z_1 e z_2.

O conjunto de todos os números complexos com as operações de adição e multiplicação é um corpo, denotado por $\mathbf{C}$ e denominado corpo dos números complexos. O elemento neutro da adição é o $0=0+\ 0i$ e o elemento neutro da multiplicação é o $1=1+\ 0i$. O 0 é denominado **zero** e 1 é denominado de **unidade**. Se $a+\ bi$ for um número complexo, então seu **inverso aditivo** ou seu **oposto**, é $-a+(\ -b)i$ e seu **inverso multiplicativo** ou seu **inverso**, é

$$(a+bi)\left(\frac{a}{a^2+b^2}-\frac{b}{a^2+b^2}i\right).$$

que existe apenas quando $a+bi$ for diferente de zero. O oposto de z é denotado por $-z$ e o inverso de z é denotado por z^{-1}.

Definimos a **subtração** $z_1-\ z_2$ de dois números complexos z_1 e z_2 por

$$z_1-z_2=z_1+(-z_2)$$

e a **divisão** z_1/z_2, onde $z_2\neq 0$, por

$$\frac{z_1}{z_2}=z_1z_2^{-1}.$$

Sendo $z=a+bi$ um número complexo, onde a e b são reais, $\overline{z}=a-\ bi$ é o seu **complexo conjugado** e o número real $|z|=\sqrt{z\overline{z}}=\sqrt{a^2+b^2}$ é o seu **módulo**. Neste caso, sendo $z\neq 0$, então

$$z^{-1}=\frac{\overline{z}}{z\ \overline{z}}$$

Exemplo 4.4 *Se* $z=3+4i$, *então* $\overline{z}=3-4i$ *e* $z\overline{z}=(3+4i)(3-4i)=25$. *Logo,* $z^{-1}=(3-4i)/25$.

Se z e w são números complexos, valem as propriedades

$$\begin{aligned}\overline{z+w}&=\bar{z}+\bar{w}\\ \overline{zw}&=\bar{z}\bar{w}\\ \overline{z^{-1}}&=\bar{z}^{-1}.\end{aligned}$$

Vamos tentar definir um produto interno em $\mathbf{C}^n$ que mantenha as propriedades do produto interno do $\mathbf{R}^n$. Se $x = (x_1, \ldots, x_n)$ e $y = (y_1, \ldots, y_n)$ forem dois elementos de $\mathbf{C}^n$, então, uma primeira tentativa seria

$$\langle x, y \rangle = x_1 y_1 + \cdots + x_n y_n.$$

Entretanto, com esta definição a primeira propriedade $\langle x, x \rangle \geq 0$ falha pois $\langle x, x \rangle$ nem sempre é real. Uma correção possível consiste em definir

$$\langle x, y \rangle = \overline{x}_1 y_1 + \cdots + \overline{x}_n y_n$$

que agora satisfaz às propriedades 1 e 3 do produto interno em espaços vetoriais sobre os reais. Entretanto, a propriedade 2 não é satisfeita. Em seu lugar vale

$$\langle x, y \rangle = \overline{\langle y, x \rangle}.$$

Aceitamos esta propriedade como uma consequência inevitável e com isto em mente definimos o produto interno em espaços vetoriais sobre o corpo dos números complexos.

Seja V um espaço vetorial sobre o corpo $\mathbf{C}$ dos números complexos. Um **produto interno** em V é uma operação

$$\langle\ ,\ \rangle : V \times V \to \mathbf{C}$$

com as propriedades abaixo, válidas para todo v, w e z em V e todo a e b em $\mathbf{C}$:

1. O produto interno é positivo definido
$$\langle v, v \rangle \geq 0 \quad \text{e} \quad \langle v, v \rangle = 0 \text{ se e só se } v = 0.$$

2. O produto interno é hermitiano
$$\langle v, w \rangle = \overline{\langle w, v \rangle}.$$

3. O produto interno é linear na segunda variável
$$\langle v\ ,\ aw + bz\ \rangle = a \langle v, w \rangle + b \langle v, z \rangle.$$

A partir das propriedades 2 e 3, se conclui que

$$\langle\ av + bw\ ,\ z \rangle = \bar{a} \langle v, z \rangle + \bar{b} \langle w, z \rangle.$$

Se v, $v_1, \ldots, v_p$ e w. $w_1, \ldots, w_q$ forem vetores de V, se $a_1, \ldots, a_p$ e $b_1, \ldots, b_q$ forem números complexos, prova-se por indução que

$$\begin{aligned} \left\langle v, \sum_j b_j w_j \right\rangle &= \sum_j b_j \langle v, w_j \rangle \\ \left\langle \sum_i a_i v_i, w \right\rangle &= \sum_i \bar{a}_i \langle v_i, w \rangle \end{aligned}$$

e, juntando as duas proprieades,

$$\left\langle \sum_i a_i v_i, \sum_j b_j w_j \right\rangle = \sum_i \sum_j \bar{a}_i b_j \langle v_i, w_j \rangle .$$

Exemplo 4.5 *Se $x = [x_1, \ldots, x_n]^T$ e $y = [y_1, \ldots, y_n]^T$ forem matrizes coluna em $\mathbf{C}^{n\times 1}$, definimos $x^* = [\bar{x}_1, \ldots, \bar{x}_n]$. A operação $\langle x, y\rangle = x^* y$, que leva duas matrizes coluna em $\mathbf{C}^{n\times 1}$ em uma matriz complexa 1×1 é um produto interno em $\mathbf{C}^n$. Aqui identificamos $[a]$, uma matriz 1×1, com o número complexo a.*

Exemplo 4.6 *Seja $V = \{f : [a, b] \to \mathbf{C} : f$ é contínua$\}$. Este conjunto com as operações de adição de funções de V e multiplicação de um número complexo por uma função de V é um espaço vetorial sobre o corpo dos números complexos. Um produto interno neste espaço vetorial é dado por*

$$\langle f, g\rangle = \int_{-L}^{L} \overline{f(t)} g(t)\ dt.$$

4.3 Funcional linear

Seja V um espaço vetorial sobre um corpo $\mathbf{C}$ dos números complexos. Uma transformação linear $f : V \to \mathbf{C}$ recebe o nome de **funcional linear** em V.

Teorema 4.7 *Seja V um espaço vetorial sobre $\mathbf{C}$, com dimensão finita e produto interno. Dado um funcional linear $f : V \to \mathbf{C}$. Então existe um vetor w em V tal que $f(v) = \langle w, v\rangle$ para todo v em V.*

Prova. Seja $\{v_1, \ldots, v_n\}$ uma base ortonormal de V. Decompondo v nesta base e calculando $f(v)$ obtemos

$$f(v) = f(a_1 v_1 + \cdots + a_n v_n) = a_1 f(v_1) + \cdots + a_n f(v_n) = \langle w, v\rangle$$

onde $w = \overline{f(v_1)} v_1 + \cdots + \overline{f(v_n)} v_n$.

Vamos mostrar que este vetor é único. Se houvesse outro vetor u tal que $\langle v, w\rangle = \langle v, u\rangle$ para todo v em V, então $\langle v, w - u\rangle = 0$ para todo v. Tomando $v = w - u$, segue $\langle w - u, w - u\rangle = 0$, mostrando que $w = u$. $\square$

O vetor w tal que $f(v) = \langle w, v\rangle$ para todo v em V pertence ao complemento ortogonal do $\ker(f)$ uma vez que $f(v) = 0$ implica em $\langle w, v\rangle = 0$.

Toda transformação linear L de um espaço vetorial complexo V de dimensão n em $\mathbf{C}^m$ é da forma

$$L(v) = (\ f_1(v), \ldots, f_m(v)\ ),$$

onde f_i é um funcional linear em V. Dado um produto interno em V, existem $w_1, \ldots, w_m$ em V tais que

$$L(v) = (\ \langle w_1, v\rangle, \ldots, \langle w_m, v\rangle\ ),$$

4.4 Norma

Seja V um espaço vetorial real ou complexo e v um vetor de V. A **norma** de v é definida por

$$\|v\| = \sqrt{\langle v, v\rangle}.$$

Se $\|v\| = 1$, diz-se que o vetor é **unitário**.

Para todo v e w em V e todo escalar a, as igualdades abaixo se verificam.

1. $\|v\| \geq 0$ e $\|v\| = 0$ se e só se $v = 0$.

2. $\|av\| = |a| \, \|v\|$.

3. $\|v + w\| \leq \|v\| + \|w\|$ (Desigualdade triangular)

Para provar esta última desigualdade, devemos provar a desigualdade de Cauchy-Schwarz. Lembramos que, se a e b são reais, a é a parte real e b a parte imaginária do número complexo $a + bi$. As notações $\mathrm{Re}\,(a + bi)$ e $\mathrm{Im}\,(a + bi)$ são usadas para designar as partes real e imaginária do número complexo $a + bi$. Observe que, se z for um número complexo, então $2\mathrm{Re}\,(z) = z + \overline{z}$ e $2\mathrm{Im}\,(z) = z - \overline{z}$.

As desigualdades abaixo são úteis. Sendo z um número complexo,

$$\begin{aligned} \mathrm{Re}\,(z) &\leq |\mathrm{Re}\,(z)| \leq |z| \\ \mathrm{Im}\,(z) &\leq |\mathrm{Im}\,(z)| \leq |z| \,. \end{aligned}$$

Teorema 4.8 *Seja V um espaço vetorial sobre o corpo $\mathbf{C}$ dos números complexos onde se definiu um produto interno. Sejam v e w dois vetores em V. Vale a desigualdade de Cauchy-Schwarz*

$$|\langle v, w\rangle| \leq \|v\| \, \|w\| \,.$$

Prova. Seja λ um número real qualquer. Então

$$\begin{aligned} 0 &\leq \langle v + \lambda w, v + \lambda w\rangle = \langle v, v\rangle + \lambda \langle v, w\rangle + \lambda \langle w, v\rangle + \lambda^2 \langle w, w\rangle \\ &= \|v\|^2 + \lambda\overline{\langle v, w\rangle} + \lambda \langle v, w\rangle + \lambda^2 \|w\|^2 \\ &= \|v\|^2 + 2\lambda \mathrm{Re}\, \langle v, w\rangle + \lambda^2 \|w\|^2 \\ &\leq \|v\|^2 + 2\lambda \, |\langle v, w\rangle| + \lambda^2 \|w\|^2 \end{aligned}$$

Como este polinômio real é maior ou igual a zero para todo λ, seu discriminante $\Delta = 4\,|\langle v, w\rangle|^2 - 4\,\|v\|^2 \, \|w\|^2$ é menor ou igual a zero, ou seja,

$$|\langle v, w\rangle|^2 \leq \|v\|^2 \, \|w\|^2 \,.$$

□

Esta desigualdade implica em

$$\frac{|\langle v, w\rangle|}{\|v\| \, \|w\|} \leq 1$$

para todo par de vetores v e w não nulos.

Seja V um espaço vetorial sobre o corpo dos números reais. A desigualdade de Cauchy-Schwarz implica em

$$-1 \leq \frac{\langle v, w\rangle}{\|v\| \, \|w\|} \leq 1$$

o que motiva a definição de **ângulo** entre v e w, como sendo aquele único número real θ, pertencente ao intervalo $[0, \pi]$, para o qual

$$\cos\theta = \frac{\langle v, w\rangle}{\|v\| \, \|w\|}.$$

Os vetores v e w são **ortogonais** quando $\theta = \pi/2$ ou $\langle v, w\rangle = 0$, fato que será indicado pelo símbolo $v \perp w$.

Quando estivermos em um espaço vetorial sobre o corpo dos números complexos, não tem sentido definir ângulo entre vetores pois, neste caso, $\langle v, w\rangle$ pode ser um número complexo. Entretanto diremos que dois vetores v e w em tal espaço são **ortogonais** quando $\langle v, w\rangle = 0$ e usaremos o símbolo $v \perp w$ para designar este fato.

Se w for ortogonal a si mesmo, $\langle w, w\rangle = 0$ e isto implica em $w = 0$. Se w for ortogonal a todo elemento de V, então é ortogonal a si mesmo e $w = 0$. Se a dimensão de V for finita e w for ortogonal a uma base de V, então $w = 0$.

Um conjunto de vetores $\{v_1, \ldots, v_p\}$ é **ortogonal** quando seus elementos forem dois a dois ortogonais entre si. Se além disto, todos os vetores possuírem norma unitária, o conjunto é **ortonormal**.

Teorema 4.9 *Seja V um espaço vetorial de dimensão n e $S = \{v_1, \ldots, v_n\}$ um conjunto ortogonal de vetores de V. Então S é uma base.*

Prova. Basta provar que S é linearmente independente. De fato, sejam $k_1, \ldots, k_n$ escalares tais que $k_1v_1 + \cdots + k_nv_n = 0$. Como $\langle v_i, 0\rangle = 0$, obtemos

$$0 = \langle v_i, 0\rangle = \langle v_i, k_1v_1 + \cdots + k_nv_n\rangle = k_i \langle v_i, v_i\rangle = k_i \|v_i\|^2.$$

Como $\|v_i\| \neq 0$, segue que $k_i = 0$, provando a independência linear de S. □

Seja $\{v_1, \ldots, v_n\}$ uma base ortonormal de um espaço vetorial V. Dado $v = x_1v_1 + \cdots + x_nv_n$ neste espaço, multiplicando-o internamente por v_i, obtemos $x_i = \langle v_i, v\rangle$ e assim,

$$v = \langle v_1, v\rangle v_1 + \cdots + \langle v_p, v\rangle v_p.$$

O próximo teorema apresenta a forma da matriz de mudança de bases quando as bases envolvidas são ortonormais.

Seja $A = [a_{ij}]$ uma matriz complexa $m \times n$. A matriz $A^* = [b_{ij}]$ onde $b_{ij} = \bar{a}_{ij}$ é a matriz adjunta de A. Se $A^* = A$, a matriz A é denominada **hermitiana**. Se $A^{-1} = A^*$ a matriz A é denominada **unitária**.

Teorema 4.10 *Sejam $\mathcal{B}_1$ e $\mathcal{B}_2$ bases ortonormais de um espaço vetorial V sobre o corpo $\mathbf{C}$ dos números complexos. A matriz M_{12} de mudança da base $\mathcal{B}_1$ para a base $\mathcal{B}_2$ é **hermitiana** e **unitária**.*

Prova. Sejam $\mathcal{B}_1 = \{v_1, \ldots, v_n\}$ e $\mathcal{B}_2 = \{w_1, \ldots, w_n\}$ as bases em questão, $M_{12} = [a_{ij}]$ e $M_{21} = [b_{ij}]$ as matrizes de mudança da base $\mathcal{B}_1$ para a base $\mathcal{B}_2$ e da base $\mathcal{B}_2$ para a base $\mathcal{B}_1$, respectivamente. Sabemos que $M_{12}^{-1} = M_{21}$ e que

$$w_j = \sum_i a_{ij} v_i \quad \text{e} \quad v_j = \sum_i b_{ij} w_i.$$

Sendo as bases ortonormais, $a_{ij} = \langle v_i, w_j \rangle = \bar{a}_{ji}$ e $b_{ij} = \langle w_i, v_j \rangle = \overline{\langle v_j, w_i \rangle} = \bar{a}_{ji}$, mostrando que $M_{12}^* = M_{12}$ e $M_{12}^{-1} = M_{12}^*$. $\square$

Exemplo 4.11 *Consideremos as bases $\mathcal{B}_1 = \{e_1,\ e_2\}$ e $\mathcal{B}_2 = \{f_1,\ f_2\}$ do $\mathbf{R}^2$, onde $e_1 = (1, 0)$, $e_2 = (0,\ 1)$, $f_1 = (1/5)(3,\ 4)$ e $f_2 = (1/5)(4,\ -3)$. Ambas são ortonormais em relação ao produto interno $\langle (x_1, y_1), (x_2, y_2) \rangle = x_1 x_2 + y_1 y_2$. Temos*

$$\begin{aligned} f_1 &= \frac{3}{5} e_1 + \frac{4}{5} e_2 \\ f_1 &= \frac{4}{5} e_1 - \frac{3}{5} e_2 \end{aligned}$$

e

$$\begin{aligned} e_1 &= \frac{3}{5} f_1 + \frac{4}{5} f_2 \\ e_1 &= \frac{4}{5} f_1 - \frac{3}{5} f_2. \end{aligned}$$

Assim

$$M_{12} = M_{21} = \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \\ \frac{4}{5} & -\frac{3}{5} \end{bmatrix}.$$

A matriz M_{12} é hermitiana e $M_{12} M_{12}^*$ é a matriz identidade, mostrando que M_{12} é unitária.

4.5 Ortogonalização de Gram-Schmidt

Podemos, a partir de uma base $\{v_1, \ldots, v_n\}$ uma base de um espaço vetorial V, obter uma base ortogonal $\{w_1, \ldots, w_n\}$ de V, seguindo o procedimento descrito em seguida e conhecido como processo de ortogonalização de Gram-Schmidt.

Defina

$$w_1 = v_1.$$

Agora escreva

$$w_2 = v_2 - \beta_{12} w_1$$

e determine o escalar β_{12} para que a condição de ortogonalidade $\langle w_1, w_2\rangle = 0$ seja satisfeita. Substitua w_2 por $v_2 - \beta_{12} w_1$ nesta condição para obter

$$\beta_{12} = \frac{\langle w_1, v_2\rangle}{\langle w_1, w_1\rangle}.$$

Em seguida, considere

$$w_3 = v_3 - \beta_{13} w_1 - \beta_{23} w_3$$

e determine β_{13} e β_{23} para tornar w_3 ortogonal a w_1 e w_2. Das condições de ortogonalidade $\langle w_1, w_3\rangle = 0$ e $\langle w_2, w_3\rangle = 0$ calcule

$$\beta_{13} = \frac{\langle w_1, v_3\rangle}{\langle w_1, w_1\rangle} \quad \text{e} \quad \beta_{23} = \frac{\langle w_2, v_3\rangle}{\langle w_2, w_2\rangle}.$$

Prosseguindo com este raciocínio, se chega a um conjunto ortogonal $\{w_1, \ldots, w_n\}$ de vetores que é base de V. Observe que os vetores $w_1, w_2, \ldots, w_n$ são definidos recursivamente por

$$w_1 = v_1$$

e

$$w_k = v_k - \beta_{1k} w_1 - \cdots - \beta_{k-1,k} w_{k-1}$$

para $k = 2, \ldots, n$, onde

$$\beta_{ik} = \frac{\langle w_i, v_k\rangle}{\langle w_i, w_i\rangle}.$$

A partir da base ortogonal $\{w_1, \ldots, w_n\}$ pode-se determinar uma base ortonormal $\{q_1, \ldots, q_n\}$, onde $q_i = w_i / \|w_i\|$. Esta base ortonormal pode ser obtida ao mesmo tempo em que se obtém a base ortogonal. Comece com

$$w_1 = v_1 \quad \text{e} \quad q_1 = w_1 / \|w_1\|$$

e continue com o processo de ortogonalização, tomando

$$w_2 = v_2 - r_{12} q_1 \quad \text{e} \quad q_2 = w_2 / \|w_2\|,$$

$$w_3 = v_3 - r_{13} q_1 - r_{23} q_2 \quad \text{e} \quad q_3 = w_3 / \|w_3\|,$$

e assim por diante, até que, num passo genérico k,

$$w_k = v_k - r_{1k} q_1 - \cdots - r_{k-1,k} q_{k-1} \quad \text{e} \quad q_k = w_k / \|w_k\|,$$

onde

$$r_{ik} = \langle q_i, v_k\rangle,$$

para $k = 2, \ldots, n$ e $i = 1, \ldots, k-1$.

Exemplo 4.12 *Os ternos ordenados* (1, 0, 0), (0, 3/5, 4/5), (0, 4/5, −3/5) *formam uma base ortonormal no espaço vetorial* $\mathbf{C}^n$ *em relação ao produto interno*

$$\langle\ (a_1, a_2, a_3),\ \ (b_1, b_2, b_3)\ \rangle = \bar{a}_1 b_1 + \bar{a}_2 b_2 + \bar{a}_3 b_3.$$

Exemplo 4.13 *Os polinômios* $1,\ x,\ x^2$ *formam uma base ortonormal no espaço vetorial sobre* $\mathbf{C}$ *dos polinômios de grau menor ou igual a* 2 *e coeficientes complexos, munido com o produto interno*

$$\left\langle a_1 + a_2 x + a_3 x^2, b_1 + b_2 x + b_3 x^2 \right\rangle = \bar{a}_1 b_1 + \bar{a}_2 b_2 + \bar{a}_3 b_3.$$

Exemplo 4.14 *Considere o espaço vetorial sobre* $\mathbf{C}$ *dos polinômios com coeficientes complexos de grau menor ou igual a* 3, *com o produto interno*

$$\langle f, g\rangle = \int_{-1}^{1} f(x) g(x) dx.$$

O conjunto $\{1,\ x,\ x^2,\ x^3\}$ *é uma base não ortogonal deste espaço vetorial. A base ortogonal obtida a partir dela, usando o procedimento de Gram-Schmidt, é*

$$\{\ 1,\ \ x,\ \ x^2 - 1/3,\ \ x^3 - (3/5)x\ \}.$$

Este procedimento pode ser estendido para o espaço vetorial dos polinômios de grau menor ou igual a n.

Denotemos por $p_0(x),\ p_1(x),\ p_2(x),\ \ldots$ *os polinômios obtidos de* $1,\ x,\ x^2,\ \ldots$ *pelo processo de ortogonalização de Gram-Schmidt, usando o produto interno definido acima. Os polinômios* $L_k(x) = p_k(x)/p_k(1)$ *continuam ortogonais dois a dois e são denominados de* ***polinômios de Legendre****. Os quatro primeiros são*

$$L_0(x) = 1, \quad L_1(x) = x, \quad L_2(x) = \frac{3}{2}x^2 - \frac{1}{2}, \quad L_3(x) = \frac{5}{2}x^3 - \frac{3}{2}x.$$

Tanto $\{1, x, x^2, x^3\}$ *quanto* $\{L_0(x), L_1(x), L_2(x), L_3(x)\}$ *são bases do espaço dos polinômios de grau menor ou igual a* 3. *A segunda possui a vantagem de ser ortogonal, o que a torna mais adequada para determinados cálculos. Os métodos espectrais usam polinômios ortogonais para resolver equações diferenciais parciais tanto analítica quanto numericamente.*

4.6 Decomposição QR

Vamos analisar o caso especial do espaço vetorial complexo $\mathbf{C}^{n\times 1}$ com o produto interno

$$\langle x, y\rangle = x^* y.$$

Seja $A = [v_1,\ \ldots,\ v_n]$ uma matriz $m \times n$ cujo coluna k é v_k. Um modo interessante de olhar para o produto Ax, onde $x = [x_1,\ \ldots,\ x_n]^T$ é uma matriz em $\mathbf{C}^{m\times 1}$ consiste em escrever

$$Ax = x_1 v_1 + \cdots + x_n v_n$$

e observar que Ax é uma combinação linear das colunas de A.

Mantendo a notação do parágrafo anterior, sendo b uma matriz coluna em $\mathbf{C}^{m\times 1}$, a igualdade matricial $Ax = b$, pode ser escrita na forma

$$b = x_1 v_1 + \cdots + x_n v_n$$

que pode ser interpretada do seguinte modo: x é a matriz de b na base formada pelas colunas de A. Se as colunas de A forem linearmente independentese b estiver na imagem de A, a decomposição é única.

Ainda uma última observação, sendo $A = [v_1, \ldots, v_n]$, então

$$A^* = \begin{bmatrix} v_1^* \\ \vdots \\ v_n^* \end{bmatrix}$$

e

$$A^*A = \begin{bmatrix} v_1^*v_1 & v_1^*v_2 & & v_1^*v_n \\ v_2^*v_1 & v_2^*v_2 & & v_2^*v_n \\ & & & \\ v_n^*v_1 & v_n^*v_2 & & v_n^*v_n \end{bmatrix} = [v_i^* v_j] \, .$$

Se $\{q_1, \ldots, q_n\}$ for uma base ortonormal em $\mathbf{C}^{n\times 1}$, então $q_i^* q_j = \delta_{ij}$. A matriz quadrada $Q = [q_1, \ldots, q_n]$, cuja coluna k é q_k, é unitária pois $Q^*Q = [q_i^* q_j] = [\delta_{ij}] = I$. Conclusão, quando as colunas de uma matriz quadrada formarem uma base ortonormal de $\mathbf{C}^{n\times 1}$, ela é unitária.

Vamos iniciar com um caso particular, em que $n = 3$. Seja $\{v_1, v_2, v_3\}$ uma base de $\mathbf{C}^{3\times 1}$ e $\{q_1, q_2, q_3\}$ a base ortonormal de $\mathbf{C}^{3\times 1}$ obtida pelo processo de ortogonalização de Gram-Schmidt. Seja $A = [v_1, v_2, v_3]$ a matriz cuja coluna k é v_k e $Q = [q_1, q_2, q_3]$ a matriz cuja coluna k é q_k. Sabemos, pelo desenvolvimento da seção anterior que

$$\begin{aligned} v_1 &= w_1 \\ v_2 &= r_{12} q_1 + w_2 \\ v_3 &= r_{13} q_1 + r_{23} q_2 + w_3 \end{aligned}$$

onde $r_{ik} = \langle q_i, v_k \rangle$ quando $i \neq k$ ou ainda,

$$\begin{aligned} v_1 &= r_{11} q_1 \\ v_2 &= r_{12} q_1 + r_{22} q_2 \\ v_3 &= r_{13} q_1 + r_{23} q_2 + r_{33} q_3 \end{aligned}$$

com $r_{kk} = \|w_k\|$. Então,

$$[v_1, v_2, v_3] = [q_1, q_2, q_3] \begin{bmatrix} r_{33} & r_{12} & r_{13} \\ 0 & r_{33} & r_{23} \\ 0 & 0 & r_{33} \end{bmatrix}$$

ou $A = QR$, onde Q é uma matriz unitária e

$$R = \begin{bmatrix} r_{33} & r_{12} & r_{13} \\ 0 & r_{33} & r_{23} \\ 0 & 0 & r_{33} \end{bmatrix}$$

é triangular superior. Esta é a chamada decomposição QR de uma matriz A,

Motivados por esse exemplo, vamos mostrar um processo para obter a decomposição QR de uma matriz A em $\mathbf{C}^{m\times n}$, analizando dois casos separadamente. No primeiro, todas as colunas de A são linearmente independentes e, no segundo caso, nem todas as colunas de A são linearmente independentes.

As colunas da matriz são linearmente independentes

Seja $A = [v_1, \ldots, v_n]$ uma matriz complexa de ordem m por n, cujas colunas $v_1, \ldots, v_n$ são vetores linearmente independentes de $\mathbf{C}^{m\times 1}$, o que exige $m \geq n$. Usando o processo de ortogonalização de Gram-Schmidt, podemos escrever obter uma matriz $Q = [q_1, \ldots, q_n]$ cujas colunas formam uma base ortonormal para o espaço gerado pelas colunas de A e onde

$$\begin{aligned} v_1 &= r_{11}q_1 \\ v_2 &= r_{12}q_1 + r_{22}q_2 \\ v_3 &= r_{13}q_1 + r_{23}q_2 + r_{33}q_3 \\ &\cdots \end{aligned}$$

Essas igualdades escritas na forma matricial resultam em

$$[v_1, v_2, v_3, \ldots, v_n] = [q_1, q_2, q_3, \ldots, q_n] \begin{bmatrix} r_{11} & r_{12} & r_{13} & \cdots \\ 0 & r_{22} & r_{23} & \cdots \\ 0 & 0 & r_{33} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}.$$

que se resume em

$$A = \hat{Q}\hat{R}$$

denominada **decomposição QR reduzida** de A.

Nesta decomposição, observe que o espaço $\langle v_1 \rangle$, gerado por v_1 é igual ao espaço $\langle q_1 \rangle$ gerado por q_1, o espaço $\langle v_1, v_2 \rangle$ gerado por v_1 e v_2 é igual ao espaço $\langle q_1, q_2 \rangle$ gerado por q_1, q_2, e assim por diante,

$$\begin{aligned} \langle q_1 \rangle &= \langle v_1 \rangle, \\ \langle q_1, q_2 \rangle &= \langle v_1, v_2 \rangle, \\ \langle q_1, q_2, q_3 \rangle &= \langle v_1, v_2, v_3 \rangle, \\ &\cdots \end{aligned}$$

Completemos a base $\{q_1, \ldots, q_n\}$ com os vetores unitários $q_{n+1}, \ldots, q_m$ de modo que $\{q_1, \ldots, q_n, \ldots, q_m\}$ seja uma base ortonormal de $\mathbf{C}^m$. A matriz $Q = [q_1, \ldots, q_n, \ldots, q_m]$ obtida pela inclusão de $m - n$ colunas à direita de $\hat{Q}$ e a matriz R obtida pela inclusão de $m - n$ linhas nulas na parte inferior de R são tais que

$$A = QR$$

que é a chamada **decomposição QR completa** de A ou **decomposição QR** de A.

Realizado este desenvolvimento, podemos descrever o algoritmo clássico de Gram-Schmidt, que possibilita a obtenção da decomposição QR de uma mariz A. Alertamos o leitor de que este algoritmo é numericamente instável.

==================================

Algoritmo 8.1. Algoritmo clássico de Gram-Schmidt (instável)
Entrada: Base $\{v_1, \ldots, v_n\}$ de $\mathbf{C}^n$
Saída: Base ortonormal $\{q_1, \ldots, q_n\}$ de $\mathbf{C}^n$

==================================

for $k = 1$ to n
    $w_k = v_k$
    for $i = 1$ to $k - 1$
        $r_{ik} = q_i^* v_k$
        $w_k = w_k - r_{ik} q_i$
    $r_{kk} = \|w_k\|$
    $q_k = w_k / r_{kk}$

==================================

Solução de $Ax = b$ usando a decomposição QR

Quando A é uma matriz quadrada de ordem m, cujas colunas são linearmente independentes, o sistema $Ax = b$ possui uma única solução. Para resolver este sistema usando a decomposição QR, procedemos do seguinte modo:

1. Calcule a decomposição $A = QR$.
2. Determine $y = Q^* b$.
3. Resolva o sistema $Rx = y$ na variável x.

As colunas da matriz são linearmente dependentes

Passemos ao caso em que $m \geq n$ e as colunas de A formam um conjunto linearmente dependente. Neste caso, lá pelas tantas, v_k depende linearmente das colunas $v_1, \ldots, v_{k-1}$, à sua esquerda, ou seja,

$$v_k \in \langle v_1, \ldots, v_{k-1} \rangle = \langle q_1, \ldots, q_{k-1} \rangle$$

e, para este valor de k,

$$w_k = v_k - r_{1k}q_1 - r_{2k}q_2 - \cdots - r_{k-1,k}q_{k-1} = 0.$$

Quando isto ocorre, escolhemos um vetor unitário q_j, ortogonal aos vetores $q_1, \ldots, q_{j-1}$, obtendo um conjunto ortonormal $\{q_1, \ldots, q_{j-1}, q_j\}$.

Vejamos um exemplo em que $A = [v_1, v_2, v_3, v_4]$. Suponha que v_1 e v_2 são linearmente independentes. Usando o método de ortogonalização de Gram-Schmidt, calculamos

$$\begin{aligned} v_1 &= r_{11}q_1 \\ v_2 &= r_{12}q_1 + r_{22}q_2 \end{aligned}$$

Supondo v_3 no espaço gerado por $\{q_1, q_2\}$, tem-se

$$w_3 = v_3 - r_{13}q_1 - r_{23}q_2 = 0$$

e

$$v_3 = r_{13}q_1 + r_{23}q_2$$

Daí, escolhe-se de modo arbitrário um q_3 unitário, ortogonal a q_1 e a q_2. Com esta escolha, $\{q_1, q_2, q_3\}$ é ortonormal e o espaço que ele gera contém o espaço gerado por $\{v_1, v_2, v_3\}$.

Se v_4 não pertencer ao espaço gerado por $\{q_1, q_2, q_3\}$, Calcula-se

$$q_4 = \frac{1}{r_{44}}(v_4 - r_{14}q_1 - r_{24}q_2 - r_{34}q_3)$$

quando então

$$[v_1, v_2, v_3, v_4] = [q_1, q_2, q_3, q_4] \begin{bmatrix} r_{11} & r_{12} & r_{13} & r_{14} \\ 0 & r_{22} & r_{23} & r_{24} \\ 0 & 0 & 0 & r_{34} \\ 0 & 0 & 0 & r_{44} \end{bmatrix}$$

onde se observa que a matriz da direita é triangular superior. Note-se que o espaço gerado por $\{q_1, q_2, q_3, q_4\}$ contém o espaço gerado por $\{v_1, v_2, v_3, v_4\}$.

No caso genérico, este procedimento continua, até obter as matrizes $\hat{Q}$ e $\hat{R}$. As colunas da matriz $\hat{Q} = [q_1, \ldots, q_n]$, de ordem m por n, são vetores ortogonais entre si e possuem módulo unitário. A matriz $\hat{R}$, de ordem n por n, é triangular superior. Para estas matrizes,

$$A = \hat{Q}\hat{R}.$$

Esta fatoração de A é conhecida como **decomposição QR reduzida** de A.

Podemos acrescentar $m-n$ colunas $q_{n+1}, \ldots, q_m$ à direita de $\hat{Q}$, de modo que $\{q_1, \ldots, q_n, \ldots, q_m\}$ seja uma base ortonormal de $\mathbf{C}^m$ e assim, obter uma matriz unitária $Q = [q_1, \ldots, q_n, \ldots, q_m]$, de ordem m por m, cujas colunas formam uma base ortonormal de $\mathbf{C}^m$. Na continuação, devemos acrescentar $m - n$ linhas nulas na parte inferior de $\hat{R}$, obtendo uma matriz R, de ordem m por n, triangular superior. As matrizes Q e R assim obtidas são de tal forma que

$$A = QR.$$

Esta é a **decomposição QR completa** de A ou apenas **decomposição QR** de A.

Quando $m < n$, o procedimento é semelhante ao anterior. A decomposição se encerra quando obtemos o conjunto de m vetores $\mathcal{B} = \{q_1, \ldots, q_m\}$, que formam uma base ortonormal de $\mathbf{C}^m$. A matriz quadrada $Q = [q_1, \ldots, q_m]$, de ordem m, e a matriz triangular superior R de ordem m por n, obtidas no desenrolar do processo são tais que

$$A = QR.$$

Este produto é conhecido como decomposição QR completa da matriz A ou decomposição QR de A.

Exemplo 4.15 *A decomposição QR de*

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

é

$$\begin{pmatrix} -1/\sqrt{2} & 3/\sqrt{22} & -1/\sqrt{11} \\ 1/\sqrt{2} & 3/\sqrt{22} & -1/\sqrt{11} \\ 0 & 2/\sqrt{22} & 3/\sqrt{11} \end{pmatrix} \begin{pmatrix} \sqrt{2} & -3/\sqrt{2} & 1/\sqrt{2} \\ 0 & 11/\sqrt{22} & 7/\sqrt{22} \\ 0 & 0 & 5/\sqrt{11} \end{pmatrix}.$$

Exemplo 4.16 *A decomposição QR de*

$$\begin{pmatrix} -1 & 3 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$$

é

$$\begin{pmatrix} -1/\sqrt{2} & 3/\sqrt{22} & 1/\sqrt{11} \\ 1/\sqrt{2} & 3/\sqrt{22} & 1/\sqrt{11} \\ 0 & \sqrt{2/11} & -3/\sqrt{11} \end{pmatrix} \begin{pmatrix} \sqrt{2} & -3/\sqrt{2} \\ 0 & \sqrt{11/2} \\ 0 & 0 \end{pmatrix}$$

Exemplo 4.17 *A decomposição QR de* $\begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 2 & 0 \end{pmatrix}$

é $\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{22} & 2/\sqrt{33} & 3/\sqrt{3} \\ 0 & 2/\sqrt{22} & 4/\sqrt{33} & -3/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{22} & -2/\sqrt{33} & -3/\sqrt{3} \\ 0 & 4/\sqrt{22} & -3/\sqrt{33} & 0 \end{pmatrix} \begin{pmatrix} \sqrt{2} & 1/\sqrt{2} & 3/\sqrt{2} \\ 0 & 1/\sqrt{22} & 3/\sqrt{22} \\ 0 & 0 & 6/\sqrt{33} \\ 0 & 0 & 0 \end{pmatrix}$

Exemplo 4.18 *A decomposição QR de*

$$\begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

$$
\text{é} \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} & 0 \\ 0 & 2/\sqrt{6} & -1\sqrt{3} & 0 \\ 1/\sqrt{2} & -1/\sqrt{6} & -1\sqrt{3} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{2} & 1/\sqrt{2} & 3/\sqrt{2} \\ 0 & 3/\sqrt{6} & 3/\sqrt{6} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}
$$

Exemplo 4.19 *A decomposição QR de*

$$
\begin{pmatrix} 1 & 1 & 2 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 3 \end{pmatrix}
$$

é $A = QR$, onde

$$
Q = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & -1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & -1/\sqrt{3} \end{pmatrix} \quad e \quad R = \begin{pmatrix} \sqrt{2} & 1/\sqrt{2} & 3/\sqrt{2} & 3/\sqrt{2} \\ 0 & 3/\sqrt{6} & 3/\sqrt{6} & -1/\sqrt{6} \\ 0 & 0 & 0 & -4/\sqrt{3} \end{pmatrix}.
$$

Capítulo 5

Soma de subespaços

Sejam $V_1, \ldots, V_k$ subespaços vetoriais de V. O conjunto

$$V_1 + \cdots + V_k = \{\ v_1 + \cdots + v_k : v_i \in V_i \text{ para } i = 1, \ldots, k\ \}$$

é um subespaço vetorial de V e recebe o nome de **soma** de $V_1, \ldots, V_k$.

Teorema 5.1 *Sejam V e W dois subespaços de um espaço vetorial U. Então*

$$\dim(V + W) = \dim(V) + \dim(W) - \dim(V \cap W).$$

Prova. Quando V está contido em W, então $V \cap W = V$ e $V + W = W$. Neste caso,

$$\dim(V) + \dim(W) - \dim(V \cap W) = \dim(V) + \dim(W) - \dim(V) = \dim(W)$$

o que prova o teorema para este caso particular. Do mesmo modo se prova que o teorema vale quando W está contido em V.

Vamos agora tratar o caso em que $V \cap W$ é diferente de V e de W. Seja $\mathcal{B}_1 = \{u_1, \ldots, u_p\}$ uma base de $V \cap W$. Vamos completá-la de modo que $\mathcal{B}_2 = \{u_1, \ldots, u_p, v_1, \ldots, v_q\}$ seja base de V e $\mathcal{B}_3 = \{u_1, \ldots, u_p, w_1, \ldots, w_r\}$ seja base de W. O conjunto $\mathcal{B}_4 = \{u_1, \ldots, u_p, v_1, \ldots, v_q, w_1, \ldots, w_r\}$ gera $V + W$ e, se for linearmente independente, será base de $V + W$. Neste caso,

$$\begin{aligned} \dim(V + W) &= p + q + r = (q + p) + (r + p) - p \\ &= \dim(V) + \dim(W) - \dim(V \cap W) \end{aligned}$$

e o teorema estará provado.

Falta provar que $\mathcal{B}_4$ é linearmente independente. Vamos mostrar que, se $x_1, \ldots, x_p, y_1, \ldots, y_q, z_1, \ldots, z_r$ forem escalares tais que

$$x_1 u_1 + \cdots + x_p u_p + y_1 v_1 + \cdots + y_q v_q + z_1 w_1 + \cdots + z_r w_r = 0,$$

então todos eles são nulos. Analisemos esta possibilidade. Se algum y_j for diferente de zero, o vetor não nulo $y_1 v_1 + \cdots + y_q v_q$ seria uma combinação linear dos elementos de

$\mathcal{B}_3$. Logo, ele estaria em W e em V ao mesmo tempo, estando na interseção $V \cap W$ e assim $y_1 v_1 + \cdots + y_q v_q$ poderia ser escrito como uma combinação linear de $u_1, \ldots, u_p$, contrariando a hipótese de $\mathcal{B}_2$ ser base. Do mesmo modo não podemos ter um z_k diferente de zero. Logo, y_j e z_k são todos nulos e a equação se reduz a $x_1 u_1 + \cdots + x_p u_p = 0$. Sendo $\mathcal{B}_1$ uma base, concluímos que $x_1, \ldots, x_p$ são todos nulos. Daí $\mathcal{B}_4$ é linearmente independente. $\square$

5.1 Soma direta

Definição 5.2 *Sejam $V_1, \ldots, V_k$ subespaços vetoriais de V. Se todo v em V puder ser escrito de forma única como uma soma do tipo*

$$v = v_1 + \cdots + v_k$$

*onde $v_i \in V_i$, diremos que V é uma **soma direta** dos subespaços $V_1, \ldots, V_k$ e escreveremos*

$$V = V_1 \oplus \cdots \oplus V_k.$$

*Se $V = V_1 \oplus V_2$, então V_1 e V_2 são denominados **complementares**.*

Dois subespaços vetoriais V_1 e V_2 de V são **disjuntos** se a interseção $V_1 \cap V_2$ contiver apenas o zero.

Teorema 5.3 *Sejam V_1 e V_2 subespaços vetorias de V tais que $V = V_1 + V_2$. Então $V = V_1 \oplus V_2$ se e só se V_1, V_2 forem disjuntos.*

Prova. Se $V = V_1 \oplus V_2$, seja v um vetor de V na interseção de V_1 e V_2. Então

$$v = \underbrace{v}_{\in V_1} + \underbrace{0}_{\in V_2} = \underbrace{0}_{\in V_1} + \underbrace{v}_{\in V_2}$$

e, como a decomposição é única, $v = 0$, provando que V_1 e V_2 são disjuntos.

Se V_1 e V_2 forem disjuntos, como $V = V_1 + V_2$, todo v em V pode ser decomposto numa soma $v = v_1 + v_2$, com v_1 em V_1 e v_2 em V_2. Se houvesse outra decomposição $v = w_1 + w_k$, com w_1 em V_1 e w_2 em V_2, então $v_1 + v_2 = w_1 + w_2$ e assim, $v_1 - w_1 = w_2 - v_2$. Sendo $v_1 - w_1$ um vetor de V_1 igual a $w_2 - v_2$, um vetor de V_2, então $v_1 - w_1$ está na interseção de V_1 com V_2 e, como estes dois subespaços são disjuntos, $v_1 - w_1 = 0$ ou $v_1 = w_1$. Com este resultado, obtemos $w_2 - v_2 = 0$ ou $v_2 = w_2$, provando que a decomposição de v numa soma de um elemento de V_1 com um elemento de V_2 é única e assim, $V = V_1 \oplus V_2$. $\square$

Quando V igual à soma de mais do que dois subespaços, o fato de os espaços envolvidos serem disjuntos dois a dois não é suficiente para garantir que V seja a soma direta desses subespaços como nos mostra o exemplo a seguir.

Exemplo 5.4 *Seja* $V = \mathbf{R}^2$ *e* $V_1 = \{(x,\, 0) \in \mathbf{R}^2 : x \in \mathbf{R}\}$, $V_2 = \{(0,\, y) \in \mathbf{R}^2 : y \in \mathbf{R}\}$, $V_3 = \{(x,\, x) \in \mathbf{R}^2 : x \in \mathbf{R}\}$. *Estes três subespaços são disjuntos dois a dois,* $V = V_1 + V_2 + V_3$, *mas* V *não é a soma direta de* V_1, V_2 *e* V_3.

A condição de serem disjuntos será substituida pela condição de serem independentes. Os subespaços vetoriais $V_1,\, \ldots,\, V_k$ de V são **independentes** se

$$v_1 + \cdots + v_k = 0,$$

com v_i em V_i, para $i = 1,\, \ldots,\, k$, então $v_1 = \cdots = v_k = 0$.

Uma caracterização da independência dos subespaços é a seguinte: Os subespaços V_1, $\ldots$, V_k são independentes se e só se V_j for disjunto da soma $V_1 + \cdots + V_{j-1}$, para $j = 2$, $\ldots$, k. Dois espaços vetoriais V_1 e V_2 de V são independentes se e só se forem disjuntos.

Teorema 5.5 *Sejam* $V_1,\, \ldots,\, V_k$ *subespaços vetoriais de um espaço vetorial* V *tais que* $V = V_1 + \cdots + V_k$. *Então* $V = V_1 \oplus \cdots \oplus V_k$ *se e só se* $V_1,\, \ldots,\, V_k$ *forem independentes.*

Prova. Se $V = V_1 \oplus \cdots \oplus V_k$, sejam $v_1,\, \ldots,\, v_k$ vetores de $V_1,\, \ldots,\, V_k$, respectivamente, e tais que $v_1 + \cdots + v_k = 0$. Como $0 = 0 + \cdots + 0$, da unicidade da decomposição, concluímos que $v_i = 0$ para $i = 1,\, \ldots,\, k$. Logo, $V_1,\, \ldots,\, V_k$ são independentes.

Se $V = V_1 + \cdots + V_k$ e $V_1,\, \ldots,\, V_k$ forem independentes, todo v em V pode ser decomposto numa soma $v = v_1 + \cdots + v_k$, com v_i em V_i, para $i = 1,\, \ldots,\, k$. Se houvesse outra decomposição $v = w_1 + \cdots + w_k$, com w_i em V_i, então $(v_1 - w_1) + \cdots + (v_k - w_k) = 0$ e, da independência dos subespaços vetoriais V_i, concluímos que $v_i = w_i$, para $i = 1$, $\ldots$, k. Logo, a decomposição de v como soma de vetores de $V_1,\, \ldots,\, V_k$ é única e $V = V_1 \oplus \cdots \oplus V_k$. $\square$

Teorema 5.6 *Sejam* $V_1,\, \ldots,\, V_k$ *subespaços vetoriais de* V, *um espaço vetorial com dimensão finita. Se* $V = V_1 \oplus \cdots \oplus V_k$ *então*

$$\dim V = \dim V_1 + \cdots + \dim V_k.$$

Prova. Seja $\mathcal{B}_i$ base de V_i para $i = 1,\, \ldots,\, k$.

Se $V = V_1 \oplus \cdots \oplus V_k$, todo v em V pode ser decomposto de forma única numa soma $v = v_1 + \cdots + v_k$, com v_i em V_i, para $i = 1,\, \ldots,\, k$. Cada v_i pode ser decomposto de forma única nos vetores da base $\mathcal{B}_i$. Logo, v pode ser escrito de forma única como uma combinação linear dos vetores da união $\mathcal{B}_1 \cup \cdots \cup \mathcal{B}_k$, provando que esta é uma base de V. $\square$

5.2 Complemento ortogonal

Definição 5.7 *Seja V um espaço vetorial com produto interno e S um subespaço vetorial de V. O conjunto*

$$S^{\perp} = \{v \in V : \langle v, s\rangle = 0 \ \ \textit{para todo} \ \ s \ \textit{em} \ S \ \}$$

*é um subespaço de V, chamado de **complemento ortogonal** de S.*

Para mostrar que um determinado vetor v está em $S^{\perp}$, basta mostrar que ele é ortogonal a todo vetor de uma base de S. O único vetor que está ao mesmo tempo em S e em $S^{\perp}$ é o vetor nulo e daí,

$$S \cap S^{\perp} = \{0\}.$$

Teorema 5.8 *Seja S um subespaço vetorial de dimensão finita de um espaço vetorial V com produto interno. Então $V = S \oplus S^{\perp}$.*

Prova. Seja $\{v_1, \ldots, v_p\}$ uma base ortonormal de S. Dado qualquer v em V, o vetor

$$w = v - \langle v_1, v\rangle\, v_1 - \cdots - \langle v_p, v\rangle\, v_p$$

é ortogonal a todo vetor de S. Desta forma, qualquer vetor v pode ser decomposto numa soma $v = s + w$, onde $s = \langle v_1, v\rangle\, v_1 + \cdots + \langle v_p, v\rangle\, v_p$ pertence a S e w pertence a $S^{\perp}$ e assim, $V = S + S^{\perp}$.

Vamos mostrar que esta decomposição é única. Se $v = s_1 + w_1$, com s_1 em S e w_1 em $S^{\perp}$, então $s + w = s_1 + w_1$ e assim, $s - s_1 = w_1 - w$, mostrando que os vetores $s - s_1$ e $w_1 - w$ estão na interseção $S \cap S^{\perp}$. Como a interseção só possui o vetor nulo, $s = s$ e $w = w$. $\square$

Se v é um vetor de um subespaço S de um espaço vetorial V, então v é ortogonal a todo vetor de $S^{\perp}$ e assim ele pertence ao $\left(S^{\perp}\right)^{\perp}$, mostrando que S está contido no complemento ortogonal do complemento ortogonal de S, isto é, $S \subset \left(S^{\perp}\right)^{\perp}$. Por outro lado, $V = S \oplus S^{\perp} = S^{\perp} \oplus \left(S^{\perp}\right)^{\perp}$ e assim, $\dim V = \dim S + \dim S^{\perp} = \dim S^{\perp} + \dim \left(S^{\perp}\right)^{\perp}$, que acarreta na igualdade $\dim S = \dim \left(S^{\perp}\right)^{\perp}$. Estando S contido em $\left(S^{\perp}\right)^{\perp}$ e possuindo ambos a mesma dimensão, eles são iguais

$$\left(S^{\perp}\right)^{\perp} = S.$$

Capítulo 6

Transformação adjunta

Sejam V e W espaços vetoriais complexos com dimensão finita e produto interno. Dado uma tansformação linear $L : V \to W$, vamos mostrar que existe uma única transformação linear $L^* : W \to V$ tal que

$$\langle Lv, w \rangle = \langle v, L^* w \rangle .$$

para todo v em V e w em W.

Primeiro a existência. Sendo $\mathcal{B}_1 = \{v_1, \ldots, v_n\}$ uma base ortonormal de V e $\mathcal{B}_2 = \{w_1, \ldots, w_m\}$ uma base ortonormal de W podemos escrever Lv_j, para $j = 1, \ldots, n$, como uma combinação linear dos elementos de $\mathcal{B}_2$

$$\begin{aligned} Lv_1 &= a_{11}w_1 + a_{21}w_2 + \cdots + a_{m1}w_m \\ Lv_2 &= a_{12}w_1 + a_{22}w_2 + \cdots + a_{m2}w_m \\ &\cdots \\ Lv_n &= a_{1n}w_1 + a_{2n}w_2 + \cdots + a_{mn}w_m \end{aligned}$$

Para definir uma transformação linear, basta estabelecer seu valor nos elementos de uma base do seu domínio. Seja $L^* : W \to V$ aquela transformação linear que leva w_i, para $i = 1, 2, \ldots, m$, nos seguintes vetores de V

$$\begin{aligned} L^*w_1 &= \bar{a}_{11}v_1 + \bar{a}_{12}v_2 + \cdots + \bar{a}_{1n}v_n \\ L^*w_2 &= \bar{a}_{21}v_1 + \bar{a}_{22}v_2 + \cdots + \bar{a}_{2n}v_n \\ &\cdots \\ L^*w_m &= \bar{a}_{m1}v_1 + \bar{a}_{m2}v_2 + \cdots + \bar{a}_{mn}v_n \end{aligned}$$

Usando o símbolo de somatório, os valores das transformações lineares L e L^* nas bases de seus respectivos domínios se escrevem

$$\begin{aligned} Lv_j &= \sum_{i=1}^{m} a_{ij}w_i \\ L^*w_i &= \sum_{j=1}^{n} \bar{a}_{ij}v_j. \end{aligned}$$

Da ortonormalidade das bases $\mathcal{B}_1$ e $\mathcal{B}_2$, segue $\langle Lv_j, w_i \rangle = \langle v_j, L^* w_i \rangle$ fazendo com que

$$\langle Lv, w \rangle = \langle v, L^* w \rangle$$

para todo v em V e w em W, o que prova a existência.

Agora a unicidade. Se $T : W \to V$ for outra transformação linear para a qual

$$\langle Lv, w \rangle = \langle v, Tw \rangle$$

para todo v em V e w em W, então

$$\langle v, L^* w \rangle = \langle v, Tw \rangle$$

ou

$$\langle v, (L^* - T) w \rangle = 0$$

ainda para todo v em V e w em W. Fazendo $v = (L^* - T)w$, obtemos

$$\langle \ (L^* - T)w, (L^* - T)w \ \rangle = 0$$

ou $(L^* - T)w = 0$ para todo w em W, mostrando que $T = L^*$, o que prova a unicidade.

A transformação linear L^* recebe o nome de **transformação adjunta** de L.

Se $L : V \to W$ e $T : W \to U$ forem duas transformações lineares, então

$$(TL)^* = L^* T^*.$$

Se $A = [a_{ij}]$ for a matriz de $L : V \to W$ e $B = [b_{ij}]$ for a matriz de $L^* : W \to V$ nas bases ortonormais $\mathcal{B}_1$ e $\mathcal{B}_2$ de V e W, respectivamente, então $b_{ij} = \bar{a}_{ji}$ e

$$B = A^*.$$

Esta relação entre as matrizes de L e L^* só se verifica se as bases forem ortonormais, como nos mostra o próximo exemplo.

Este conceito de transformação adjunta se aplica a transformações lineares entre espaços vetoriais reais. Neste caso, os escalares serão reais e $\bar{a}_{ij} = a_{ij}$ e

Exemplo 6.1 *Seja $L(x, y) = (2x + 3y,\ 5x + 7y,\ 11x + 13y)$ uma transformação linear do $\mathbf{R}^2$ no $\mathbf{R}^3$. Consideremos nestes dois espaços seus respectivos produtos internos euclidianos*

$$\langle \ (x_1, y_1),\ (x_2, y_2) \ \rangle = x_1 x_2 + y_1 y_2$$

e

$$\langle \ (x_1, y_1, z_1),\ (x_2, y_2, z_2) \ \rangle = x_1 x_2 + y_1 y_2 + z_1 z_2.$$

Seja $\mathcal{B}_1 = \{e_1, e_2\}$ a base canônica do $\mathbf{R}^2$ e $\mathcal{B}_2 = (f_1, f_2, f_3)$ a base canônica do $\mathbf{R}^3$ que são ortonormais em relação aos produtos internos euclidianos de $\mathbf{R}^2$ e $\mathbf{R}^3$, respectivamente. Temos

$$\begin{aligned} Le_1 &= 2f_1 + 5f_2 + 11f_3 \\ Le_2 &= 3f_1 + 7f_2 + 13f_3 \end{aligned}$$

e

$$\begin{aligned} L^* f_1 &= 2e_1 + 3e_2 \\ L^* f_2 &= 5e_1 + 7e_2 \\ L^* f_3 &= 11e_1 + 13e_2 \end{aligned}$$

de modo que

$$[L]_{12} = \begin{bmatrix} 2 & 3 \\ 5 & 7 \\ 11 & 13 \end{bmatrix} \quad , \qquad [L^*]_{21} = \begin{bmatrix} 2 & 5 & 11 \\ 3 & 7 & 13 \end{bmatrix}$$

onde uma é a transposta da outra.

Entretanto, se $v_1 = (1,\, 2)$, $v_2 = (0,\, 1)$, $w_1 = (1,\, 1,\, 1)$, $w_2 = (0,\, 1,\, 2)$ *e* $w_3 = (0,\, 0,\, 1)$ *então* $\mathcal{B}_3 = \{v_1,\, v_2\}$ *será base de* $\mathbf{R}^2$ *e* $\mathcal{B}_4 = \{w_1,\, w_2,\, w_3\}$ *será base de* $\mathbf{R}^3$*. Nenhuma das duas é ortonormal em relação ao produto interno euclidiano do* $\mathbf{R}^2$ *e* $\mathbf{R}^3$*. O leitor poderá verificar que*

$$\begin{aligned} Lv_1 &= 8w_1 + 11w_2 + 7w_3 \\ Lv_2 &= 3w_1 + 4w_2 + 2w_3 \end{aligned}$$

e

$$\begin{aligned} L^* w_1 &= 18v_1 - 13v_2 \\ L^* w_2 &= 27v_1 - 21v_2 \\ L^* w_3 &= 11v_1 - 9v_2 \end{aligned}$$

As matrizes

$$[L]_{34} = \begin{bmatrix} 8 & 3 \\ 11 & 4 \\ 7 & 2 \end{bmatrix} \qquad e \qquad [L^*]_{43} = \begin{bmatrix} 18 & 27 & 11 \\ -13 & -21 & -9 \end{bmatrix}$$

não são mais uma a adjunta da outra.

Dada a transformação linear L, a transformação adjunta L^* é a única que satisfaz à igualdade $\langle Lv, w\rangle = \langle v, L^* w\rangle$ para todo v em V e todo w em W. De fato, se T for outra transformação linear para a qual $\langle Lv, w\rangle = \langle v, Tw\rangle$ para todo v e todo w, então $\langle v, Tw\rangle = \langle v, L^* w\rangle$ e $\langle v, Tw - L^* w\rangle = 0$ para todo v o que acarreta na igualdade $Tw = L^* w$ para todo w, nos conduzindo à igualdade $T = L^*$.

Para todo v em V e w em W, tem-se

$$\langle L^* w, v\rangle = \overline{\langle v, L^* w\rangle} = \overline{\langle Lv, w\rangle} = \langle w, Lv\rangle\,,$$

mostrando que a adjunta da adjunta é igual a L, isto é,

$$(L^*)^* = L.$$

Um operador linear $L : V \to V$ é **auto-adjunto** quando $L^* = L$.

Os operadores LL^* e $L^* L$ são auto-adjuntos.

Exemplo 6.2 *Sejam* $x = (x_1, x_2, x_3)$ *e* $y = (y_1, y_2, y_3)$ *dois pontos do* $\mathbf{R}^3$, *consideremos o produto interno* $\langle x, y\rangle = x_1y_2 + x_2y_2 + x_2y_3$. *Em relação a este produto interno, o operador linear* $L : \mathbf{R}^3 \to \mathbf{R}^3$ *definido por* $L(x, y, z) = (2x + 3y + 5z, 3x + y, 5x + 4z)$ *é auto-adjunto pois* $L^* = L$.

Teorema 6.3 *Sejam* V *e* W *espaços vetoriais com produto interno e* $L : V \to W$ *linear. Sendo* $L^* : W \to V$ *a adjunta de* L, *vale a relação*

$$\mathrm{Im}\,(L)^\perp = \ker(L^*).$$

Prova. Se $w \in \mathrm{Im}\,(L)^\perp$, então $\langle Lv, w\rangle = 0$ ou $\langle v, L^*w\rangle = 0$ para todo $v \in V$ de onde se conclui que $L^*w = 0$ mostrando que w está no $\ker(L^*)$.

Reciprocamente, se $w \in \ker(L^*)$, então $\langle Lv, w\rangle = \langle v, L^*w\rangle = \langle v, 0\rangle = 0$ para todo $v \in V$, provando, deste modo, que w é ortogonal a todo elemento da imagem de L. Conclui-se que w pertence ao complemento ortogonal da $\mathrm{Im}\,(L)$. $\square$

Como $L^{**} = L$, substituindo L por L^* na igualdade acima, segue

$$\mathrm{Im}\,(L^*)^\perp = \ker(L)$$

ou

$$\mathrm{Im}\,(L^*) = \ker(L)^\perp$$

Sejam V e W espaços vetoriais com produto interno e $L : V \to W$ linear. Sendo $L^* : W \to V$ a adjunta de L,

$$V = \ker(L) \oplus \mathrm{Im}\,(L^*).$$

Esta igualdade ocorre porque $V = \ker(L) \oplus \ker(L)^\perp = \ker(L) \oplus \mathrm{Im}\,(L^*)$.

Quando L é auto-adjunta,

$$\mathrm{Im}\,(L)^\perp = \ker(L).$$

Teorema 6.4 *Sejam* V *e* W *espaços vetoriais com produto interno e* $L : V \to W$ *linear. Sendo* $L^* : W \to V$ *a adjunta de* L,

1. $\mathrm{Im}\,(L) = \mathrm{Im}\,(LL^*)$.
2. $\ker(L^*) = \ker(LL^*)$.

Prova. 1a. Inicialmente provaremos que $\mathrm{Im}\,LL^* \subset \mathrm{Im}\,L$. Se $w \in \mathrm{Im}\,(LL^*)$, então existe um w_1 tal que $w = LL^*w_1 = L(L^*w_1)$ provando que $w \in \mathrm{Im}\,(L)$.

1b. Vamos provar agora que $\mathrm{Im}\,(L) \subset \mathrm{Im}\,(LL^*)$. Se $w \in \mathrm{Im}\,(L)$, então $w = Lv$ para algum v em V. Podemos escrever de modo único $v = v_1 + v_2$ onde $v_1 \in \mathrm{Im}\,(L^*)$ e $v_2 \in \ker(L)$. Logo $w = Lv = Lv_1 + Lv_2 = Lv_1$. Como $v_1 \in \ker(L)^\perp = \mathrm{Im}\,(L^*)$, existe w_1 tal que $v_1 = L^*w_1$ e assim $w = LL^*w_1$, mostrando que $w \in \mathrm{Im}\,(LL^*)$, o que completa a prova da recíproca.

2a. Se $w \in \ker(L^*)$ então $L^*w = 0$ e, em consequência, $LL^*w = 0$, provando que $\ker(L^*) \subset \ker(LL^*)$.

2b. Se $w \in \ker(LL^*)$ então $L(L^*w) = 0$ e L^*w pertence ao $\ker(L)$ e à $\text{Im}\,(L^*)$ cuja interseção contém apenas o zero. Logo, $L^*w = 0$, provando que w está no $\ker(L^*)$. Com isto, provamos que $\ker(LL^*) \subset \ker(L^*)$, o que completa a prova da parte 2 do teorema. $\square$

Resumindo: Para uma transformação linear $L : V \to W$ e sua adjunta $L^* : W \to V$ valem as identidades

$$\begin{aligned} (L^*)^* &= L \\ \text{Im}\,(L^*) &= \ker(L)^\perp \\ \text{Im}\,(LL^*) &= \text{Im}\,(L), \\ \ker(LL^*) &= \ker(L^*). \end{aligned}$$

Definição 6.5 *Seja V um espaço vetorial com produto interno. O operador linear $L : V \to V$ é* ***antiadjunto*** *quando $L^* = -L$ e* ***unitário*** *quando $L^* = L^{-1}$.*

Teorema 6.6 *Numa base ortonormal, a matriz $A = [a_{ij}]$ de um operador auto-adjunto é* ***hermitiana*** *($A = A^*$), a de um operador antiadjunto é* ***antihermitiana*** *($A = -A^*$) e a de um operador unitário é* ***unitária*** *($A^* = A^{-1}$).*

Teorema 6.7 *O operador linear $L : V \to V$ é unitário se e só se, para todo v_1 e v_2 em V,*

$$\langle Lv_1, Lv_2 \rangle = \langle v_1, v_2 \rangle .$$

Para fixar a nomenclatura, apresentamos o quadro abaixo.

Espaço Real		Espaço Complexo	
Operador	Matriz	Operador	Matriz
auto-adjunto	simétrica	auto-adjunto	hermitiana
antiadjunto	anti-simétrica	antiadjunto	antihermitiana
ortogonal	ortogonal	unitário	unitária

6.1 Posto de uma transformação linear

Definição 6.8 *Sejam V e W espaços vetoriais e $L : V \to W$ uma transformação linear. A dimensão da imagem de L é chamada de* ***posto*** *de L.*

Sendo A uma matriz complexa de ordem m por n. Considerada como uma transformação linear de $\mathbf{C}^n$ em $\mathbf{C}^m$, seu ***posto*** *é igual ao número de colunas linearmente independentes que A possui.*

O número de colunas linearmente independentes de uma matriz é igual ao número de suas linhas linearmente independente.

Teorema 6.9 *Sejam V e W espaços vetoriais de dimensão finita e $L : V \to W$ uma transformação linear. Seja A a representação matricial de L em relação a bases de V e de W. O posto de L é igual ao posto de A.*

Exemplo 6.10 *Seja $L : \mathbf{R}^3 \to \mathbf{R}^3$ definida por*

$$\begin{aligned} L(x,y,z) &= (\ x+2y-z,\ 2x+4y-2z,\ y+2z\ ) \\ &= (x+2y-z)(1,2,0)+(y+2z)(0,0,1), \end{aligned}$$

cujo posto é 2. A matriz de L em relação à base canônica do $\mathbf{R}^3$ é

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 4 & -2 \\ 0 & 1 & 2 \end{bmatrix}$$

a primeira e terceira linha são linearmente independentes e a segunda é o dobro da primeira. As duas primeiras colunas são linearmente independentes e a terceira é igual a -5 vezes a primeira mais 2 vezes a segunda. O posto de A é dois.

Teorema 6.11 *Sejam V e W espaços vetoriais com produto interno e dimensão finita. Seja $L : V \to W$ linear. O posto das transformações lineares L, L^*, L^*L e LL^* são iguais.*

Prova. Sabemos que

$$V = \ker(L) \oplus \operatorname{Im}(L^*),$$

de onde obtemos

$$\dim\ker(L) + \dim\operatorname{Im}(L^*) = \dim V.$$

Por outro lado,

$$\dim\ker(L) + \dim\operatorname{Im}(L) = \dim V.$$

Dessas duas igualdades concluímos que $\dim\operatorname{Im}(L^*) = \dim\operatorname{Im}(L)$ provando que o posto de L é igual ao posto de L^*.

Como $\operatorname{Im}(LL^*) = \operatorname{Im}(L)$ e $\operatorname{Im}(L^*L) = \operatorname{Im}(L^*)$, o teorema está provado. $\square$

Corolário 6.12 *Seja A uma matriz complexa. As matrizes A, A^*, AA^* e A^*A possuem o mesmo posto.*

Sejam V e W espaços vetoriais, ambos com dimensão finita. A imagem de uma transformação linear $L : V \to W$ está contida em W. Portanto, o posto de L deve ser menor ou igual do que a dimensão de W. Se $\{v_1, \ldots, v_n\}$ for uma base de V, qualquer vetor v em V pode ser decomposto de modo único como uma combinação linear $v = x_1v_1 + \cdots + x_nv_n$ e assim, $Lv = L(x_1v_1 + \cdots + x_nv_n) = x_1Lv_1 + \cdots + x_nLv_n$, mostrando que $\{Lv_1, \ldots, Lv_n\}$ gera a imagem de L o que assim o posto de L deve ser menor ou igual do que a dimensão de V. Concluímos que o posto de L não pode ser maior do que a dimensão de V nem maior do que a dimensão de W. Motivados por este comentário, diremos que L tem **posto máximo** quando o posto de L for igual ao mínimo entre a dimensão de V e a dimensão de W.

Teorema 6.13 *Sejam V e W espaços vetoriais com dimensão finita, ambos com produto interno. Seja $L : V \to W$ uma transformação linear com posto máximo.*

1. *Quando* $\dim V \leq \dim W$, *a transformação linear* $L^*L : V \to V$ *é um isomorfismo.*

2. *Quando* $\dim W \leq \dim V$, *a transformação linear* $LL^* : W \to W$ *é um isomorfismo.*

Prova. Quando

1. *posto*$(L) = \dim V \leq \dim W$, então $\dim \operatorname{Im}(L^*L) = \dim \operatorname{Im}(L) = \dim(V)$ e assim L^*L é sobrejetora e, portanto, um isomorfismo.

2. *posto*$(L) = \dim W \leq \dim V$, então $\dim \operatorname{Im}(LL^*) = \dim \operatorname{Im}(L^*) = \dim(W)$ e assim LL^* é sobrejetora e, portanto, um isomorfismo.

□

6.2 Existência de solução dos sistemas lineares

As igualdades $\operatorname{Im}(L)^\perp = \ker(L^*)$ e $\operatorname{Im}(L^*) = \ker(L)^\perp$ possuem uma consequência interessante para sistemas de equações lineares $Ax = b$, onde A é uma matriz complexa $m \times n$ e b é uma matriz complexa $m \times 1$. Nestes sistemas, as matrizes A e b são dadas e o que se deseja é determinar se existem matrizes coluna complexas x de tamanho $n \times 1$ tais que $Ax = b$. Tais matrizes x são denominadas soluções do sistema $Ax = b$. Existindo soluções, é importante determiná-las. Um método usado na obtenção das soluções é o da eliminação de Gauss.

A matriz A é uma transformação linear de $\mathbf{C}^{n\times 1}$ em $\mathbf{C}^{m\times 1}$.

O sistema $Ax = b$ tem solução se e só se b estiver na imagem de A. Da igualdade $\operatorname{Im}(A) = \ker(A^*)^\perp$ conclímos que $Ax = b$ tem solução se e só se b for ortogonal a todo y no núcleo de A^* isto é,

$$\langle b, y\rangle = 0$$

para todo y em $\mathbf{C}^{m\times 1}$ solução do sistema homogêneo $A^*y = 0$.

O sistema homogêneo $Ax = 0$ tem solução x não nula se e só se x pertencer ao núcleo de A. Da igualdade $\ker(A) = \operatorname{Im}(A^*)^\perp$, concluímos que x é solução do sistema homogêneo $Ax = 0$ se e só se x for ortogonal à imagem de A^*, isto é $\langle x, A^*y\rangle =$ para todo y em $\mathbf{C}^{m\times 1}$.

Percebe-se do comentado acima que há uma estreita relação entre os sistemas lineares $Ax = b$ e $A^*y = c$.

Capítulo 7

Projetores

Seja V um espaço vetorial. Um operador linear $P : V \to V$ é um **projetor** em V se $P^2 = P$. Sendo $I : V \to V$ o operador identidade, o operador linear $I - P$ também é um projetor, denominado **projetor complementar** de P. Os projetores também recebem o nome de operadores **idempotentes**.

Sejam S_1 e S_2 dois subespaços de V tais que $V = S_1 \oplus S_2$. Considere o operador $P : V \to V$ definido por $P(v_1 + v_2) = v_1$, para todo v_1 em S_1 e v_2 em S_2. O operador assim definido é um projetor, denominado **projetor sobre** S_1 **ao longo de** S_2. Sob estas condições, S_1 é a imagem de P e S_2 é o núcleo de P.

Se v estiver na imagem de P, então existe w em V tal que $Pw = v$. Sendo P uma projeção, $P^2w = Pw$ o que implica em $Pv = v$. A imagem de $(I - P)$ é igual ao núcleo de P e a imagem de P é igual ao núcleo de $I - P$.

Teorema 7.1 *Seja $P : V \to V$ um projetor. Então $V = \text{Im}\,(P) \oplus \ker(P)$.*

Prova. (1) Seja v um vetor em V. Podemos escrever $v = Pv + (I - P)v$. Como Pv está na imagem de P e $(I - P)v$ está no núcleo de V, segue $V = \text{Im}\,(P) + \ker(P)$.

(2) Se v_1 na $\text{Im}\,(P)$ e v_2 no $\ker(P)$ forem tais que $v = v_1 + v_2$, então $Pv = Pv_1 + Pv_2$. Como $Pv_1 = v_1$ e $Pv_2 = 0$, segue $Pv = v_1$ e $(I - P)v = v_2$, mostrando que a decomposição de v numa soma de um elemento da $\text{Im}\,(P)$ com um elemento do $\ker(P)$ é única. Logo, $V = \text{Im}\,(P) \oplus \ker(P)$. $\square$

De acordo com este teorema, todo projetor P é um projetor sobre sua imagem ao longo do seu núcleo.

7.1 Projetores ortogonais

Seja V um espaço vetorial com produto interno. Um projetor P em V é **ortogonal** se a sua imagem e seu núcleo forem ortogonais. Quando este for o caso, se diz que P **projeta ortogonalmente** sobre sua imagem.

Seja $P : V \to V$ um projetor ortogonal e S sua imagem. Se a dimensão de S for finita, $V = S \oplus S^{\perp}$. Dado v em V, existe um único s em S e um único w em $S^{\perp}$ para os quais $v = s + w$ e

$$P(v) = s.$$

Se $\mathcal{B} = \{q_1, \ldots, q_k\}$ for uma base ortonormal de S, podemos decompor Pv nesta base e escrever

$$Pv = x_1 q_1 + \cdots + x_k q_k.$$

Para determinar $x_1, \ldots, x_k$, usamos o fato de $v - Pv$ ser ortogonal a todo vetor de S. Isto significa que $\langle q_i, v - Pv \rangle = 0$ para $i = 1, \ldots, k$. Destas relações e da ortonomalidade da base $\mathcal{B}$ segue

$$\begin{aligned} 0 &= \langle q_i, v - Pv \rangle = \langle q_i, v \rangle - \langle q_i, x_1 q_1 + \cdots + x_k q_k \rangle \\ &= \langle q_i, v \rangle - x_1 \langle q_i, q_1 \rangle - \cdots - x_i \langle q_i, q_i \rangle - \cdots - x_k \langle q_i, q_k \rangle \\ &= \langle q_i, v \rangle - x_i \langle q_i, q_i \rangle = \langle q_i, v \rangle - x_i \end{aligned}$$

ou

$$x_i = \langle q_i, v \rangle$$

o que nos permite escrever

$$Pv = \langle q_1, v \rangle q_1 + \cdots + \langle q_k, v \rangle q_k.$$

e provamos o próximo teorema.

Teorema 7.2 *Seja S um subespaço de dimensão finita de um espaço vetorial V com produto interno. Seja $\{q_1, \ldots, q_k\}$ uma base ortonormal de S. Se P for o projetor ortogonal sobre S, então, para todo v em V,*

$$Pv = \langle q_1, v \rangle q_1 + \cdots + \langle q_k, v \rangle q_k.$$

A partir deste teorema obtemos outro de imediato para projetores em $\mathbf{C}^{n \times 1}$. Vamos lembrar que toda transformação linear L de $\mathbf{C}^{n \times 1}$ em $\mathbf{C}^{n \times 1}$ é do tipo $L(x) = Ax$, onde A é uma matriz quadrada $n \times n$ e iremos identificar a transformação linear L com a matriz A. Se P for uma projeção em $\mathbf{C}^{n \times 1}$, então P será uma matriz $n \times n$.

Corolário 7.3 *Considere o espaço vetorial $\mathbf{C}^{n \times 1}$ com o produto interno $\langle x, y \rangle = x^* y$. Seja P um projetor ortogonal em $\mathbf{C}^{n \times 1}$ e $\{q_1, \ldots, q_k\}$ uma base ortonormal da imagem de P. Então*

$$P = q_1 q_1^* + \cdots + q_k q_k^*.$$

Prova. Observe que $q_1, q_2, \ldots, q_k$ são matrizes coluna do $\mathbf{C}^{n \times 1}$. Sendo x uma matriz coluna em $\mathbf{C}^{n \times 1}$, podemos escrever

$$\langle q_i, x \rangle q_i = (q_i^* x) q_i = q_i (q_i^* x) = (q_i q_i^*) x$$

e assim,

$$\begin{aligned} Px &= \langle q_1, x\rangle\, q_1 + \cdots + \langle q_k, x\rangle\, q_k \\ &= (q_1 q_1^*)x + \cdots + (q_k q_k^*)x \\ &= (q_1 q_1^* + \cdots + q_k q_k^*)\, x. \end{aligned}$$

Como esta igualdade vale para todo x,

$$P = q_1 q_1^* + \cdots + q_k q_k^*.$$

□

Quando P projeta ortogonalmente sobre o espaço gerado por um único vetor unitário q, temos

$$P = qq^*.$$

Se x for uma matriz coluna não nula em $\mathbf{C}^{n\times 1}$, não necessariamente unitário, então $q = x/\|x\|$ é unitário e a projeção ortogonal P sobre o espaço gerado por x é

$$P = qq^* = \frac{x}{\|x\|}\frac{x^*}{\|x\|} = \frac{xx^*}{x^*x}.$$

Vamos relembrar que, sendo x uma matriz coluna do $\mathbf{C}^{n\times 1}$, então x^*x é um número real e xx^* é uma matriz complexa n por n.

O projetor complementar de P é

$$I - \frac{xx^*}{x^*x},$$

onde I é a matriz identidade $n \times n$ e sua imagem é o núcleo de P.

Teorema 7.4 *Seja V um espaço vetorial com dimensão finita e um produto interno. Um projetor $P : V \to V$ é ortogonal se e só se for auto-adjunto.*

Prova. (1) Seja P uma projeção ortogonal e $S = \operatorname{Im}(P)$. Então P é uma projeção sobre S ao longo de $S^\perp$. Sabemos que $V = S \oplus S^\perp$ e, para todo v e w em V, existem e são únicos v_1 e w_1 em S, v_2 e w_2 em $S^\perp$ para os quais, $v = v_1 + v_2$ e $w = w_1 + w_2$. Assim, da ortogonalidade dos vetores,

$$\langle Pv, w\rangle = \langle v_1, w_1 + w_2\rangle = \langle v_1, w_1\rangle$$

e

$$\langle v, Pw\rangle = \langle v_1 + v_2, w_1\rangle = \langle v_1, w_1\rangle$$

provando que P é auto-adjunto.

(2) Seja P uma projeção auto-adjunta de modo que

$$\langle Pv, w\rangle = \langle v, Pw\rangle$$

para todo v e w em V. Se w estiver no $\ker(P)$, então

$$\langle Pv, w\rangle = \langle v, Pw\rangle = \langle v, 0\rangle = 0,$$

e, como Pv está na imagem de P, provamos que o núcleo e a imagem de P são ortogonais. Logo, P é um projetor ortogonal. □

Teorema 7.5 *Seja V um espaço vetorial com produto interno e $P : V \to V$ um projetor ortogonal. O vetor Pv é o vetor da* $\mathrm{Im}\,(P)$ *mais próximo de v.*

Prova. Todo vetor u da imagem de P pode ser escrito na forma $Pv+ w$, com w na imagem de P. Assim, $v - Pv$ pertence ao $\ker(P)$ que é ortogonal à $\mathrm{Im}\,(P)$ e acarretando na ortogonalidade $\langle v - Pv, w\rangle = 0$ e

$$\begin{aligned} \|v-u\|^2 &= \|v - Pv - w\|^2 = \langle v - Pv - w, v - Pv - w\rangle \\ &= \langle v - Pv, v - Pv\rangle - \langle v - Pv, w\rangle - \langle w, v - Pv\rangle + \langle w, w\rangle \\ &= \|w\|^2 + \|v - Pv\|^2 \geq \|v - Pv\|^2 \end{aligned}$$

provando que Pv é o ponto da imagem de P mais próximo de v. □

7.2 Projetores ortogonais em $\mathbf{C}^{m\times 1}$

Nesta seção vamos considerar o espaço vetorial $\mathbf{C}^{m\times 1}$ com o produto interno $\langle x, y\rangle = x^* y$.

Usando uma base ortonormal

Seja P um projetor ortogonal em $\mathbf{C}^{m\times 1}$ e S sua imagem. Sendo P ortogonal, seu núcleo é $S^\perp$. Seja $\{q_1, \ldots, q_n\}$ uma base ortonormal de S e $\{q_{n+1}, \ldots, q_m\}$ uma base ortonormal de $S^\perp$. Pelas propriedades provadas para uma projeção ortogonal P,

$$\begin{aligned} Pq_i &= q_i \quad \text{para} \quad i = 1, \ldots, n \\ Pq_i &= 0 \quad \text{para} \quad i = n+1, \ldots, m \end{aligned}$$

Seja

$$Q = [\ q_1,\ \ldots,\ q_n,\ q_{n+1},\ \ldots,\ q_m\ ]$$

a matriz cujas colunas são os vetores da base ortonormal $\{q_1, \ldots, q_n, q_{n+1}, \ldots, q_m\}$ do $\mathbf{C}^{m\times 1}$ e para a qual

$$PQ = Q\Sigma,$$

onde Σ é uma matriz quadrada $m \times m$, onde os n primeiros elementos da diagonal são iguais a 1 e todos os demais elementos são nulos. Podemos escrevê-la usando blocos

$$\Sigma = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}$$

onde I é a matriz identidade de tamanho $n \times n$. A matriz Q é unitária pois suas colunas são formadas a partir de uma base ortonormal de $\mathbf{C}^{m\times 1}$. Sendo QQ^* a matriz identidade, chega-se a

$$P = Q\Sigma Q^*.$$

Observe que apenas as n primeiras colunas de Q são relevantes neste produto. Eliminando-as obtemos a matriz $\hat{Q} = [q_1, \ldots, q_n]$ com m linhas e n colunas para a qual

$$P = \hat{Q}\hat{Q}^*.$$

Como já se provou, $P = \sum_{i=1}^{k} q_i q_i^*$ e dela obtemos a identidade

$$P = \hat{Q}\hat{Q}^* = \sum_{i=1}^{k} q_i q_i^*$$

Usando uma base qualquer

Seja P um projetor ortogonal em $\mathbf{C}^{m\times 1}$ e $v_1, \ldots, v_n$ matrizes coluna em $\mathbf{C}^{m\times 1}$ tais que $\mathcal{B} = \{v_1, \ldots, v_n\}$ é uma base da imagem de P. Para todo v em $\mathbf{C}^{m\times 1}$, podemos escrever

$$Pv = x_1 v_1 + \cdots + x_k v_n = Ax,$$

onde $x = [x_1, \ldots, x_n]^T$ é a matriz das coordenadas de Pv na base $\mathcal{B}$ e $A = [v_1, \ldots, v_n]$ é uma matriz de ordem m por n, cujas colunas são $v_1, \ldots, v_n$. Esta matriz A define uma transformação linear de $\mathbf{C}^{n\times 1}$ em $\mathbf{C}^{m\times 1}$.

O vetor $v - Pv$ está no complemento ortogonal da imagem de P e, para $j = 1, \ldots, n$,

$$\langle v_j, v - Pv\rangle = 0 \qquad \text{ou} \qquad v_j^* P v = v_j^* v,$$

de onde segue

$$v_j^* A x = v_j^* v.$$

Como a $j-$ ésima linha de A^* é v_j^*, a identidade acima resulta em

$$A^* A x = A^* v.$$

Como as colunas de A são linearmente independentes, seu posto é máximo e $A^*A : \mathbf{C}^{n\times 1} \to \mathbf{C}^{n\times 1}$ é um isomorfismo, possuindo assim uma inversa, o que nos permite escrever

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

de onde resulta

$$Pv = Ax = A(A^*A)^{-1}A^*v.$$

Como esta igualdade vale para todo v em $\mathbf{C}^{m\times 1}$,

$$P = A(A^*A)^{-1}A^*.$$

Se a base $\mathcal{B}$ for ortonormal, a matriz A é unitária e daí $A^* = A^{-1}$. Com isto reobtemos $P = AA^*$, válida quando usamos uma base ortonormal.

Exemplo 7.6 *Determine o projetor ortogonal P sobre o espaço gerado pelos vetores $v_1 = (1,\ 2,\ 0)^T$ e $v_2 = (0,\ 1,\ 1)^T$.*

Inicialmente estabelecemos

$$A = [v_1, v_2] = \begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 0 & 1 \end{pmatrix}$$

e calculamos

$$P = A(A^*A)^{-1}A^* = \begin{pmatrix} 1/3 & 1/3 & -1/3 \\ 1/3 & 5/6 & 1/6 \\ -1/3 & 1/6 & 5/6 \end{pmatrix}.$$

Observe que $Pv_1 = v_1$ e $Pv_2 = v_2$.

7.3 Ortogonalização de Gram-Schmidt em $\mathbf{C}^{m\times 1}$

Vamos considerar o espaço vetorial $\mathbf{C}^{m\times 1}$ com o produto interno $\langle x, y\rangle = x^*y$. Pelo processo de ortogonalização de Gram-Schmidt, partindo de uma base $\{v_1,\ \ldots,\ v_m\}$ de $\mathbf{C}^{m\times 1}$, pode-se obter uma base ortonormal $\{q_1,\ \ldots,\ q_m\}$, de modo iterativo

$$\begin{aligned} w_1 &= v_1 \ \text{ e } \ q_1 = \frac{w_1}{\|w_1\|} \\ w_2 &= v_2 - \langle q_1, v_2\rangle\, q_1 \ \text{ e } \ q_2 = \frac{w_2}{\|w_2\|} \\ w_3 &= v_3 - \langle q_1, v_3\rangle\, q_1 - \langle q_2, v_3\rangle\, q_2 \ \text{ e } \ q_3 = \frac{w_3}{\|w_3\|} \\ &\cdots \end{aligned}$$

A partir de $q_1 = v_1 / \|v_1\|$ determinam-se recursivamente os demais elementos da base ortonormal, mediante a fórmula

$$w_j = v_j - \sum_{i=1}^{j-1} \langle q_i, v_j\rangle\, q_i \ \text{ e } \ q_j = \frac{w_j}{\|w_j\|},$$

válida para $j = 2,\ \ldots,\ m$. Como $\langle q_i, v_j\rangle\, q_i = q_i q_i^* v_j$, esta recorrência pode ser reescrita na forma

$$w_j = v_j - \sum_{i=1}^{j-1} q_i q_i^* v_j = \left(1 - \sum_{i=1}^{j-1} q_i q_i^*\right) v_j$$

A projeção ortogonal sobre o subespaço S_j gerado por $\{q_1,\ \ldots,\ q_j\}$ é

$$\hat{Q}_j \hat{Q}_j^* = \sum_{i=1}^{j} q_i q_i^*$$

onde $\hat{Q}_j = [q_1, \ldots, q_j]$ é a matriz cujas colunas são $q_1, \ldots, q_j$. A projeção ortogonal sobre o complemento orgogonal de S_j é

$$P_j = I - \hat{Q}_j \hat{Q}_j^*$$

e a fórmula recursiva pode ser escrita de modo conciso como

$$w_j = P_{j-1} v_j \quad \text{e} \quad q_j = \frac{w_j}{\|w_j\|}.$$

Conclui-se que o algoritmo clássico de Gram-Schmidt pode ser expresso em termos destes projetores

$$q_1 = \frac{P_0 v_1}{\|P_0 v_1\|}, \quad q_2 = \frac{P_1 v_2}{\|P_1 v_2\|}, \quad \ldots, \quad q_m = \frac{P_{m-1} v_m}{\|P_{m-1} v_m\|},$$

onde P_0 é a identidade e P_j, para $j = 1, \ldots, n-1$, é a projeção ortogonal sobre o complemento ortogonal do espaço gerado por $\{q_1, \ldots, q_j\}$.

7.4 Ortogonalização modificada de Gram-Schmidt

Se q for uma matriz coluna unitária em $\mathbf{C}^{m\times 1}$, a projeção ortogonal sobre o complemento ortogonal do espaço gerado por q é

$$P_{\perp q} = I - qq^*.$$

Vamos, a partir de um conjunto linearmente independente $\{v_1, \ldots, v_n\}$ em $\mathbf{C}^{m\times 1}$, obter um conjunto ortonormal de matrizes coluna $\{q_1, \ldots, q_n\}$ em $\mathbf{C}^{m\times 1}$. Observe que

$$P_{\perp q_2} P_{\perp q_1} = (I - q_2^* q_2)(I - q_1 q_1^*) = I - q_1 q_1^* - q_2 q_2^* + q_1 q_1^* q_2 q_2^* = I - q_1 q_1^* - q_2 q_2^*$$

pois $q_1 q_1^* q_2 q_2^* = 0$, uma vez que $q_1^* q_2 = 0$. Prosseguindo com esse raciocínio, obtemos

$$\begin{aligned} P_{\perp q_j} \cdots P_{\perp q_2} P_{\perp q_1} &= \prod_{i=1}^{j} (I - q_i q_i^*) = \\ &= I - \sum_{i=1}^{j} q_i q_i^* = I - \hat{Q}_j \hat{Q}_j^* \end{aligned}$$

uma vez que $q_r q_r^* q_s q_s^* = 0$ para todo $r \neq s$. O projetor $P_j = I - \hat{Q}_j \hat{Q}_j^*$ é exatamente aquele usado no algoritmo de Gram-Schmidt. A identidade

$$P_j = P_{\perp q_j} \cdots P_{\perp q_2} P_{\perp q_1}$$

será usada no algoritmo modificado. A obtenção de P_j através das projeções sucessivas $P_{\perp q_j} \cdots P_{\perp q_2} P_{\perp q_1}$ é mais estável numericamente do que o cálculo clássico através da matriz P_j. Em lugar de calcular w_j pela fórmula,

$$w_j = P_{j-1} v_j$$

podemos usar outra

$$w_j = P_{\perp q_{j-1}} \cdots P_{\perp q_2} P_{\perp q_1} v_j.$$

O algoritmo modificado calcula w_j usando a seqüência

$$\begin{aligned}
w_j^{(1)} &= v_j \\
w_j^{(2)} &= P_{\perp q_1} w_j^{(1)} = w_j^{(1)} - q_1 q_1^* w_j^{(1)} \\
w_j^{(3)} &= P_{\perp q_2} w_j^{(2)} = w_j^{(2)} - q_2 q_2^* w_j^{(2)}, \\
&\vdots \\
w_j &= w_j^{(j)} = P_{\perp q_{j-1}} w_j^{(j-1)} = w_j^{(j-1)} - q_{j-1} q_{j-1}^* w_j^{(j-1)}.
\end{aligned}$$

Na aritmética computacional de precisão finita, este algoritmo introduz erros menores do que o algoritmo clássico.

================================

Algoritmo 8.2 Gram-Schmidt modificado (estável)
Entrada: Um conjunto $\{v_1, \ldots, v_n\}$ em $\mathbf{C}^{m \times 1}$ linearmente independente
Saída: Um conjunto ortonormal $\{q_1, \ldots, q_m\}$ em $\mathbf{C}^{m \times 1}$

================================

for $j = 1$ to n
  $w_j = v_j$
for $j = 1$ to n
  $r_{jj} = w_j^* w_j$
  $q_j = w_j / r_{jj}$
  for $k = j + 1$ to n
    $r_{jk} = q_j^* w_k$
    $w_k = w_k - r_{jk} q_j$

================================

Na prática, pode-se sobrescrever v_j com w_j e sobrescrever w_j com q_j para economizar memória.

7.5 Contagem das operações

Vamos calcular o número de flops realizados na execução do algoritmo modificado de Gram-Schmidt. Cada operação realizada contabilizará um flop em nossa contagem. Esta operação pode ser uma adição, uma subtração, uma multiplicação, uma divisão ou a extração de uma raiz quadrada. Quando m e n forem grandes, o loop que domina o algoritmo é o mais interno

  for $k = j + 1$ to n
    $r_{jk} = q_j^* w_k$
    $w_k = w_k - r_{jk} q_j$

O produto interno $q_j^* w_k$ requer m multiplicações e $m-1$ adições. O cálculo de $w_k - r_{jk} q_j$ necessita de m multiplicações e um igual número de subtrações. Somamos $4m$ flops para um único laço do loop. Como o laço em k, que varia de $j+1$ a n, está dentro de outro em j que varia de 1 a n, o número de flops usado neste algoritmo é

$$\sum_{j=1}^{n} \sum_{k=j+1}^{n} 4m = 4m \sum_{j=1}^{n} (n-j) = 4m \frac{n^2 - n}{2} = 2mn^2 - 2mn \sim 2mn^2,$$

onde o símbolo $\sim$ tem o seguinte significado

$$\lim_{m,n \to \infty} \frac{\text{número de flops}}{2mn^2} = 1.$$

Concluimos que a fatoração QR usando Gram-Schmidt modificado demanda a realização de $\sim 2mn^2$ flops.

Capítulo 8

Refletor de Householder

Seja v um vetor não nulo de $\mathbf{C}^m$. A matriz

$$H_v = I - 2\frac{vv^*}{v^*v}$$

é chamada de **matriz de Householder** ou **refletor de Householder**. Se u for múltiplo de v e w for ortogonal a v, então

$$H_v u = -u \qquad e \qquad H_v w = w.$$

Todo u que está em S é refletido em $-u$ e todo w no complemento ortogonal de S se mantém inalterado. Esta observação nos permite dizer que a matriz H_v reflete os vetores de $\mathbf{C}^m$ no complemento ortogonal de S.

Teorema 8.1 *Seja v um vetor não nulo em $\mathbf{C}^m$. O refletor*

$$H_v = I - 2\frac{vv^*}{v^*v}$$

é hermitiano e unitário.

O refletor H_v é hermitiano e unitário mas não é um projetor, visto que

$$H_v^2 = I.$$

Sejam x e y dois vetores do $\mathbf{C}^m$ com normas iguais e $\langle x, y\rangle = \langle y, x\rangle$. Os vetores

$$\begin{aligned} v &= \frac{1}{2}(x+y) \\ w &= \frac{1}{2}(x-y) \end{aligned}$$

são ortogonais e o refletor de Householder H_v é tal que

$$H_v x = -y.$$

Este fato pode ser provado escrevendo x e y em termos de v e w

$$x = v + w \qquad \text{e} \qquad y = v - w.$$

Nota 8.2 *Se os elementos de x e y forem todas reais, basta ter $\|x\| = \|y\|$ para garantir a igualdade entre $\langle x, y\rangle$ e $\langle y, x\rangle$.*

Podemos definir um refletor de Householder que leva um vetor x do $\mathbf{C}^m$ num outro $y = (y_1, 0, \ldots, 0)$ onde apenas a primeira coordenada y_1 é não nula. Iremos usá-lo para calcular a decomposição QR de uma matriz usando refletores de Householder, que são operadores auto-adjuntos. Vamos à sua descrição.

O sinal de um número complexo z é definido por $sign(0) = 1$ e, quando $z \neq 0$,

$$sign(z) = \frac{z}{|z|}.$$

Observe que o sinal é um número complexo de módulo unitário. Se z for real, seu sinal será $+1$ ou -1. Para todo número complexo z,

$$sign(z)\overline{sign(z)} = 1.$$

Teorema 8.3 *Seja $x = (x_1, x_2, \ldots, x_m)^T$ um vetor não nulo em $\mathbf{C}^m$, com x_1 complexo não nulo e $e_1 = (1, 0, \ldots, 0)^T$ o primeiro elemento da base canônica do $\mathbf{C}^m$. O refletor H_v, onde*

$$v = \frac{1}{2}\left(x + sign(x_1)\,\|x\|\,e_1\right),$$

leva x em $y = -sign(x_1)\,\|x\|\,e_1$, cujo único elemento não nulo é o primeiro.

Prova. O vetor

$$w = \frac{1}{2}\left(x - sign(x_1)\,\|x\|\,e_1\right)$$

é ortogonal a v e

$$\begin{aligned} x &= u + w, \\ sign(x_1)\,\|x\|\,e_1 &= u - w. \end{aligned}$$

Portanto,

$$H_v(x) = H_v v + H_v w = -v + w = -sign(x_1)\,\|x\|\,e_1.$$

□

O y definido neste teorema tem a forma $(y_1, 0, \ldots, 0)^T$, onde apenas $y_1 = -sign(x_1)\,\|x\|$ não é nulo. Esta escolha de u assegura que $\|x\| \leq \|y\|$, o que fornece uma maior estabilidade numérica à decomposição QR usando refletores de Householder descrita em seguida.

8.1 Decomposição QR usando o refletor de Householder

A decomposição QR, baseada no processo de ortogonalização de Gram-Schmidt é o resultado de sucessivas multiplicações à direita de $A = [v_1, \ldots, v_n]$ por matrizes elementares, todas triangulares superiores, resultando numa matriz ortogonal $Q = [q_1, \ldots, q_n]$

$$A \times R_1 \times \cdots \times R_n = Q.$$

A matriz $R_1 \times \cdots \times R_n$ é triangular superior e sua inversa R nos fornece a decomposição $A = QR$.

Por seu turno, a decomposição QR baseada nas matrizes de Householder será o resultado da multiplicação à esquerda de A por uma seqüência de matrizes ortogonais $Q_1, \ldots,$ Q_n, que resultarão numa matriz triangular R

$$Q_n \times \cdots \times Q_1 \cdot A = R.$$

A matriz $Q_1 \times \cdots \times Q_n$ é ortogonal e sua inversa Q nos fornecerá a decomposição $A = QR$.

A idéia de Householder, proposta em 1958, foi a de escolher Q_k para zerar aqueles elementos da coluna k de A, situados abaixo da diagonal. A multiplicação pela matriz Q_k opera sobre A realizando uma combinação linear das linhas k, $k+1, \ldots, m$, mantendo as primeiras $k-1$ colunas inalteradas e anulando os elementos da coluna k situados abaixo da diagonal. A matriz Q_1 é uma matriz de Householder H_1 e a forma geral de Q_k, para $k = 2, \ldots, n$, é

$$Q_k = \begin{bmatrix} I & 0 \\ 0 & H_k \end{bmatrix}$$

onde I é a matriz identidade de ordem $k-1$ e H_k é uma matriz de Householder cuja ordem é $m - k + 1$.

Se x for o vetor de $\mathbf{C}^{m-k+1}$ formado pelos elementos da coluna k de A, extraídos da diagonal principal para baixo, o refletor de Householder H_k procurado é $I - vv^*/v^*v$, onde

$$v = \frac{1}{2}\left(x + sign(x_1)\,\|x\|\,e_1\right),$$

e $e_1 = (1, 0, \ldots, 0)$ é o primeiro elemento da base canônica do $\mathbf{C}^{m-k+1}$. Pelo que foi visto anteriormente,

$$H_k\, x = (y_1, 0, \ldots, 0)^T$$

onde $y_1 = -sign(x_1)\,\|x\|$ é o único elemento não nulo de $H_k x$.

Sendo

$$A = [a_1^{(1)}, a_2^{(1)}, \ldots, a_n^{(1)}] = \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} & \cdots & a_{1n}^{(1)} \\ a_{21}^{(1)} & a_{22}^{(1)} & a_{23}^{(1)} & \cdots & a_{2n}^{(1)} \\ a_{31}^{(1)} & a_{32}^{(1)} & a_{33}^{(1)} & \cdots & a_{3n}^{(1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1}^{(1)} & a_{m2}^{(1)} & a_{m3}^{(1)} & \cdots & a_{mn}^{(1)} \end{bmatrix}$$

então $Q_1 = H_1 = I - v_1 v_1^* / v_1^* v_1$ onde

$$v_1 = \frac{1}{2}\begin{bmatrix} a_{11}^{(1)} + sign(a_{11}^{(1)}) \left\| a_1^{(1)} \right\| \\ a_{21}^{(1)} \\ a_{31}^{(1)} \\ \vdots \\ a_{m1}^{(1)} \end{bmatrix} \quad \text{e} \quad a_1^{(1)} = \begin{pmatrix} a_{11}^{(1)} \\ a_{21}^{(1)} \\ a_{31}^{(1)} \\ \vdots \\ a_{m1}^{(1)} \end{pmatrix}$$

estão em $\mathbf{C}^m$. Assim,

$$Q_1 A = \begin{bmatrix} -sign(a_{11}) \left\| a_1^{(1)} \right\| & a_{12}^{(2)} & a_{13}^{(2)} & \cdots & a_{1n}^{(2)} \\ 0 & a_{22}^{(2)} & a_{23}^{(2)} & \cdots & a_{2n}^{(2)} \\ 0 & a_{32}^{(2)} & a_{33}^{(2)} & \cdots & a_{3n}^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & a_{m2}^{(2)} & a_{m3}^{(2)} & \cdots & a_{mn}^{(2)} \end{bmatrix}.$$

Eliminando a primeira linha e a primeira coluna de $Q_1 A$, obtemos a matriz

$$A_1 = [a_2^{(2)}, a_3^{(2)}, \ldots, a_n^{(2)}] = \begin{bmatrix} a_{22}^{(2)} & a_{23}^{(2)} & \cdots & a_{2n}^{(2)} \\ a_{32}^{(2)} & a_{33}^{(2)} & \cdots & a_{3n}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m2}^{(2)} & a_{m3}^{(2)} & \cdots & a_{mn}^{(2)} \end{bmatrix}$$

que é de ordem $m - 1$ por $n - 1$. Toma-se $H_2 = I - v_2 v_2^* / v_2^* v_2$, onde

$$v_2 = \frac{1}{2}\begin{bmatrix} a_{22}^{(2)} + sign\left(a_{22}^{(2)}\right) \left\| a_2^{(2)} \right\| e_1^{(2)} \\ a_{32}^{(2)} \\ \vdots \\ a_{m2}^{(2)} \end{bmatrix} \quad \text{e} \quad a_2^{(2)} = \begin{pmatrix} a_{22}^{(2)} \\ a_{32}^{(2)} \\ \vdots \\ a_{m2}^{(2)} \end{pmatrix}$$

estão em $\mathbf{C}^{m-1}$. Toma-se

$$Q_2 = \begin{bmatrix} 1 & 0 \\ 0 & H_2 \end{bmatrix}$$

e, com esta escolha,

$$Q_2 Q_1 A = \begin{bmatrix} -sign(a_{11}^{(1)}) \left\| a_1^{(1)} \right\| & a_{12}^{(2)} & a_{13}^{(2)} & \cdots & a_{1n}^{(2)} \\ 0 & -sign(a_{22}^{(2)}) \left\| a_2^{(2)} \right\| & a_{23}^{(3)} & \cdots & a_{2n}^{(3)} \\ 0 & 0 & a_{33}^{(3)} & \cdots & a_{3n}^{(3)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & a_{m3}^{(3)} & \cdots & a_{mn}^{(3)} \end{bmatrix}.$$

Na terceira etapa, toma-se

$$Q_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & H_3 \end{bmatrix}$$

sendo $H_3 = I - v_3 v_3^* / v_3^* v_3$, onde

$$v_3 = \frac{1}{2} \begin{bmatrix} a_{33}^{(3)} + sign\left(a_{33}^{(3)}\right) \left\| a_3^{(3)} \right\| \\ a_{43}^{(3)} \\ \vdots \\ a_{m3}^{(3)} \end{bmatrix} \quad \text{e} \quad a_3^{(3)} = \begin{bmatrix} a_{33}^{(3)} \\ a_{43}^{(3)} \\ \vdots \\ a_{m3}^{(3)} \end{bmatrix}.$$

estão em $\mathbf{C}^{m-2}$. Com esta escolha,

$$Q_2 Q_1 A = \begin{bmatrix} -sign\left(a_{11}^{(1)}\right) \left\| a_1^{(1)} \right\| & a_{12}^{(2)} & a_{13}^{(2)} & \cdots \\ 0 & -sign\left(a_{22}^{(2)}\right) \left\| a_2^{(2)} \right\| & a_{23}^{(3)} & \cdots \\ 0 & 0 & -sign\left(a_{33}^{(3)}\right) \left\| a_3^{(3)} \right\| & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ 0 & 0 & 0 & \cdots \end{bmatrix}$$

e o processo continua até obter uma matriz triangular superior

$$R = Q_n \cdots Q_2 Q_1 A.$$

As matrizes Q_1, Q_2, ..., Q_n são todas unitárias, de modo que seu produto $Q_n \cdots Q_2\, Q_1$, que denotaremos por Q^*, também é unitária sendo Q sua inversa. Assim, $Q^* A = R$, que resulta na decomposição

$$A = QR$$

onde Q é unitária e R é triangular superior.

8.2 O algoritmo para calcular R

O algoritmo seguinte calcula a matriz R, m por n, triangular superior da decomposição QR de uma matriz A de ordem m por n, com $m \geq n$. Além de R, este algoritmo constrói os vetores de reflexão v_1, ..., v_n.

====================================

Algoritmo 10.1 Fatoração QR de Householder

Entrada: $A = (a_{ij})$, de ordem m por n.

Saída: Substitui a matriz A por R, que é triangular superior. Calcula ainda as reflexões v_1, ..., v_n.

====================================

for $k = 1$ to n
    $x = A_{k:m,k}$
    $v_k = x + sign(x_1) \|x\| e_1$
    $v_k = v_k / \|v_k\|$
    $A_{k:m,\,k:n} = A_{k:m,\,k:n} - 2v_k(v_k^* A_{k:m,\,k:n})$
=====================================

8.3 Contagem das operações

Em cada loop, o cálculo dominante é o de $A_{k:m,\,k:n}$. O vetor v_k tem comprimento $l = m - k + 1$. Para cada coluna de $A_{k:m,\,k:n}$, é preciso calcular o produto interno $v_k^* A_{k:m,\,j}$ o que exige $2l - 1$ flops, sendo l multiplicações e $l - 1$ adições. Calculado este produto interno, precisamos de 1 flop para calcular $2(v_k^* A_{k:m,\,j})$ e l multiplicações para calcular o produto deste escalar, $2(v_k^* A_{k:m,\,j})$ pelo vetor v_k. Finalmente, l subtrações para obter $A_{k:m,\,j} - 2v_k(v_k^* A_{k:m,\,j})$. Efetuamos assim $4l$ operações para calcular $A_{k:m,\,j}$ em cada loop. Para calcular as $n - k + 1$ colunas de $A_{k:m,\,k:n}$ serão necessários $4l(n - k + 1)$ flops e na execução dos n loops exigirá

$$\sum_{k=1}^{n} 4l(n-k+1) = \sum_{k=1}^{n} 4(m-k+1)(n-k+1) = 2mn^2 - \frac{2}{3}n^3 + 2mn + \frac{2}{3}n$$

flops. Usamos no cálculo deste somatório os seguintes resultados

$$\sum_{k=1}^{n} 1 = n, \quad \sum_{k=1}^{n} k = \frac{1}{2}n(n+1), \quad \sum_{k=1}^{n} k^2 = \frac{1}{6}(n + 3n^2 + 2n^3).$$

Admitindo que m e n crescem na mesma razão, fazendo m e $n \to \infty$, obtemos

$$\sim 2mn^2 - \frac{2}{3}n^3$$

flops para executar o algoritmo de Householder.

8.4 O algoritmo para calcular Q^*

A matriz Q ainda não foi calculada. Podemos usar as fórmulas

$$Q^* = Q_n \cdots Q_2 Q_1$$

ou

$$Q = Q_1^* Q_2^* \cdots Q_n^*$$

para obtê-la. Lembramos aqui que as matrizes Q_i são unitárias e assim $Q_i^* = Q_i^{-1}$. Podemos calcular Q^*b ou Qx. O próximo algoritmo calcula Q^*b.

====================================

Algoritmo 10.2 Fatoração QR de Householder
Entrada: As reflexões $v_1, \ldots, v_n$ de ordem m por 1 e b de ordem m por 1
Saída: Substitui b pelo vetor Q^*b de ordem m por 1.

====================================

for $k = 1$ to n
$\quad b_{k:m} = b_{k:m} - 2v_k(v_k^* b_{k:m})$

====================================

Podemos usar este algoritmo para obter Q^*, calculando suas colunas $Q^*e_1, \ldots, Q^*e_m$ de Q^*.

8.5 O algoritmo para calcular Q

O próximo algoritmo calcula Qx.

====================================

Algoritmo 10.3 Fatoração QR de Householder
Entrada: As reflexões $v_1, \ldots, v_n$, de ordem m por 1 e o vetor x de ordem n por 1.
Saída: Substitui x pelo vetor Qx de ordem n por 1.

====================================

for $k = n$ downto 1
$\quad x_{k:m} = x_{k:m} - 2v_k(v_k^* x_{k:m})$

====================================

Podemos usar este algoritmo para calcular as colunas $Qe_1, \ldots, Qe_m$ de Q. Este é o método de escolha quando se deseja calcular a matriz unitária Q.

Quando se deseja calcular apenas $\hat{Q}$ da decomposição reduzida, basta calcular as colunas $Qe_1, \ldots, Qe_n$.

Capítulo 9

Mínimos quadrados

Seja A uma matriz complexa m por n e b um vetor coluna do $\mathbf{C}^m$. O sistema algébrico $Ax = b$ tem solução se e somente se b for um ponto da imagem de A. Quando, além disso A for inversível, $Ax = b$ possui uma única solução dada por $x = A^{-1}b$.

Quando b não pertencer à imagem de A, o sistema algébrico $Ax = b$ não tem solução. Entretanto, podemos escolher c na imagem de A que minimiza $\|c - b\|^2$ e resolver o sistema $Ax = c$. O ponto c da imagem de A mais próximo de b é a projeção ortogonal de b sobre a imagem de A. Seja Pb esta projeção. As soluções de $Ax = Pb$, minimizam $\|Ax - b\|^2$.

Quando o $\ker(A) = \{0\}$, o sistema $Ax = Pb$ tem solução única. Entretanto, quando $\ker(A)$ contiver vetores não nulos, o sistema $Ax = Pb$ possuirá infinitas soluções. De fato, se n for um elemento não nulo do núcleo de A e x_1 uma solução de $Ax = Pb$, então, para todo número complexo c, $A(x_1 + cn) = Pb$.

Se x for uma solução de $Ax = Pb$, podemos decompô-la numa soma $x_1 = r + n$ onde r está na imagem de A^* e n está no núcleo de A, uma vez que $\mathbf{C}^n = \operatorname{Im}(A^*) \oplus \ker(A)$. Observe que r é a projeção ortogonal de x sobre a imagem de A^*. Sendo x_1 qualquer outra solução do sistema $Ax = Pb$, então $A(x_1 - x) = 0$ e $x_1 - x = n_1$ pertence ao núcleo de A e assim $x_1 = x + n_1 = r + (n + n_1)$, onde $n + n_1$ pertence ao núcleo de A e r é a projeção ortogonal sobre a imagem de A^* de uma solução qualquer do sistema $Ax = Pb$. Uma das lições que se tira do raciocínio acima é que projeção ortogonal sobre a imagem de A^* de uma solução x qualquer de $Ax = Pb$ sempre resulta no mesmo vetor r. Determinando uma única solução do sistema $Ax = Pb$ e todas as soluções do sistema homogêneo $Ax = 0$, teremos em mãos todas as soluções do sistema não homogêneo $Ax = Pb$.

Cada solução de $Ax = Pb$ é **uma solução por mínimos quadrados** de $Ax = b$. A projeção ortogonal r de uma solução do sistema $Ax = Pb$ sobre a imagem de A^* é chamada de **solução principal por mínimos quadrados** de $Ax = b$.

Quando o sistema linear $Ax = b$ tem solução, suas soluções coincidem com as soluções por mínimos quadrados de $Ax = b$.

Seguindo o esquema estabelecido até o momento, para obter a solução principal por mínimos quadrados de $Ax = b$, precisamos seguir a seguinte receita:

1. Determine Pb, a projeção ortogonal de b sobre a imagem de A.

2. Determine x, uma solução de $Ax = Pb$.

3. Determine r, a projeção ortogonal de x sobre a imagem de A^*, que é ortogonal ao núcleo de A.

Exercício 9.1 *Para as matrizes abaixo, determine: a projeção ortogonal Pb de b sobre a imagem de A, uma solução de $Ax = Pb$ e a projeção ortogonal r de x sobre a imagem de A^*.*

1. $A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 2 & 3 \end{pmatrix}$ *e* $b = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$.

2. $A = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & 0 \\ 3 & 1 & 1 \end{pmatrix}$ *e* $b = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.

Exercício 9.2 *Na decomposição QR de* $A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 2 \end{pmatrix}$, $Q = \begin{pmatrix} 1/\sqrt{2} & 0 \\ 0 & 1 \\ 1/\sqrt{2} & 0 \end{pmatrix}$. *Determine a matriz que projeta sobre a imagem de A. Resolva o problema de mínimos quadrados $Ax = (1,\ 0,\ 0)^T$. (Sugestão: note que as colunas de Q dão base ortogonal para a imagem de A.)*

A resolução de um problema de mínimos quadrados é bastante árdua. Felizmente existe um atalho que nos é dado pela proposição seguinte.

Teorema 9.3 *Seja A uma matriz m por n e b um vetor coluna m por 1. Seja P a matriz que projeta ortogonalmente sobre a imagem de A. Os sistemas lineares $Ax = Pb$ e $A^*Ax = A^*b$ possuem as mesmas soluções.*

Prova. Observe inicialmente que $Pb - b \in \operatorname{Im}(A)^{\perp} = \ker(A^*)$.

1. Se x for solução de $Ax = Pb$, então $Ax - b = Pb - b$ pertence ao núcleo de A^* de onde se conclui que $A^*(Ax - b) = 0$ ou

$$A^*Ax = A^*b.$$

2. Se x for solução de $A^*Ax = A^*b$, então $A^*(b - Ax) = 0$ e $b - Ax$ pertence ao núcleo de A^*, que é ortogonal à imagem de A. Desta forma, $b = Ax + (b - Ax)$ é uma decomposição de b em um vetor Ax na imagem de A e outro $b - Ax$ no seu complemento ortogonal. Logo, $Pb = Ax$. $\square$

Definição 9.4 *A equação*

$$A^*Ax = A^*b$$

*é chamada de **equação normal** do problema de mínimos quadrados para $Ax = b$.*

Embora seja redundante, nunca é demais reafirmar que, se x for uma solução da equação normal, então a projeção ortogonal de b sobre a imagem de A é igual a Ax

$$Pb = Ax.$$

Quando $m \geq n$ e a matriz A, de ordem m por n, possuir posto máximo n, suas colunas serão linearmente independentes e o seu núcleo conterá apenas o zero. A matriz A^*A será inversível e o problema de mínimos quadrados para $Ax = b$ terá uma única solução

$$x = (A^*A)^{-1}A^*b.$$

Como $Pb = Ax$, a projeção ortogonal P sobre a imagem de A será dada pelo produto

$$P = A(A^*A)^{-1}A^*.$$

9.1 Mínimos quadrados e a decomposição QR

Seja $A = \hat{Q}\hat{R}$ a decomposição QR reduzida da matriz A. As colunas da matriz $\hat{Q} = [q_1, \ldots, q_k]$ formam uma base ortonormal da imagem de A e $\hat{Q}^*\hat{Q}$ é a matriz identidade k por k. A matriz $\hat{R}$, de ordem k por n, é triangular superior,

$$A^*A = \hat{R}^*\hat{Q}^*\hat{Q}\hat{R} = \hat{R}^*\hat{R},$$

e a equação normal $A^*Ax = A^*b$ toma a forma

$$\hat{R}^*\hat{R}x = \hat{R}^*\hat{Q}^*b.$$

Este sistema pode ser resolvido com facilidade, posto que $\hat{R}$ é triangular superior e $\hat{R}^*$ é triangular inferior.

Quando $m \geq n$ e a matriz A de ordem m por n possuir posto máximo n, as matrizes A^*A e $\hat{R}$ são inversíveis o que permite reduzir a equação normal à forma

$$\hat{R}x = \hat{Q}^*b.$$

Neste caso, $\hat{R}$ é inversível e esta equação tem solução única. A projeção $P = A\ (A^*A)^{-1}$ A^* assumirá uma forma mais simples

$$P = A(A^*A)^{-1}A^* = \hat{Q}\hat{R}(\hat{R}^*\hat{R})^{-1}\hat{R}^*\hat{Q}^* = \hat{Q}\hat{Q}^*.$$

9.2 Pseudo inversa

Quando $m \geq n$ e a matriz A de ordem m por n possuir posto máximo n, o problema de mínimo quadrado para $Ax = b$ tem uma única solução que coincide com a única solução da equação normal

$$A^*Ax = A^*b.$$

A matriz A^*A é inversível neste caso e

$$x = (A^*A)^{-1}A^*b.$$

A matriz n por m

$$A^+ = (A^*A)^{-1}A^*,$$

é chamada de **pseudo-inversa** de A. Ela recebe este nome pois $x = A^+b$ é a solução por mínimos quadrados de $Ax = b$.

Sendo $\hat{Q}\hat{R}$ for a decomposição QR reduzida de A, então

$$x = \hat{R}^{-1}\hat{Q}^*b$$

e a pseudo-inversa será dada por

$$A^+ = \hat{R}^{-1}\hat{Q}^*.$$

Continuando com A de ordem m por n e posto máximo n, a matriz A^*A, além de ser inversível, é hermitiana e possui uma decomposição de Cholesky $A^*A = R^*R$, onde R é inversível e triangular superior. Neste caso, a equação normal se reduz à forma

$$R^*Rx = A^*b$$

e, sendo R triangular superior e R^* triangular inferior, é muito simples resolver o sistema.

9.3 Reta de regressão

Sejam $(x_1,\, y_1),\, \ldots,\, (x_m,\, y_m)$ pontos de $\mathbf{C}^2$. Determine a reta

$$y = c_0 + c_1 x$$

que minimiza o resíduo $\sum_{i=1}^m (c_0 + c_1 x_i - y_i)^2$.

Consideremos em $\mathbf{C}^m$ o produto interno definido por $\langle x, y\rangle = x^*y$ e a norma definida por $\|x\|^2 = \sqrt{x^*x}$. Sendo

$$A = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix}, \quad c = \begin{pmatrix} c_0 \\ c_1 \end{pmatrix}, \quad d = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix},$$

então

$$\|Ac - d\|^2 = \langle Ac - d,\ Ac - d\rangle = \sum_{i=1}^m (c_0 + c_1 x_i - y_i)^2.$$

Portanto, o problema proposto é equivalente ao problema de mínimos quadrados para $Ac = d$ no produto interno de $\mathbf{C}^m$ definido por $\langle x, y\rangle = x^*y$.

Se $x_1, \ldots, x_m$ não forem todos iguais, a matriz A tem posto 2. A equação normal $A^*Ac = A^*d$ do problema de mínimos quadrados para $Ac = d$ é

$$\begin{pmatrix} m & \sum_{i=1}^m x_i \\ \sum_{i=1}^m x_i & \sum_{i=1}^m x_i^2 \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^m y_i \\ \sum_{i=1}^m x_i y_i \end{pmatrix}$$

que terá solução única.

Sejam

$$x = \frac{1}{m}(x_1 + \cdots + x_m) \quad \text{e} \quad y = \frac{1}{m}(y_1 + \cdots + y_m)$$

os valores médios de $x_1, \ldots, x_m$ e $y_1, \ldots, y_m$, respectivamente. Podemos fazer a decomposição $A = WT$, onde

$$W = \begin{pmatrix} 1 & (x_1 - x) \\ \vdots & \vdots \\ 1 & (x_m - x) \end{pmatrix}, \qquad T = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$$

onde as colunas de W formam uma base ortogonal para a imagem de A e T é inversível. Com esta decomposição, a equação normal $A^*Ac = A^*d$ assume a forma

$$\begin{pmatrix} m & 0 \\ 0 & \sum_{i=1}^m (x_i - x)^2 \end{pmatrix} \begin{pmatrix} c_0 + c_1 x \\ c_1 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^m y_i \\ \sum_{i=1}^m (x_i - x) y_i \end{pmatrix}$$

cuja solução é imediata

$$c_1 = \frac{\sum_{i=1}^m (x_i - x) y_i}{\sum_{i=1}^m (x_i - x)^2} \qquad \text{e} \qquad c_0 = y - c_1 x.$$

Exercício 9.5 *Calcule a reta de regressão para os dados* $(1; 3)$, $(2; 6)$, $(3; 5,5)$, $(4; 6,5)$.

9.4 Interpolação polinomial

Dados os pontos $(x_1,\ y_1), \ldots, (x_m,\ y_m)$, determine $c_0, c_1, \ldots, c_{m-1}$, de modo que o polinômio

$$y = p(x) = c_0 + c_1 x + \cdots + c_{m-1} x^{m-1}$$

seja tal que $p(x_i) = y_i$, para $i = 1, 2, \ldots, m-1$.

Estas condições nos levam ao sistema de equações algébricas lineares $Ac = d$ onde

$$A = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^k \\ 1 & x_2 & x_2^2 & \cdots & x_2^k \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^k \end{pmatrix}, \quad c = \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_k \end{pmatrix}, \quad d = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}.$$

Se os pontos $x_1, x_2, \ldots, x_m$ forem todos distintos, a matriz A é inversível e o problema tem solução única para qualquer $y_1, y_2, \ldots, y_m$.

O polinômio $p(x)$ assim obtido é chamado de polinômio interpolador dos pares de pontos $(x_1, y_1), \ldots, (x_m, y_m)$.

À medida que o m cresce, o polinômio p oscila mais e mais. Se os pontos (x_i, y_i) tiverem sido obtidos a partir de experimentos laboratoriais, $p(x)$ pode não representar o fenômeno que se pretende descrever.

9.5 Ajuste polinomial

Se os dados (x_i, y_i), $i = 1, 2, \ldots, m$, forem provenientes de experimentos, a interpolação polinomial fornece um polinômio que oscila muito e não representa adequadamente os dados obtidos experimentalmente. Quanto maior for o conjunto de dados, mais oscilante é o polinômio. Desta forma, é muito mais interessante procurar um polinômio de grau menor, que oscila menos e se ajuste melhor aos dados observados, embora não passe necessariamente por nenhum deles. Observando que dados experimentais são passíveis de erros, não é um absurdo obter um polinômio que se ajuste a eles sem passar necessariamente por nenhum deles. Fica no ar a pergunta: qual o grau do polinômio a ser usado. Em geral, quando se faz um experimento, existem modelos matemáticos que descrevem o fenômeno. Se não houver, busca-se um por tentativas e erros. O critério de ajuste pode ser aquele fornecido pelo problema de mínimos quadrados.

Dados os pontos $(x_1, y_1), \ldots, (x_m, y_m)$, determine o polinômio

$$y = p(x) = c_0 + c_1 x + \cdots + c_k x^k$$

de grau k menor do que $m - 1$ que minimiza o resíduo $\sum_{i=1}^{m} (p(x_i) - y_i)^2$.

Minimizar este resíduo é equivalente à minimização da norma $\|Ac - d\|^2$, onde

$$A = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^k \\ 1 & x_2 & x_2^2 & \cdots & x_2^k \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^k \end{pmatrix}, \quad c = \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_k \end{pmatrix}, \quad d = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix},$$

de modo que o problema proposto é equivalente àquele dos mínimos quadrados para $Ac = d$ com o produto interno

$$\langle x, y \rangle = x^* y.$$

Quando $k = m - 1$ caímos no caso anterior da interpolação polinomial.

9.6 Aproximação polinomial de funções

Um problema semelhante ao anterior consiste em determinar o polinômio

$$y = p(x) = c_0 + c_1 x + \cdots + c_k x^k$$

de grau menor ou igual a k que melhor aproxima uma função $g(x)$ definida no intervalo $[a, b]$. Logo surge a pergunta: em que sentido $p(x)$ é o melhor polinômio que aproxima $g(x)$?

Numa tentativa de responder a esta indagação, legítima por sinal, poderíamos pegar m pontos $a = x_1 \leq x_2 \leq \cdots \leq x_m = b$ igualmente espaçados em $[a, b]$ e assim determinar o polinômio $p(x)$ que minimiza a soma

$$S = \sum_{i=1}^{m} \left[p(x_i) - g(x_i)\right]^2 .$$

Denotando $g(x_i)$ por y_i, podemos resolver este problema usando a técnica anterior de ajuste polinomial.

Entretanto, antes de encerrar o caso, vamos elaborar um pouco mais este problema. Para m fixo, seja $\Delta x = (b - a)/m$. Minimizar S ou $S \cdot \Delta x$ é a mesma coisa. À medida que o m cresce,

$$S \cdot \Delta x = \sum_{i=1}^{m} \left|p(x_i) - g(x_i)\right|^2 \Delta x$$

converge para a integral $\int_a^b \left[f(x) - g(x)\right]^2 dx$. Tal fato motiva a definição do produto interno

$$\langle f, g \rangle = \int_a^b f(x)\bar{g}(x)\, dx$$

e a norma correspondente a ele

$$\|f\| = \sqrt{\int_a^b \left|f(x)\right|^2 dx}.$$

Nesta norma, $\|p - g\|^2 = \int_a^b \left|p(x) - g(x)\right|^2 dx$.

O problema agora pode ser reformulado como segue: Seja g uma função contínua, definida no intervalo real $[a, b]$. Determine o polinômio $p(x) = c_0 + c_1 x + \cdots + c_k x^k$ de grau menor ou igual a k, que minimiza

$$\|p - g\|^2 = \int_a^b \left|p(x) - g(x)\right|^2 dx.$$

Sabemos que p é obtido projetando g ortogonalmente sobre o espaço dos polinômios de grau menor ou igual a k. Podemos determinar este polinômio a partir de uma base ortogonal para este espaço vetorial. Sabemos que quando $[a, b] = [-1, 1]$, estes polinômios estão relacionados aos polinômios de Legendre.

9.7 Aproximação trigonométrica

Podemos agora resolver um problema semelhante ao anterior, aproximando uma função real $g(x)$, definida no intervalo $[-L, L]$, por uma função trigonomérica

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{m} a_k \cos\left(\frac{k\pi}{L}x\right) + \sum_{k=1}^{m} b_k \text{sen}\left(\frac{k\pi}{L}x\right)$$

fazendo com que

$$\|f - g\|^2 = \int_{-L}^{L} |f(x) - g(x)|^2 \, dx$$

seja o menor possível. A norma acima é proveniente do produto interno

$$\langle f, g \rangle = \int_{-L}^{L} f(x)\bar{g}(x) \, dx$$

e, o que é interessante observar, o conjunto de funções

$$\left\{ \ 1, \ \cos\left(\frac{k\pi}{L}x\right), \ \text{sen}\left(\frac{k\pi}{L}x\right) : k = 1, 2, \ldots, m \ \right\}$$

é ortogonal em relação a este produto interno. Pode-se calcular que $\langle 1, 1 \rangle = 2L$ e

$$\left\langle \cos\left(\frac{k\pi}{L}x\right), \ \cos\left(\frac{k\pi}{L}x\right) \right\rangle = \left\langle \text{sen}\left(\frac{k\pi}{L}x\right), \ \text{sen}\left(\frac{k\pi}{L}x\right) \right\rangle = L.$$

Consequentemente,

$$\begin{aligned} a_0 &= \frac{1}{L} \int_{-L}^{L} g(x) dx \\ a_k &= \frac{1}{L} \int_{-L}^{L} g(x) \cos\left(\frac{k\pi}{L}x\right) dx \\ b_k &= \frac{1}{L} \int_{-L}^{L} g(x) \text{sen}\left(\frac{k\pi}{L}x\right) dx \end{aligned}$$

para $k = 1, 2, \ldots, m$.

Seja $g(x)$ uma função contínua por partes no intervalo $[-L, L]$ e $h(x)$ sua extensão periódica para toda a reta. À medida que o m cresce, a aproximação trigonométrica converge para o valor médio

$$\frac{g(x^-) + g(x^+)}{2}$$

onde

$$g(x^-) = \lim_{s \to x^-} g(s) \quad \text{e} \quad g(x^+) = \lim_{s \to x^+} g(s).$$

A teoria que trata das aproximações por funções trigonométricas é denominada de Análise de Fourier e a ciência que estuda as aproximações por seqüências de funções ortogonais é denominada de Análise Harmônica.

Capítulo 10

Autovalores e autovetores

Definição 10.1 *Seja V um espaço vetorial e $L : V \to V$ um operador linear. Um escalar λ é um **autovalor** de L se existir um vetor v, não nulo, tal que $Lv = \lambda v$. O vetor v é chamado de **autovetor** de L correspondente ao autovalor λ.*

Uma matriz quadrada A define um operador linear de $\mathbf{R}^n$ em $\mathbf{R}^n$ se for real ou de $\mathbf{C}^n$ em $\mathbf{C}^n$ se for complexa. Um número real λ é um **autovalor** de A se existir uma matriz coluna x de ordem n por 1, não nula, tal que $Ax = \lambda x$. O vetor coluna x é chamado de **autovetor** de A correspondente ao autovalor λ.

Sendo x um autovetor de A correspondente ao autovalor λ então

$$(\lambda I - A)x = 0,$$

onde I é a matriz identidade. A equação matricial acima possui solução não trivial se e só se

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

Se A for uma matriz de ordem n, o $\det(\lambda I - A)$ é um polinômio de grau n em λ. Sobre o corpo dos números complexos, a equação polinomial

$$\det(tI - A) = 0$$

possui pelo menos uma raiz λ. Substituindo este valor na equação matricial $(\lambda I - A)x = 0$, determinamos os autovalores x. Lembre-se: quando uma equação matricial homogênea possui solução não trivial, esta solução não é única. Se x_1 e x_2 são dois autovetores de A correspondentes ao mesmo autovalor λ e α_1, α_2 forem dois escalares, então $\alpha_1 x_1 + \alpha_2 x_2$ será autovetor de A correspondentes ao autovalor λ. O conjunto

$$auto(\lambda) = \{x \in \mathbf{C}^n : Ax = \lambda x\}$$

é um subespaço vetorial de $\mathbf{C}^n$, chamado de **autoespaço** de A correspodente ao autovalor λ.

A matriz $tI - A$ é chamada de **matriz característica** de A, o polinômio $\Delta(t) = \det(tI - A)$ é chamado de **polinômio característico** de A e a equação polinomial $\det(tI - A) = 0$ é chamada de **equação característica** de A.

Para obter os autovalores e autovetores de A, calcule as raízes $\lambda_1, \ldots, \lambda_s$ da equação característica $\det(\lambda I - A) = 0$ e, para cada autovalor λ_i, determine o conjunto solução do sistema homogêneo

$$(\lambda I - A)x = 0$$

para obter o autoespaço de λ.

Se duas matrizes quadradas A e B forem semelhantes, então existe uma matriz inversível P tal que $B = PAP^{-1}$ e

$$\begin{aligned} \det(tI - B) &= \det(tP^{-1}P - P^{-1}AP) = \det(P^{-1}(tI - A)P) = \\ &= \det(P^{-1})\det(tI - A)\det(P) = \det(tI - A), \end{aligned}$$

mostrando que matrizes semelhantes possuem a mesma equação característica e, portanto, os mesmos autovalores.

Para determinar autovalores e autovetores de um operador linear $L : V \to V$, onde V é um espaço vetorial de dimensão finita n, escolhemos uma base $\mathcal{B} = \{v_1, \ldots, v_n\}$ de V e calculamos a matriz de L na base $\mathcal{B}$ que denotamos por A. Um escalar λ é autovalor de L se e só se for autovalor de A. Um vetor v não nulo é autovetor de L correspondente ao autovalor λ se e só se a matriz coluna x das coordenadas de v na base $\mathcal{B}$ for um autovetor de A correspondente ao autovalor λ.

De fato, se $A = (a_{ij})$, então

$$Lv_j = \sum_{i=1}^{n} a_{ij} v_i$$

e, se $x = (x_1, \ldots, x_n)^T$, então

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

e

$$Lv = \sum_{j=1}^{n} x_j L v_j = \sum_{j=1}^{n} x_j \sum_{i=1}^{n} a_{ij} v_i = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} a_{ij} x_j \right) v_i.$$

Assim, $Lv = \lambda v$ se e só se $\sum_{i=1}^{n} \left(\sum_{j=1}^{n} a_{ij} x_j \right) v_i = \sum_{i=1}^{n} \lambda x_i v_i$ e, da independência linear dos elementos de $\mathcal{B}$, esta igualdade se verifica se e só se $\sum_{j=1}^{n} a_{ij} x_j = \sum_{i=1}^{n} \lambda x_i$, para $i = 1, \ldots, n$, que corresponde à equação matricial $Ax = \lambda x$.

Desta forma, para calcular os autovalores e autovetores de um operador L num espaço vetorial de dimensão finita, basta calcular sua matriz A numa base $\mathcal{B}$, determinar seus autovalores $\lambda_1, \ldots, \lambda_s$, que serão os autovetores de L. A cada autovalor λ_i, determine o autoespaço de A correspondente a este autovalor

$$auto(\lambda_i) = \{x \in \mathbf{C}^n : Ax = \lambda_i x\}$$

para obter os autovetores v de L correspondente ao autovalor λ_i que serão dados por

$$v = x_1 v_1 + \cdots + x_n v_n$$

onde $x = (x_1, \ldots, x_n)^T$ é autovetor de A correspondente ao autovalor λ_i.

Seja λ um autovalor de L. Tal como no caso de matrizes, o conjunto

$$auto(\lambda) = \{v \in V : Lv = \lambda v\}$$

é um subespaço vetorial de V, denominado de **autoespaço** de L correspondente ao autovalor λ.

Se A for a matriz de L numa base $\mathcal{B}$ de V, então $tI - A$ é chamada de **matriz característica** de L, o polinômio $\det(tI - A)$ é chamado de **polinômio característico** de L e a equação polinomial $\det(tI - A) = 0$ é chamada de **equação característica** de L.

Como as matrizes de uma transformação linear em bases diferentes são semelhantes, o polinômio característico de L não depende da base escolhida e, portanto, seus autovalores não dependem da base escolhida. A mesmo acontece com os autovetores de L. Sua determinação não depende da base escolhida.

Teorema 10.2 *Autovetores de L correspondentes a autovalores distintos são linearmente independentes.*

Prova. Sejam $v_1, \ldots, v_r$ os autovetores de L correspondentes aos autovalores distintos $\lambda_1, \ldots, \lambda_r$. Vamos provar que estes autovetores formam um conjunto linearmente independente.

1. Inicialmente provaremos que $\{v_1, v_2\}$ é linearmente independente. Sejam a_1 e a_2 dois escalares tais que $a_1 v_1 + a_2 v_2 = 0$. Multiplicando por λ_1 e por A vem

$$a_1 \lambda_1 v_1 + a_2 \lambda_1 v_2 = 0$$

e

$$a_1 \lambda_1 v_1 + a_2 \lambda_2 v_2 = 0.$$

Subtraindo uma da outra chega-se a $a_2(\lambda_1 - \lambda_2)v_2 = 0$. Como $\lambda_1 \neq \lambda_2$ e $v_2 \neq 0$, obtemos $a_2 = 0$ e, consequentemente, $a_1 = 0$, provando que o conjunto $\{v_1, v_2\}$ é linearmente independente. Do mesmo modo se prova que um conjunto formado por dois autovetores $\{v_i, v_j\}$ são linearmente independentes.

2. Vamos supor, como hipótese de indução, que qualquer subconjunto de $\{v_1, \ldots, v_r\}$ com menos de r elementos é linearmente independente.

3. Vamos provar que $\{v_1, \ldots, v_r\}$ é linearmente independente. Consideremos a equação

$$a_1 v_1 + a_2 v_2 + \cdots + a_r v_r = 0,$$

onde a_i são escalares. Multiplicando-a por A e por λ_1, obtemos

$$a_1 \lambda_1 v_1 + a_2 \lambda_2 v_2 + \cdots + a_r \lambda_r v_r = 0$$

e

$$a_1 \lambda_1 v_1 + a_2 \lambda_1 v_2 + \cdots + a_r \lambda_1 v_r = 0.$$

Subtraindo uma da outra vem

$$a_2(\lambda_2 - \lambda_1)v_2 + \cdots + a_r(\lambda_r - \lambda_1)v_r = 0$$

Sendo o conjunto $\{v_2, \ldots, v_r\}$ linearmente independente, $a_2(\lambda_2 - \lambda_1) = \cdots = a_r(\lambda_r - \lambda_1) = 0$. Como os autovalores são distintos, $a_2 = \cdots = a_r = 0$ e, em conseqüência, a_1 também é nulo, completando a prova de que o conjunto $\{v_1, \ldots, v_r\}$ é linearmente independente. $\square$

Teorema 10.3 *Seja V um espaço com produto interno e $L : V \to V$ auto-adjunto. Os autovalores de L são reais e os autovetores correspondentes a autovalores distintos são ortogonais.*

Prova. Se $Lv = \lambda v$, onde v é um vetor não nulo e λ escalar, então $\langle Lv, v\rangle = \langle v, Lv\rangle$ o que implica em $\bar{\lambda}\langle v, v\rangle = \lambda\langle v, v\rangle$ ou $(\bar{\lambda} - \lambda)\langle v, v\rangle = 0$, de onde se conclui que $\bar{\lambda} = \lambda$.

Se v e w forem autovetores correspondentes aos autovalores reais distintos λ e μ, então, de $\langle Lv, w\rangle = \langle v, Lw\rangle$ o que implica em $\lambda\langle v, w\rangle = \mu\langle v, w\rangle$ ou $(\lambda - \mu)\langle v, w\rangle = 0$. Como $\lambda \neq \mu$, $\langle v, w\rangle = 0$. $\square$

Teorema 10.4 *Seja V um espaço com produto interno e $L : V \to V$ linear. Os autovalores de L^*L são reais e não negativos, isto é, se λ for autovalor de L^*L, então $\lambda \geq 0$.*

Prova. Como L^*L é auto-adjunto, seus autovalores são reais. Seja v um autovetor de L^*L associado ao autovalor λ. De $\langle Lv, Lv\rangle \geq 0$, segue

$$0 \leq \langle Lv, Lv\rangle = \langle v, L^*\, Lv\rangle = \langle v, \lambda v\rangle = \lambda\langle v, v\rangle$$

Como $\langle v, v\rangle > 0$, conclui-se que $\lambda \geq 0$. $\square$

Teorema 10.5 *Seja V um espaço com produto interno e $L : V \to V$ antiadjunto. Os autovalores de L são números imaginários puros (números complexos com parte real nula) e os autovetores correspondentes a autovalores distintos são ortogonais.*

Prova. Se v for um autovetor de L, com $Lv = \lambda v$ então

$$\langle Lv, v\rangle = \langle v, -Lv\rangle \Longrightarrow \bar{\lambda}\langle v, v\rangle = -\lambda\langle v, v\rangle \Longrightarrow \bar{\lambda} = -\lambda$$

provando que λ é um número imaginário com parte real nula.

Sejam v e w autovetores de L com $Lv = \lambda v$, $Lw = \sigma w$ e $\lambda \neq \sigma$. Então

$$\langle Lv, w\rangle = \langle v, -Lw\rangle \Longrightarrow \bar{\lambda}\langle v, w\rangle = -\sigma\langle v, w\rangle .$$

Sendo λ e σ imaginários puros e distintos, $\bar{\lambda} \neq -\sigma$ e assim, $\langle v, w\rangle = 0$. $\square$

Teorema 10.6 *Seja V um espaço com produto interno e $L : V \to V$ unitário. Os autovalores de L são números complexos com módulo unitário e os autovetores correspondentes a autovalores distintos são ortogonais.*

Prova. Seja v um autovetor de L, com $Lv = \lambda v$. Então

$$\langle Lv, Lv\rangle = \langle v, v\rangle \Longrightarrow \bar{\lambda}\lambda \langle v, v\rangle = \langle v, v\rangle \Longrightarrow \bar{\lambda}\lambda = 1,$$

mostrando que λ é um número complexo de módulo unitário.

Sejam v e w autovetores de L com $Lv = \sigma v$, $Lw = \lambda w$ e $\sigma \neq \lambda$. Se $\langle v, w\rangle \neq 0$, então

$$\langle Lv, Lw\rangle = \langle v, w\rangle \Longrightarrow \bar{\sigma}\lambda \langle v, w\rangle = \langle v, w\rangle \Longrightarrow \bar{\sigma}\lambda = 1.$$

Como λ possui módulo unitário, podemos escrevê-lo na forma $\lambda = \exp(i\theta)$, onde θ é um número real. Assim $\bar{\sigma}\exp(i\theta) = 1$ e $\bar{\sigma} = \exp(-i\theta) = \bar{\lambda}$ de onde se conclui que $\lambda = \sigma$, o que contraria a hipótese. $\square$

Definição 10.7 *Seja $p(t) = c_0 + c_1 t + \cdots + c_k t^k$ um polinômio em t e A uma matriz quadrada. Define-se $p(A)$,* ***o valor do polinômio*** *$p(t)$ na* ***matriz*** *A, por*

$$p(A) = c_0 I + c_1 A + \cdots + c_k A^k$$

onde I é a matriz identidade com ordem igual à de A. Se $p(A) = 0$, diremos que a matriz A é um ***zero do polinômio*** *$p(t)$.*

Sejam A_k, $k = 0, 1, \ldots, r$, matrizes quadradas de mesma ordem. Os elementos da matriz

$$A(t) = A_0 + A_1 t + \cdots + A_r t^r$$

são polinômios de grau menor ou igual a r.

Teorema 10.8 *(Cayley-Hamilton) Toda matriz é um zero do seu polinômio característico.*

Prova. Seja A uma matriz quadrada de ordem n e $B = tI - A$ sua matriz característica. Seu polinômio característico

$$\det B = \det(tI - A) = t^n + c_{n-1}t^{n-1} + \cdots + c_1 t + c_0$$

tem grau n. Cada elemento da adjunta clássica de B, denotada por $adj(B)$, é um polinômio de grau $n-1$ ou inferior. Logo, $adj(B) = B_{n-1}t^{n-1} + \cdots + B_1 t + B_0$ é um polinômio matricial de grau $n-1$ ou inferior. Sabe-se que $B^{-1} = adj(B)/\det(B)$ e assim,

$$\det(B)I = adj(B) \cdot B.$$

Como

$$\det(B)I = It^n + c_{n-1}It^{n-1} + \cdots + c_1 It + c_0 I$$

e

$$\begin{aligned} adj(B)\cdot B &= \left(B_{n-1}t^{n-1}+\cdots+B_1t+B_0\right)(tI-A) \\ &= B_{n-1}t^n+(B_{n-2}-B_{n-1}A)t^{n-1}+\cdots+(B_0-B_1A)t+B_0A \end{aligned}$$

segue

$$\begin{aligned} I &= B_{n-1} \\ c_{n-1}I &= B_{n-2}-B_{n-1} \\ &\cdots \\ c_1I &= B_0-B_1A \\ c_0I &= B_0A. \end{aligned}$$

Multiplicando as igualdades, da primeira até a última por A^n, A^{n-1}, ..., A e I, respectivamente, pela direita e adicionando os resultados, obtemos zero do lado direito e, do lado esquerdo, o polinômio característico

$$A^n+c_{n-1}A^{n-1}+\cdots+c_1A+c_0I$$

calculado na matriz A. Isto prova o teorema de Cayley-Hamilton. □

Matrizes triangulares em bloco são aquelas do tipo

$$A=\begin{pmatrix} A_1 & B \\ 0 & A_2 \end{pmatrix}$$

onde A_1 e A_2 são matrizes quadradas, podendo ter a mesma ordem ou não. Então

$$\det(A)=\det(A_1)\det(A_2).$$

A matriz característica de A também é triangular por blocos e

$$\det(tI-A)=\det(tI-A_1)\det(tI-A_2).$$

Teorema 10.9 *Seja A uma matriz triangular em blocos, cujos blocos diagonais são A_1, ..., A_r. Então o polinômio característico de A é o produto dos polinômios característicos de A_1, ..., A_r, isto é,*

$$\det(tI-A)=\det(tI-A_1)\cdots\det(tI-A_r).$$

O polinômio característico de uma matriz quadrada complexa A pode ser fatorado em fatores lineares,

$$\Delta(t)=(t-\lambda_1)^{p_1}\cdots(t-\lambda_k)^{p_k},$$

onde λ_1, ..., λ_k são suas raízes e p_1, ..., p_k suas multiplicidades. O expoente p_k é a multiplicidade algébrica do autovalor λ_k.

Definição 10.10 *Seja λ um autovalor de uma matriz quadrada A de ordem n. A* ***multiplicidade algébrica*** *de λ é a multiplicidade de λ como raíz da equação característica. A* ***multiplicade geométrica*** *de λ é a dimensão do seu autoespaço.*

Exemplo 10.11 *O polinômio característico de* $A = \begin{pmatrix} 5 & 1 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{pmatrix}$ *é $(t-5)^3$. Logo, 5 é um autovalor de A de multiplicidade algébrica 3. O autoespaço deste autovalor é gerado por $(1,0,0)^T$ e $(0,0,1)^T$. Logo, a multiplicidade geométrica do autovalor 5 é 2.*

Teorema 10.12 *A multiplicidade geométrica de um autovalor nunca excede sua multiplicidade algébrica.*

Prova. Seja V um espaço vetorial de dimensão n e $L : V \to V$ um operador linear. Seja g a multiplicidade geométrica de um autovalor λ de L. Existem g autovetores $v_1, \ldots,$ v_g linearmente independentes correspondentes ao autovalor λ. Completemos este conjunto para obter uma base de V. Seja $B = \{v_1, \ldots, v_g, w_1, \ldots, w_r\}$ esta base. A matriz de L nesta base é

$$M = \begin{pmatrix} \lambda & \cdots & 0 & c_{11} & \cdots & c_{1r} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda & c_{p1} & \cdots & c_{pr} \\ 0 & \cdots & 0 & b_{11} & \cdots & b_{1r} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & b_{r1} & \cdots & b_{rr} \end{pmatrix} = \begin{pmatrix} A & C \\ 0 & B \end{pmatrix}$$

onde $A = \lambda I$. O polinômio característico de M é $\Delta(t) = \det(tI - M) = \det(tI - A)\det(tI - B) = (t-\lambda)^g \det(tI - B)$. Portanto, $(t-\lambda)^g$ deve dividir $\Delta(t)$. Para que isto ocorra, a multiplicidade algébrica de λ deve ser maior ou igual a g. $\square$

Teorema 10.13 *Uma matriz quadrada A de ordem n é semelhante a uma matriz diagonal D se e só se A tem n autovetores linearmente independentes.*

Prova. 1. Se A for semelhante a uma matriz diagonal D, então existe uma matriz inversível P tal que $AP = PD$. Os elementos diagonais de D são os autovalores de A e as colunas de P os autovetores. Sendo inversível, as colunas de P são vetores linearmente independentes.

2. Se A tem n autovetores linearmente independentes, sejam eles $\{v_1, \ldots, v_n\}$ que formam uma base de V e para os quais $Av_i = \lambda_i v_i$. Se P for a matriz cujas colunas são formadas por esses vetores, temos $AP = PD$ onde $D = diag(\lambda_1, \ldots, \lambda_n)$. $\square$

Nota 10.14 *No teorema anterior, os n autovalores $\lambda_1, \ldots, \lambda_n$ não precisam ser todos distintos.*

Seja P a matriz inversível para a qual $D = P^{-1}AP$ é diagonal. Então $A = PDP^{-1}$ é a chamada de **fatoração diagonal** de A.

Seja V um espaço vetorial com dimensão n. Seja $L : V \to V$ uma transformação linear, cujo polinômio característico

$$\Delta(t) = (t - \lambda_1)(t - \lambda_2) \cdots (t - \lambda_n)$$

pode ser fatorado em n fatores distintos do primeito grau. Então L possui n autovetores linearmente independentes e portanto, possui uma representação matricial diagonal, na base formada pelos autovetores. Os elementos da diagonal desta representação são os autovalores λ_i.

Se D for diagonal e seus elementos diagonais forem d_{11}, d_{22}, $\ldots$, d_{nn} denotaremos esta matriz por $D = diag(d_{11},\, d_{22},\, \ldots,\, d_{nn})$. Sendo $A = PDP^{-1}$, então

$$A^m = \left(PDP^{-1}\right)^m = PD^mP^{-1} = Pdiag(d_{11}^m, d_{22}^m, \ldots, d_{nn}^m)P^{-1}.$$

Sendo $f(t)$ um polinômio,

$$f(A) = f\left(P\,D\,P^{-1}\right) = P\,f(D)\,P^{-1} = Pdiag(\ f(d_{11}),\ f(d_{22}),\ \ldots,\ f(d_{nn})\ )P^{-1}.$$

Além disso, se os elementos diagonais de D forem não negativos, então

$$B = Pdiag\left(\sqrt{k_1}, \ldots, \sqrt{k_n}\right) P^{-1}$$

é uma **raiz quadrada** de A pois $B^2 = A$.

Capítulo 11

Espaços Invariantes

Definição 11.1 *Seja $L : V \to V$ um operador linear e W um subespaço de V. Se $L(W) \subset W$, diremos que W é* ***invariante*** *sob L.*

Exemplo 11.2 *Seja $L(x, y, z) = (\ x - y,\ x + y,\ z\ )$ um operador definido no espaço vetorial dos ternos ordenados. O subespaço $W = \{\ (x,\ y,\ 0) : x,\ y \in \mathbf{R}\ \}$ é invariante sob L.*

O subespaço gerado por um vetor w não nulo é invariante sob L se e só se w for um autovetor de L.

Seja $L : V \to V$ linear e $f(t)$ um polinômio. O $\ker f(L)$ é invariante sob L. Se W for invariante sob L, então W é invariante sob $f(L)$.

Definição 11.3 *Seja W um subespaço vetorial de V e $L : V \to V$ um operador linear. Se W for invariante sob L podemos definir $T : W \to W$ por $T(w) = L(w)$ para todo w em W. O operador T é linear em W e recebe o nome de* ***restrição*** *de L em W.*

Sendo W invariante sob L, então é invariante sob L^k, para todo k inteiro positivo. Se $f(t)$ for um polinômio, então W é invariante sob $f(L)$. Sendo T a restrição de L a W, para todo w em W, tem-se $f(T)w = f(L)w$.

Teorema 11.4 *Seja $L : V \to V$ linear e W subespaço de V invariante sob L. Então L possui uma representação matricial em bloco*

$$\begin{pmatrix} A & C \\ 0 & B \end{pmatrix}$$

onde A é uma representação matricial da restrição de L em W.

Prova. Seja $\{u_1, \ldots, u_j\}$ uma base de W e $\{u_1, \ldots, u_j, v_1, \ldots, v_k\}$ uma base de V. Como W é invariante frente a L,

$$\begin{aligned}
L(u_1) &= a_{11}u_1 + \cdots + a_{j1}u_j \\
&\cdots \\
L(u_j) &= a_{1j}u_1 + \cdots + a_{jj}u_j \\
L(v_1) &= c_{11}u_1 + \cdots + c_{j1}u_j + b_{11}v_1 + \cdots + b_{k1}v_k \\
&\cdots \\
L(v_k) &= c_{1k}u_1 + \cdots + c_{jk}u_j + b_{1k}v_1 + \cdots + b_{kk}v_k
\end{aligned}$$

e, portanto, a representação matricial de L nesta base é

$$\begin{pmatrix}
a_{11} & \cdots & a_{1j} & c_{11} & \cdots & c_{1k} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
a_{j1} & \cdots & a_{jj} & c_{j1} & \cdots & c_{jk} \\
0 & \cdots & 0 & b_{11} & \cdots & a_{1j} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & b_{k1} & \cdots & b_{kk}
\end{pmatrix} = \begin{pmatrix} A & C \\ 0 & B \end{pmatrix}.$$

□

Definição 11.5 *Seja $L : V \to V$ linear e $V = W_1 \oplus \cdots \oplus W_r$, onde W_i é invariante sob L. Seja L_i a restrição de L a W_i. Neste caso se diz que L é a* ***soma direta*** *dos L_i ou que L é* ***decomponível*** *nos operadores L_i, quando então se escreve*

$$L = L_1 \oplus \cdots \oplus L_r.$$

Ainda se diz que os subespaços $W_1, \ldots, W_r$ ***reduzem*** *L.*

Teorema 11.6 *Seja $L : V \to V$ linear e $W_1, \ldots, W_r$ subespaços de V invariantes sob L e tais que $V = W_1 \oplus \cdots \oplus W_r$. Neste caso, L possui uma representação matricial diagonal em bloco*

$$A = \begin{pmatrix}
A_1 & 0 & \cdots & 0 \\
0 & A_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & A_r
\end{pmatrix}$$

onde A_i é uma representação matricial da restrição L_i de L no subespaço W_i.

Prova. Provaremos o teorema no caso em que $V = W_1 \oplus W_2$. Sejam $\mathcal{B}_1 = \{u_1, \ldots, u_r\}$ base de W_1 e $\mathcal{B}_2 = \{w_1, \ldots, w_s\}$ base de W_2. Como W_1 e W_2 são invariantes frente

L,

$$\begin{aligned}
L(u_1) &= a_{11}u_1 + \cdots + a_{r1}u_r \\
&\cdots \\
L(u_r) &= a_{1r}u_1 + \cdots + a_{rr}u_r \\
L(w_1) &= b_{11}w_1 + \cdots + b_{s1}w_s \\
&\cdots \\
L(w_s) &= b_{1s}w_1 + \cdots + b_{ss}w_s
\end{aligned}$$

Desta forma, a representação matricial de L nesta base é

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1r} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{r1} & \cdots & a_{rr} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & b_{11} & \cdots & b_{1s} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & b_{s1} & \cdots & b_{ss} \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}.$$

□

Nas condições do teorema anterior, temos $L = L_1 \oplus \cdots \oplus L_r$. A matriz A é chamada de **soma direta** de $A_1, \ldots, A_r$ e se escreve

$$A = A_1 \oplus \cdots \oplus A_r.$$

11.1 Polinômio mínimo

Para obter representações matriciais simplificadas de um operador linear é preciso obter subespaços invariantes. Se $f(t)$ for um polinômio e L for um operador linear, então o núcleo de $f(L)$ é invariante sob L. Este fato nos fornece um modo sistemático de obter subespaços invariantes.

Em particular vamos provar que

Teorema 11.7 *Seja $L : V \to V$ linear e $g(t)$, $h(t)$ polinômios mônicos primos entre si, tais que L é um zero do polinômio $f(t) = g(t)h(t)$. Então os subespaços* $\ker g(L)$ *e* $\ker g(L)$ *são invariantes sob L e*

$$V = \ker g(L) \oplus \ker h(L)$$

Sabemos que L é um zero de seu polinômio característico. Vamos provar que todo polinômio que tem L como zero possui os mesmos fatores irredutíveis. Dentre eles, destaca-se o de menor grau, denominado de polinômio mínimo de L.

Definição 11.8 *Seja $L : V \to V$ um operador linear. O polinômio mínimo de L é aquele polinômio mônico de menor grau para o qual $m(L) = 0$.*

Teorema 11.9 *Toda matriz quadrada possui um único polinômio mínimo.*

Prova. (Existência) Sabemos que a matriz L é um zero do seu polinômio característico e assim, existe pelo menos um polinômio não nulo que possui L como zero. Considere o conjunto de todos os polinômios mônicos que se anulam em L. Existe pelo menos um polinômio não nulo de grau mínimo neste conjunto. Este é um polinômio mínimo de L.

(Unicidade) Seja $m(t)$ o polinômio mônico de menor grau para o qual $m(L) = 0$. Seja $f(t)$ outro polinômio mônico de mesmo grau que $m(t)$ e que possui L como zero. Então $g(t) = f(t) - m(t)$ não é nulo e seu grau é menor que o grau de $m(t)$. Ao mesmo tempo, $g(L) = 0$, contrariando a hipótese de $m(t)$ ser o polinômio de menor grau que possui L como zero. □

Teorema 11.10 *O polinômio mínimo de L divide todo polinômio que tem L como zero. Em particular, o polinômio mínimo de L divide o polinômio característico de L.*

Prova. Seja $m(t)$ o polinômio mínimo de L e $f(t)$ um polinômio mônico para o qual $f(L) = 0$. O grau de $f(t)$ é maior do que o grau de $m(t)$ e, pelo algoritmo da divisão, $f(t) = q(t)m(t) + r(t)$, onde $grau(r) < grau(m)$. Existem duas possibilidades: $r(t) = 0$ ou $r(t) \neq 0$. Esta possibilidade, $r(t) \neq 0$, deve ser descartada pois nos leva a uma contradição, considerando-se que $r(A) = 0$ e o $grau(r) < grau(m)$. Logo, $r(t) = 0$ e $m(t)$ divide $f(t)$. □

Teorema 11.11 *Seja $m(t)$ o polinômio mínimo e $\Delta(t)$ o polinômio característico de um operador linear $L : V \to V$. Se a dimensão de V for n, então $\Delta(t)$ divide $m(t)^n$.*

Prova. Seja $f(t) = c_0 + c_1 t + c_2 t^2 + \cdots + t^r$, qualquer polinômio mônico para o qual $f(L) = 0$ ou

$$c_0 I + c_1 L + c_2 L^2 + \cdots + L^r = 0,$$

onde se pode explicitar $c_0 I$

$$c_0 I = -c_1 L - c_2 L^2 - \cdots - L^r.$$

Esta expressão pode ser usada para eliminar $c_0 I$ em $f(t)I$.

$$\begin{aligned} f(t)I &= c_0 I + c_1 tI + c_2 t^2 I + \cdots + t^r I \\ &= c_1(tI - L) + c_2(t^2 I - L^2) + \cdots + (t^r I - L^r) \\ &= (tI - L)(c_1 + c_2(tI + L) + \cdots + (t^{r-1} I + t^{r-2} L + \cdots + L^{r-1})) \\ &= (tI - L)B(t). \end{aligned}$$

Se $f(t)$ for o polinômio mínimo $m(t)$ de L, vale $(tI - L)B(t) = m(t)I$. Calculando o determinante dos dois membros, segue

$$\Delta(t) \det B(t) = m(t)^n.$$

□

Teorema 11.12 *Os polinômios mínimo e característico de um operador linear L possuem os mesmos fatores irredutíveis. Logo, possuem as mesmas raízes.*

Prova. Seja $m(t)$ o polinômio mínimo e $\Delta(t)$ o polinômio característico de L.

Pelo teorema anterior, $\Delta(t)$ divide $m(t)^n$. Assim, os fatores irredutíveis de $\Delta(t)$ devem ser fatores irredutíveis de $m(t)^n$ que possui os mesmos fatores irredutíveis que $m(t)$.

Por outro lado, $m(t)$ divide $\Delta(t)$. Logo, os fatores irredutíveis de $m(t)$ também devem ser fatores irredutíveis de $\Delta(t)$. Isto prova o teorema. □

Corolário 11.13 *Um escalar λ é um autovalor de um operador linear L se e só se λ for uma raiz do polinômio mínimo de L.*

Exemplo 11.14 *Considere a matriz*

$$A = \begin{pmatrix} 2 & 2 & -5 \\ 3 & 7 & -15 \\ 1 & 2 & -4 \end{pmatrix}.$$

, eigenvectors: $\left\{ \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 5 \\ 0 \\ 1 \end{pmatrix} \right\} \leftrightarrow 1, \left\{ \begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix} \right\} \leftrightarrow 3$*Seu polinômio característico é*

$$\Delta(t) = (t-1)^2(t-3).$$

Os candidatos a polinômio mínimo são $\Delta(t)$ *e*

$$f(t) = (t-1)(t-3).$$

Já sabemos que $\Delta(A) = 0$. *Vamos verificar se* $f(A) = 0$. *Sendo* $f(t) = t^2 - 4t + 3$, *um cálculo simples mostra que* $f(A) = A^2 - 4A + 3I = 0$. *Logo,* $f(t)$ *é o polinômio mínimo de* A.

Exemplo 11.15 *O polinômio característico e mínimo da matriz*

$$\begin{pmatrix} 5 & 2 & 0 \\ 0 & 5 & 2 \\ 0 & 0 & 5 \end{pmatrix}$$

são ambos iguais a $(t-5)^3$.

Exemplo 11.16 *Dada a matriz*

$$A = \begin{pmatrix} 5 & 2 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{pmatrix},$$

seu polinômio característico é $(t-5)^3$ *e seu polinômio mínimo é* $(t-5)^2$.

Exemplo 11.17 *Dada a matriz*

$$A = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{pmatrix},$$

seu polinômio característico é $(t-\lambda)^3$ *e seu polinômio mínimo é* $(t-\lambda)$.

Exemplo 11.18 *Os polinômios característico e mínimo de*

$$A = \begin{pmatrix} 0 & 0 & 2 \\ 1 & 0 & 3 \\ 0 & 1 & 5 \end{pmatrix}$$

são ambos iguais a $t^3 - 5t^2 - 3t - 2$.

Exemplo 11.19 *Os polinômios característico e mínimo de*

$$\begin{pmatrix} 0 & 0 & 0 & 2 \\ 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & 7 \end{pmatrix}$$

são ambos iguais a $t^4 - 7t^3 - 5t^2 - 3t - 2$.

Exemplo 11.20 *Os polinômios característico e mínimo de*

$$\begin{pmatrix} 0 & 0 & 0 & a_0 \\ 1 & 0 & 0 & a_1 \\ 0 & 1 & 0 & a_2 \\ 0 & 0 & 1 & a_3 \end{pmatrix}$$

são ambos iguais a $t^4 - a_3t^3 - a_2t^2 - a_1t - a_0$.

Exemplo 11.21 *A matriz* $A = \begin{pmatrix} 5 & 6 \\ 3 & -2 \end{pmatrix}$ *tem dois autovalores:* -4 *e* 7. *Os autovetores correspondentes a eles são* $\begin{pmatrix} 2 \\ -3 \end{pmatrix}$ *e* $\begin{pmatrix} 3 \\ 1 \end{pmatrix}$. *Os polinômios característico e mínimo são iguais* $\Delta(t) = m(t) = t^2 - 3t - 28$. *Assim*

$$\frac{1}{11}\begin{pmatrix} 1 & -3 \\ 3 & 2 \end{pmatrix}\begin{pmatrix} 5 & 6 \\ 3 & -2 \end{pmatrix}\begin{pmatrix} 2 & 3 \\ -3 & 1 \end{pmatrix} = \begin{pmatrix} -4 & 0 \\ 0 & 7 \end{pmatrix}.$$

Exemplo 11.22 *Os autovalores de* $\begin{pmatrix} 5 & 6 \\ 3 & 2 \end{pmatrix}$ *são* -1 *e* 8 *e os autovetores correspondentes a eles são, respectivamente,* $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$ *e* $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$.

Exemplo 11.23 *A matriz* $C = \begin{pmatrix} 5 & -1 \\ 1 & 3 \end{pmatrix}$ *possui um único autovalor real* $\lambda = 4$ *e um único autovetor linearmente independente* $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$. *Os polinômios característico e mínimo são iguais* $\Delta(t) = m(t) = t^2 - 8t + 16$.

Exemplo 11.24 *Os autovalores de* $\begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix}$ *são* 1 *e* 4. *Os autovetores correspondentes são* $\begin{pmatrix} 2 \\ -1 \end{pmatrix}$ *e* $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$. *Os polinômios carcterístico e mínimo são iguais:* $t^2 - 5t + 4$.

Exemplo 11.25 *Os autovalores de* $\begin{pmatrix} 4 & 1 & -1 \\ 2 & 5 & -2 \\ 1 & 1 & 2 \end{pmatrix}$ *são* 3 *e* 5. *Correspondente a* $\lambda_1 = 3$ *temos dois autovetores LI* $\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$ *e* $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$. *Correspondente a* $\lambda_2 = 5$ *temos um autovetor LI* $\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$. *Seu polinômio característico é* $\Delta(t) = (t-3)^2 (t-5)$ *e mínimo é* $m(t) = (t-3)(t-5)$.

Exemplo 11.26 *Os autovalores de* $A = \begin{pmatrix} 3 & -1 & 1 \\ 0 & 6 & -3 \\ 0 & 1 & 2 \end{pmatrix}$ *são* 3 *e* 5. *Correspondentes ao autovalor* $\lambda_1 = 3$ *temos dois autovetores linearmente independentes,* $\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$ *e* $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$. *Todos os autovetores correspondentes ao autovalor* $\lambda_2 = 5$ *são múltiplos de* $\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$. *O polinômio característico de* A *é* $\Delta(t) = (t-3)^2 (t-5)$ *e o polinômio mínimo é* $m(t) = (t-3)(t-5)$.

Exemplo 11.27 *Os autovalores de* $\begin{pmatrix} -3 & 1 & -1 \\ -7 & 5 & -1 \\ -6 & 6 & -2 \end{pmatrix}$ *são* -2 *e* 4 *e os autovetores correspondentes são* $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ *e* $\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$. *O polinômio característico e mínimo são* $\Delta(t) = m(t) = (t-4)(t+2)^2$.

Exemplo 11.28 *Dada a matriz* $\begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$, *seu polinômio característico é* $(t-2)^3$ *e seu polinômio mínimo é* $t - 2$.

Exemplo 11.29 *Dada a matriz* $\begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$, *seu polinômio característico é* $(t-2)^3$ *e seu polinômio mínimo é* $(t-2)^2$.

Exemplo 11.30 *Dada a matriz* $\begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}$, *seu polinômio característico é* $(t-2)^3$ *e seu polinômio mínimo é* $(t-2)^3$

Exemplo 11.31 *Dada a matriz* $\begin{pmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}$, *seus autovetores* $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$ *formam uma base de* $\mathbf{C}^5$. *Seu polinômio característico é* $(t-2)^5$ *e seu polinômio mínimo é* $t-2$.

Exemplo 11.32 *Dada a matriz* $\begin{pmatrix} 2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}$, *seus autovetores são* $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$, *seu polinômio característico é* $(t-2)^5$ *e seu polinômio mínimo é* $(t-2)^2$.

Exemplo 11.33 *Dada a matriz* $\begin{pmatrix} 2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}$, *seus autovetores são* $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$, *seu polinômio característico é* $(t-2)^5$ *e seu polinômio mínimo é* $(t-2)^3$.

Exemplo 11.34 *Dada a matriz* $\begin{pmatrix} 2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}$, *seus autovetores são* $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$, *polinômio característico:* $(t-2)^5$, *polinômio mínimo:* $(t-2)^4$.

Exemplo 11.35 *Dada a matriz* $\begin{pmatrix} 2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}$, *seus autovetores são* $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$, *seu polinômio característico é* $(t-2)^5$ *e seu polinômio mínimo é* $(t-2)^5$

Nota 11.36 *Parece-me que dá para tirar uma regra para calcular a multiplicidade geométrica de um autovalor: Calcule os polinômios característico e mínimo da matriz e fatore-os. Se o polinômio característico possuir o fator* $(t-\lambda)^n$ *o polinômio característico terá o fator* $(t-\lambda)^m$ *com* $m \leq n$. *O número de autovetores correspondentes a* λ *(multiplicidade geométrica de* λ*) é* $n-m+1$.

11.2 Matrizes em bloco

*** Incluir casos mais gerais do que apenas diagonais em bloco.

Se A e B forem matrizes quadradas, então

$$D = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix},$$

onde 0 são matrizes retangulares nulas, é uma **matriz diagonal em bloco**. Se $f(t)$ for um polinômio em t, então

$$f(D) = \begin{pmatrix} f(A) & 0 \\ 0 & f(B) \end{pmatrix}.$$

Teorema 11.37 *Sejam* A_1 *e* A_2 *são matrizes quadradas e*

$$A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$$

uma matriz diagonal em bloco. Então o polinômio mínimo de A *é igual ao mínimo múltiplo comum dos polinômios mínimos dos blocos diagonais* A_1 *e* A_2.

Prova. Sejam $m(t)$, $m_1(t)$ e $m_2(t)$ os polinômios mínimos de A, A_1 e A_2, respectivamente. Vamos provar que $m(t) = \text{mmc}\{m_1(t),\, m_2(t)\}$. Sendo $m(t)$ o polinômio mínimo de A, então

$$m(A) = \begin{pmatrix} m(A_1) & 0 \\ 0 & m(A_2) \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Conclui-se que $m(A_1) = m(A_2) = 0$. Logo, $m_1(t)$ e $m_2(t)$ dividem $m(t)$ e daí, $m(t)$ é multiplo comum de $m_1(t)$ e $m_2(t)$.

Vamos mostrar que A é uma raiz de todo polinômio múltiplo de $m_1(t)$ e $m_2(t)$. Sendo $f(t)$ um múltiplo comum de $m_1(t)$ e $m_2(t)$, então

$$f(A) = \begin{pmatrix} f(A_1) & 0 \\ 0 & f(A_2) \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = 0$$

pois $f(A_1) = 0$ e $f(A_2) = 0$. Daí A é um zero de todo múltiplo comum de $m_1(t)$ e $m_2(t)$. Sendo o polinômio mínimo o poliômio de menor grau que possui A como raiz, $m(t)$ é o mínimo múltiplo comum de $g(t)$ e $h(t)$. □

Teorema 11.38 *Seja A uma matriz diagonal em bloco, cujos blocos diagonais são A_1, A_2, ..., A_r. Então o polinômio mínimo de A é igual ao mínimo múltiplo comum (mmc) dos polinômios mínimos dos blocos diagonais A_i*

$$m_A(t) = \text{mmc}\{\ m_{A_1}(t),\ m_{A_2}(t),\ \ldots,\ m_{A_r}(t)\ \}.$$

Prova. A prova usa o teorema anterior e indução em r. □

Nota 11.39 *Frisamos que este teorema se aplica a matrizes **diagonais em blocos**, enquanto o Teorema (10.9) que é análogo a este e se refere aos polinômios característicos, aplica-se a matrizes **triangulares em blocos**.*

11.3 Decomposição primária

Teorema 11.40 *Seja $L : V \to V$ uma transformação linear e W invariante sob L. Seja $L_W : W \to W$ a restrição de L em W. Sob estas hipóteses, o polinômio mínimo de L_W divide o polinômio mínimo de L.*

Prova. Se $m(t)$ for o polinômio mínimo de L, então $m(L) = 0$. Para todo $w \in W$, temos $m(L_W)(w) = m(L)(w) = 0$. Sendo L_W um zero de $m(t)$, o polinômio mínimo de L_W divide o polinômio mínimo de L. □

Lembramos que, se $f(t)$ e $g(t)$ forem dois polinômios quaisquer,

$$f(t)g(t) = \text{mdc}(f(t), g(t))\text{mmc}(f(t), g(t))$$

Teorema 11.41 *Seja $L : V \to V$ uma transformação linear, V_1 e V_2 invariantes sob L de modo tal que $V = V_1 \oplus V_2$. Sejam L_1 e L_2 as restrições de L a V_1 e V_2.*

1. *O polinômio mínimo de L é o mínimo múltiplo comum dos polinômios mínimos de L_1 e L_2.*

2. *O polinômio característico de L é o produto dos polinômios característicos de L_1 e de L_2.*

Prova. 1. Sejam $m(t)$, $m_1(t)$ e $m_2(t)$ os polinômios mínimos de L, L_1 e L_2, respectivamente. Pelo teorema anterior, $m(t)$ é divisível por $m_1(t)$ e por $m_2(t)$.

Seja $f(t)$ outro polinômio que tem L por raiz. Então $f(L_1) = f(L_2) = 0$ e f é divisível por $m_1(t)$ e $m_2(t)$. Provamos que todo polinômio que tem L por raiz é múltiplo comum de $m_1(t)$ e $m_2(t)$. Sendo $m(t)$ o polinômio de menor grau que tem L como raiz,

$$m(t) = \operatorname{mmc}(m_1(t), m_2(t)).$$

2. Se $\mathcal{B}_1 = \{u_1, \ldots, u_r\}$ for uma base de V_1 e $\mathcal{B}_2 = \{w_1, \ldots, w_s\}$ for uma base de V_2, então a união $\mathcal{B}_1 \cup \mathcal{B}_2$ é uma base de V. Como V_1 e V_2 são invariantes sob L, para $j = 1, \ldots, r$, tem-se

$$Lu_j = \sum_{i=1}^{r} a_{ij} u_i$$

e, para $j = 1, \ldots, s$,

$$Lw_j = \sum_{i=1}^{r} b_{ij} w_i.$$

A matriz da transformação linear L nesta base é triangular em bloco

$$M = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}.$$

Consequentemente,

$$\det(tI - M) = \det(tI - A)\det(tI - B) = \Delta_{L_1}(t)\Delta_{L_2}(t).$$

□

Nota 11.42 *No teorema anterior, se $m_1(t)$ e $m_2(t)$ forem primos entre si, então $m(t) = m_1(t)m_2(t)$.*

Teorema 11.43 *Seja $L : V \to V$ linear e $g(t)$, $h(t)$ polinômios mônicos primos entre si, tais que L é um zero do polinômio $f(t) = g(t)h(t)$. Então:*

1. *Os subespaços $\ker g(L)$ e $\ker h(L)$ são invariantes sob L e*

$$V = \ker g(L) \oplus \ker h(L)$$

2. *Se $f(t)$ for o polinômio mínimo de L, então $g(t)$ e $h(t)$ são os polinômios mínimos das restrições de L ao* $\ker g(L)$ *e ao* $\ker h(L)$*, respectivamente.*

Prova. Inicialmente, observamos que $\ker g(L)$ e $\ker h(L)$ são invariantes sob L.

1. Como $g(t)$ e $h(t)$ são primos entre si, existem polinômios $r(t)$ e $s(t)$ tais que

$$r(t)g(t) + s(t)h(t) = 1,$$

acarretando na igualdade

$$I = r(L)g(L) + s(L)h(L).$$

Sendo v um elemento de V, podemos escrever

$$v = r(L)g(L)v + s(L)h(L)v.$$

Nesta soma, $r(L)g(L)v$ pertence ao $\ker h(L)$ e $w = s(L)h(L)v$ pertence ao $\ker g(L)$. De fato,

$$h(L)r(L)g(L)v = r(L)g(L)h(L)v = r(L)f(L)v = 0$$

pois $f(L) = 0$. De modo análogo se prova que $s(L)h(L)v$ pertence ao $\ker g(L)$. Consequentemente,

$$V = \ker g(L) + \ker h(L).$$

Para completar a prova deste item, falta mostrar que esta decomposição é única. Seja $v = u + w$, uma decomposição de v, com $u \in \ker g(L)$ e $w \in \ker h(L)$. Então,

$$\begin{aligned} u &= r(L)g(L)u + s(L)h(L)u = s(L)h(L)u \\ &= s(L)h(L)(u+w) = s(L)h(L)v. \end{aligned}$$

De modo semelhante se prova que $w = s(L)h(L)v$.

Como a decomposição é única,

$$V = \ker g(L) \oplus \ker h(L).$$

2. Se $f(t) = g(t)h(t)$ for o polinômio mínimo de L, então ele é divisível pelos polinômios mínimos $m_1(t)$ e $m_2(t)$ de L_1 e L_2, como já se provou.

Por outro lado, $g(L_1) = 0$ e $h(L_2) = 0$. Portanto, $m_1(t)$ divide $g(t)$ e $m_2(t)$ divide $h(t)$. Como $g(t)$ e $h(t)$ são primos entre si, o mesmo ocorre com $m_1(t)$ e $m_2(t)$ que os divide.

Sendo $m_1(t)$ e $m_2(t)$ primos entre si e $f(t) = \text{mmc}\{m_1(t), m_2(t)\}$, segue

$$f(t) = m_1(t)m_2(t).$$

Por outro lado, $f(t) = g(t)h(t)$, de onde segue que $m_1(t) = g(t)$ e $m_2(t) = h(t)$.

□

Teorema 11.44 *(Decomposição primária) Seja $L : V \to V$ linear cujo polinômio mínimo é igual a*

$$m(t) = f_1(t)^{n_1} \cdots f_r(t)^{n_r}$$

onde os polinômios $f_i(t)$ são mônicos, distintos e irredutíveis. Então $f_i(t)^{n_i}$ são os polinômios mínimos das restrições de L a $W_i = \ker\left(f_i(L)^{n_i}\right)$ e

$$V = W_1 \oplus \cdots \oplus W_r.$$

Prova. Para simplificar a prova, vamos abordar o caso particular em que $m(t) = f_1(t)^{n_1} f_2(t)^{n_2}$ onde $f_1(t)$, $f_2(t)$ são polinômios mônicos, distintos e irredutíveis. Certamente, $f_1(t)^{n_1}$ e $f_2(t)^{n_2}$ são primos entre si. Este teorema decorre imediatamente do anterior.

De fato, sabemos que $V = W_1 \oplus W_2$, onde $W_i = \ker f_i(t)^{n_i}$ e que $f_i(t)^{n_i}$, é o polinômio mínimo da restrição de L em W_i. □

Teorema 11.45 *Uma transformação linear $L : V \to V$ possui uma representação matricial diagonal se e só se seu polinômio mínimo,*

$$m(t) = (t - \lambda_1) \cdots (t - \lambda_r)$$

for um produto de polinômios lineares distintos. Os elementos da diagonal principal desta matriz são os autovalores $\lambda_1, \ldots, \lambda_r$ de L.

Prova. 1. Se L possui uma representação matricial diagonal $D = diag(\ \lambda_1, \ldots, \lambda_n)$, admitamos que apenas os r primeiros λ_i são distintos. O polinômio característico desta transformação linear é da forma $\Delta(t) = (t - \lambda_1)^{n_1} \cdots\ (t - \lambda_r)^{n_r}$. O polinômio mínimo possui os mesmos fatores de $\Delta(t)$. Como $(D - \lambda_1 I) \cdots\ (D - \lambda_r I) = 0$ o polinômio mínimo de L é $(t - \lambda_1) \cdots\ (t - \lambda_r)$.

2. Sendo $m(t) = (t - \lambda_1) \cdots (t - \lambda_r)$, então existe uma decomposição de V numa soma direta $W_1 \oplus \cdots \oplus W_r$, onde $W_i = \ker(L - \lambda_i I)$ é o autoespaço de V correspondente ao autovalor λ_i. Se $B_i = \{v_{i1}, v_{i2}, \ldots, v_{is_i}\}$ for uma base de W_i, então $L(v_{ij}) = \lambda_i v_{ij}$. Seja B a união *das* bases B_i, para $i = 1, \ldots, r$, que é uma base de V. Nesta base, a matriz que representa L é diagonal. □

11.4 Diagonalização de operadores normais

Um operador $L : V \to V$ é **diagonalizável** se existir uma base $\{v_1, \ldots, v_n\}$ de V tal que a matriz de L nesta base é diagonal. Isto significa que a matriz de L numa base qualquer é semelhante a uma matriz diagonal. Isto significa que, se A for a matriz de L numa base de V, então existe uma matriz inversível P e uma matriz diagonal D tais que

$$A = PDP^{-1}.$$

Podemos nos perguntar qual a condição mais geral que um operador linear deve satisfazer para ser diagonalizável. Quando o operador linear $L : V \to V$ é auto-adjunto, antiadjunto ou unitário, então $LL^* = L^*L$. Esta é a condição mais geral sob a qual um operador é diagonalizável.

Definição 11.46 *Um operador linear $L : V \to V$ é **normal** quando*

$$LL^* = L^*L.$$

*A matriz quadrada A que satisfaz $AA^* = A^*A$ é denominada **matriz normal**.*

Exemplo 11.47 *Os operadores auto-adjuntos, antiadjuntos e unitários são normais.*

Exemplo 11.48 *Se L for normal, I for o operador identidade e λ um escalar, então o operador $L - \lambda I$ é normal.*

Se λ for autovalor de L, então $\bar{\lambda}$ é autovalor de L^*. Se L for um operador normal, provaremos logo em seguida que v é autovetor de L correspondente ao autovalor λ se e só se v for autovetor de L^* correspondente ao autovalor $\bar{\lambda}$.

Quando L não é normal, os autovetores de L não são, necessariamente, os autovetores de L^*.

Seja V um espaço vetorial de dimensão finita com produto interno e $L : V \to V$ um operador linear. Se S for um subespaço de V invariante sob L então $S^\perp$ é invariante sob L^*. Agora, quando L for normal e S for invariante sob L, provaremos que tanto S quanto $S^\perp$ são invariantes sob L e L^*.

Teorema 11.49 *Seja L um operador normal sobre um espaço vetorial de dimensão finita com produto interno. Se v for autovetor de L correspondente ao autovalor λ, então v é autovetor de L^* correspondente ao autovalor $\bar{\lambda}$.*

Prova. Se L é normal, $LL^* = L^*L$ e, portanto, $\langle Lv, Lw\rangle = \langle L^*v, L^*w\rangle$ para todo v e w em V. Em particular, para todo v em V, $\langle Lv, Lv\rangle = \langle L^*v, L^*v\rangle$ que nos fornece a igualdade $\|Lv\| = \|L^*v\|$. Para todo escalar λ, $T = L - \lambda I$ é normal pois $T^* = L^* - \bar{\lambda} I$. Sendo T normal, $\|Tv\| = \|T^*v\|$ para todo v em V. Desta forma, $Tv = 0$ se e só se $T^*v = 0$. Conclui-se que v é autovetor de L se e só se for autovetor de L^*.

Se λ for autovalor de L, seja v o autovetor correspondente. No item anterior provou-se que v é autovetor de L^*. Sendo μ o autovalor correspondente segue

$$\bar{\lambda}\langle v, v\rangle = \langle \lambda v, v\rangle = \langle Lv, v\rangle = \langle v, L^*v\rangle = \mu \langle v, v\rangle .$$

Sendo $\langle v, v\rangle \neq 0$, segue $\mu = \bar{\lambda}$. $\square$

Teorema 11.50 *Seja L uma transformação normal sobre um espaço vetorial de dimensão finita com produto interno V. Então*

1. *V possui uma base ortonormal formada por autovetores de L.*

2. *Seja A uma matriz complexa normal de ordem n. Existe uma matriz unitária U e uma matriz diagonal D para as quais*

$$A = UDU^{-1}.$$

Prova. 1. Seja n a dimensão de V. A transformação L possui pelo menos um autovetor v de módulo unitário. O subespaço vetorial $W = ger(v)$ e $W^{\perp}$ são invariantes sob L e sob L^*. A dimensão de W é 1 e a de $W^{\perp}$ é $n-1$.

Vamos provar que a restrição T de L em $W^{\perp}$ é normal. Para todo v e w em $W^{\perp}$ temos $\langle v, T^*w\rangle = \langle Tv, w\rangle = \langle Lv, w\rangle = \langle v, L^*w\rangle$ de onde concluímos que o adjunto de T é a restrição de L^* a $W^{\perp}$. Para todo v em $W^{\perp}$, $TT^*v = LL^*v = L^*Lv = T^*Tv$, mostrando que T é normal.

O teorema agora será demonstrado por indução na dimensão n do espaço.

Se $n = 1$, nada resta a provar: a base de V contém apenas um autovetor v_1 de norma unitária.

Vamos supor, como hipótese de indução, que o teorema vale para todos os operadores normais em espaços vetoriais de dimensão $n-1$.

Provemos que o teorema vale para todo operador normal L definido em um espaço vetorial V de dimensão n. Seja v_1 um autovetor de L e $W = ger(v_1)$. Seja T a restrição de L a $W^{\perp}$. O operador T é normal em $W^{\perp}$ que tem dimensão $n-1$. Pela hipótese de indução, $W^{\perp}$ possui uma base ortonormal $\{v_2, \ldots, v_n\}$ cujos elementos são autovetores de T. Os autovetores de T são autovetores de L. Ao incluirmos v_1 a este conjunto, obtemos a base ortonormal $\{v_1, v_2, \ldots, v_n\}$ formada por autovetores de L. Nela, a representação matricial de L é diagonal. □

Teorema 11.51 *Seja L um operador linear sobre um espaço vetorial V de dimensão finita. Se V possuir uma base ortonormal formada por autovetores de L, então L é normal.*

Prova. Seja $\mathcal{B} = \{v_1, \ldots, v_n\}$ uma base ortonormal de V cujos elementos são autovetores de L, de modo que $Lv_i = \lambda_i v_i$ para $i = 1, \ldots, n$. O escalar λ_i é o autovalor de L correspondente ao autovetor v_i. Se decompondo L^*v_i na base $\mathcal{B}$, podemos escrever

$$L^*v_i = \sum_j a_{ij} v_j$$

onde, graças à ortonormalidade da base $\mathcal{B}$, $\bar{a}_{ij} = \langle L^*v_i, v_j\rangle$. Por outro lado,

$$\bar{a}_{ij} = \langle L^*v_i, v_j\rangle = \langle v_i, Lv_j\rangle = \langle v_i, \lambda_j v_j\rangle = \lambda_j \langle v_i, v_j\rangle = \lambda_j \delta_{ij}.$$

Isto significa que $a_{ij} = 0$ quando $i \neq j$ e $a_{ii} = \bar{\lambda}_i$, de modo que $L^* v_i = \bar{\lambda}_i v_i$. Mostramos assim que v_i é autovetor de L^* correspondente ao autovalor $\bar{\lambda}_i$. Portanto, $L^* L v_i = \lambda_i \bar{\lambda}_i v_i = |\lambda_i|^2 v_i$ e $LL^* v_i = \bar{\lambda}_i \lambda_i v_i = |\lambda_i|^2 v_i$. Sendo as transformações $L^* L$ e LL^* iguais em cada elemento da base, são iguais no espaço todo, provando que L é normal. □

Seja V um espaço vetorial de dimensão finita com produto interno e $L : V \to V$ uma transformação normal. Sejam $\lambda_1, \ldots, \lambda_r$ os autovalores de L e $S_i = \{v \in V : Lv = \lambda_i v\}$ o autoespaço do autovalor λ_i. Se $\{v_{1i}, \ldots, v_{si}\}$ for uma base ortonormal de S_i, então, para todo v em S_i temos $L(v) = \lambda_i \langle v_{1i}, v \rangle v_{1i} + \cdots + \lambda_i \langle v_{si}, v \rangle v_{si} = \lambda_i P_i(v)$, onde P_i é a projeção ortogonal de V sobre S_i. Logo L coincide com $\lambda_i P_i$ em S_i. Se a soma dos autoespaços S_i resultar numa decomposição de V em soma direta, então

$$L = \lambda_1 P_1 + \cdots + \lambda_n P_n.$$

Teorema 11.52 *(Versão projetiva do teorema espectral) Seja V um espaço vetorial complexo com produto interno e dimensão n. Seja $L : V \to V$ um operador normal e $\{v_1, \ldots, v_n\}$ uma base ortonormal de V, formada pelos autovalores de L. Os autoespaços $S_i = auto(\lambda_i)$, $i = 1, \ldots, r$, são ortogonais dois a dois e sua soma é igual a V. Se $P_i : V \to V$ for a projeção ortogonal sobre S_i, então*

$$P_1 + \cdots + P_n = I$$

e

$$L = \lambda_1 P_1 + \cdots + \lambda_n P_n.$$

Teorema 11.53 *(Versão do teorema espectral para matriz real simétrica) Seja A uma matriz normal de ordem n. Sejam $\lambda_1, \ldots, \lambda_r$ seus autovalores distintos. Então*

$$A = \lambda_1 P_1 + \cdots + \lambda_r P_r$$

onde P_i são as matrizes que projetam ortogonalmente sobre o autoespaço de λ_i. Estes autoespaços são ortogonais e sua soma direta é igual a $\mathbf{C}^n$, de modo que

$$P_1 + \cdots + P_k = I.$$

Exemplo 11.54 *A matriz* $A = \begin{pmatrix} 7 & 4 & -5 \\ 4 & -2 & 4 \\ -5 & 4 & 7 \end{pmatrix}$ *é simétrica. Seus autovalores são* 6, 12 *e* -6. *Os autovetores correspondentes são* $(1,\ 1,\ 1)^T$, $(-1,\ 0,\ 1)^T$ *e* $(1,\ -2,\ 1)^T$. *Para montar S tal que $A = SDS^{-1}$, tomamos as colunas de S iguais aos autovetores normalizados*

$$S = \begin{pmatrix} 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0 & -2/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \end{pmatrix} \quad e \quad D = \begin{pmatrix} 6 & 0 & 0 \\ 0 & 12 & 0 \\ 0 & 0 & -6 \end{pmatrix}.$$

Podemos escrever A como uma combinação linear de matrizes de projeção

$$A = 6P_1 + 12P_2 - 6P_3,$$

onde

$$P_1 = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}, \quad P_2 = \begin{pmatrix} 1/2 & 0 & -1/2 \\ 0 & 0 & 0 \\ -1/2 & 0 & 1/2 \end{pmatrix} \quad e \quad P_3 = \frac{1}{6}\begin{pmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{pmatrix}$$

Exemplo 11.55 *A matriz* $A = \begin{pmatrix} 0 & 1 & 1 \\ -1 & 0 & 2 \\ -1 & -2 & 0 \end{pmatrix}$ *é anti-simétrica. Seus autovalores são* 0, $i\sqrt{5}$ *e* $-i\sqrt{5}$. *Os autovetores correspondentes são* $(2, -1, 1)$, $(-.4 - .4899i, .2 - .9798i, 1)$ *e* $(-.4 + .4899i, .2 + .9798i, 1)$.

Exemplo 11.56 *A matriz* $A = \begin{pmatrix} 1/\sqrt{3} & 2/\sqrt{6} & 9 \\ 1/\sqrt{3} & -1/\sqrt{6} & -1/\sqrt{2} \\ 1/\sqrt{3} & -1/\sqrt{6} & 1/\sqrt{2} \end{pmatrix}$ *é ortogonal. Seus autovalores são* -1; $0,9381+\ 0,3464i$ *e* $0,9381-\ 0,3464i$. *Os autovetores correspondentes são*

$$\begin{aligned} &(0,5176; 1; 0,4142), \\ &(0,1691 + 0,9463i; -0,3267 + 0,4898i; 1), \\ &(0,1691 - 0,9463i; -0,3267 - 0,4898i; 1). \end{aligned}$$

11.5 Decomposição de Schur

Teorema 11.57 *Dada uma matriz complexa A de ordem n, existe uma matriz unitária U tal que*

$$T = U^* A U$$

é triangular superior.

Prova. Será usada a indução sobre n. Se $n = 1$, A é triangular superior e nada resta a provar. Se $n > 1$, assuma, como hipótese de indução, que o teorema é verdadeiro para toda matriz quadrada de ordem $n-1$. Seja q_1 um autovetor unitário correspondente a um autovalor λ_1 de A. Construa uma base de $\mathbf{C}^n$ onde um dos elementos é q_1. Use o processo de ortogonalização de Gram-Schmidt para obter uma base ortonormal $\{q_1, \ldots, q_n\}$ de $\mathbf{C}^n$. A matriz

$$U_1 = [q_1, \ldots, q_n]$$

é unitária e

$$U_1^* A U_1 = \begin{bmatrix} \lambda_1 & b_1 \\ 0 & A_1 \end{bmatrix}$$

onde A_1 é uma matriz quadrada de ordem $n-1$. De acordo com a hipótese de indução, existe uma matriz unitária V_1 de ordem $n-1$ tal que

$$T_1 = V_1^* A_1 V_1$$

é triangular superior. A matriz

$$U_2 = \begin{bmatrix} 1 & 0 \\ 0 & V_1 \end{bmatrix}$$

é unitária e, sendo $U = U_1 U_2$, segue

$$\begin{aligned} U^*AU &= U_2^* \left(U_1^* A U_1\right) U_2 \\ &= \begin{bmatrix} 1 & 0 \\ 0 & V_1^* \end{bmatrix} \begin{bmatrix} \lambda_1 & b_1 \\ 0 & A_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & V_1 \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 & b_1 a_1 \\ 0 & V_1^* A_1 V_1 \end{bmatrix} = \begin{bmatrix} \lambda_1 & b_1 a_1 \\ 0 & T_1 \end{bmatrix} = T \end{aligned}$$

que é triangular superior. $\square$

Como A e T são semelhantes, possuem os mesmos autovalores e com a mesma multiplicidade. Sendo T triangular superior, os elementos da diagonal principal são seus autovalores e, consequentemente, autovalores de A.

Nota 11.58 *Sejam $\lambda_1, \ldots, \lambda_n$ os elementos da diagonal principal de T. Como matrizes semelhantes possuem o mesmo determinante e o mesmo traço, segue*

$$\det(A) = \lambda_1 \times \cdots \times \lambda_n \qquad e \qquad tr(A) = \lambda_1 + \cdots + \lambda_n.$$

Teorema 11.59 *(Teorema Espectral). Se H é uma matriz hermitiana, existe uma matriz unitária U tal que U^*HU é diagonal.*

Prova. Pelo teorema de Schur, existe U unitária tal que $U^*HU = T$ é triangular superior. Como $T^* = U^*H^*U = U^*HU = T$, vemos que T é simética. Logo, também é triangular inferior, ou seja, é diagonal. $\square$

A matriz $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ não é hermitiana mas é ortogonalmente diagonalizável, com

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & i \\ i & -1 \end{bmatrix}.$$

A matriz mais geral diagonalizável é a normal.

Teorema 11.60 *Uma matriz de ordem n é diagonalizável unitariamente se e só se for normal.*

Prova. Se A for uma matriz de ordem n diagonalizável unitáriamente, existe uma matriz unitária U tal que $D = U^*AU$ é diagonal. Como D e D^* são diagonais, segue que $DD^* = D^*D$, ou $(U^*AU)(U^*A^*U) = (U^*A^*U)(U^*AU)$ de onde segue $U^*AA^*U = U^*A^*AU$. Multiplicando à esquerda por U e à direita por U^*, obtemos $AA^* = A^*A$, provando que A é normal.

Se A for normal, então $AA^* = A^*A$. Pelo teorema de Schur, existe uma matriz unitária U tal que $T = U^*AU$ é triangular superior. Vamos provar por indução em n que T é diagonal. Se $n = 1$, a matriz T é diagonal. Se $n > 1$, então

$$TT^* = T^*T.$$

Sendo $T = [t_{ij}]$, então igualando os elementos $(1, 1)$ em na igualdade acima, obtemos

$$|t_{11}|^2 + |t_{12}|^2 + \cdots + |t_{1n}|^2 = |t_{11}|^2$$

pois T é triangular superior. Segue que $t_{12} = \cdots = t_{1n} = 0$. Logo T tem a forma de blocos

$$T = \begin{bmatrix} t_{11} & 0 \\ 0 & T_1 \end{bmatrix}$$

onde T_1 é normal e, por hipótese de indução, é diagonal. Cnclui-se daí que T é diagonal, completando a prova do teorema. $\square$

Nota 11.61 *As matrizes normais reais* 2×2 *são as matrizes simétricas e aquelas com a forma*

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

11.6 Decomposição em valores singulares

Seja A uma matriz complexa m por n. A matriz A^*A é auto-adjunta e seus autovalores são todos positivos ou nulos. Podemos posicioná-los em ordem descrescente

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0.$$

Sendo A^*A auto-adjunta, existe uma base ortonormal $\{q_1, \ldots, q_n\}$ de $\mathbf{C}^n$ onde q_i é o autovetor correspondente ao autovalor λ_i. A matriz

$$Q = [q_1, \ldots, q_n]$$

cujas colunas são os autovetores de A^*A, é unitária e

$$A^*AQ = QD$$

onde $D = diag(\lambda_1, \ldots, \lambda_n)$ é uma matriz quadrada diagonal $n \times n$. Como $\lambda_i \geq 0$, definimos $\sigma_i = \sqrt{\lambda_i} \geq 0$. Seja $r \leq n$ o número de valores σ_i diferentes de zero, de modo que

$$\sigma_1 \geq \cdots \geq \sigma_r > 0 \qquad \text{e} \qquad \sigma_{r+1} = \cdots = \sigma_n = 0.$$

Para $i = 1, \ldots, r$ os vetores de $\mathbf{C}^m$ definidos por

$$p_i = \frac{1}{\sigma_i} A q_i$$

formam um conjunto ortonormal. Uma vez que $q_{r+1}, \ldots, q_n$ estão no núcleo de A (pois são levados por A em 0), o conjunto $\{p_1, \ldots, p_r\}$ é uma base ortonormal da imagem de A e, por este motivo r é o posto de A que deve ser menor ou igual a m.

Para provar que $\{p_1, \ldots, p_r\}$ é um conjunto ortonormal, basta calcular, para $1 \leq i$, $j \leq r$, o produto interno

$$\begin{aligned} \langle p_i, p_j \rangle &= \frac{1}{\sigma_i \sigma_j} \langle A q_i, A q_j \rangle = \frac{1}{\sigma_i \sigma_j} \langle q_i, A^* A q_j \rangle \\ &= \frac{\sigma_j^2}{\sigma_i \sigma_j} \langle q_i, q_j \rangle = \frac{\sigma_j}{\sigma_i} \langle q_i, q_j \rangle = \delta_{ij}. \end{aligned}$$

Pode-se completar $\{p_1, \ldots, p_r\}$ de modo a obter uma base ortonormal $\{p_1, \ldots, p_r, p_{r+1}, \ldots, p_m\}$ de $\mathbf{C}^m$.

A matriz $m \times m$

$$P = [p_1, \ldots, p_r, p_{r+1}, \ldots, p_m]$$

cujas colunas são os vetores p_i, é unitária e, pelo modo como foi construída,

$$AQ = P\Sigma$$

onde

$$\Sigma = diag\{\sigma_1, \ldots, \sigma_r, 0, \ldots, 0\}.$$

é uma matriz $m \times n$ onde todos os elementos são nulos, à excessão dos r primeiros elementos da diagonal principal que são $\sigma_1 \geq \cdots \geq \sigma_r$. Lembramos que $r \leq \min(m, n)$.

Como Q é unitária, obtemos a **decomposição** de A **em valores singulares**

$$A = P\Sigma Q^*$$

Os números reais $\sigma_1, \ldots, \sigma_r$, 0, são denominado de **valores singulares** de A.

Usando a expressão de A deduzida acima, prova-se que

$$A^* A Q = Q \Sigma^* \Sigma \qquad \text{e} \qquad A A^* P = P \Sigma \Sigma^*$$

mostrando que as colunas q_i de Q são autovetores da matriz $A^* A$, que as colunas p_i de P são autovetores da matriz AA^* e, tanto no primeiro quanto no segundo caso, autovetores correspondentes aos autovalores σ_i^2.

Lembramos que $\Sigma\Sigma^*$ é uma matriz diagonal $m \times m$ e $\Sigma^*\Sigma$ é uma matriz diagonal $n \times n$. Se A tiver posto máximo, então $r = \min\{m, n\}$ e uma das duas matrizes, $\Sigma\Sigma^*$ ou $\Sigma^*\Sigma$, possui todos os termos diagonais diferentes de zero e, portanto, é não singular.

Provamos o seguinte teorema:

Teorema 11.62 *(Decomposição em valores singulares) Seja A uma matriz $m \times n$ cujo posto é r $(r \leq \min\{m, n\})$. Então:*

1. *Esta matriz possui r valores singulares $\sigma_1 \geq \cdots \geq \sigma_r$ não nulos (incluindo sua multiplicidade).*

2. *Existe um conjunto ortonormal $\{q_1, \ldots, q_r\}$ formado por autovetores de A^*A, correspondentes aos autovalores $\sigma_1^2 \geq \cdots \geq \sigma_r^2$ tais que o conjunto*

$$\{p_1, \ldots, p_r\} = \{\frac{1}{\sigma_1} A q_1, \ldots, \frac{1}{\sigma_r} A q_r\}$$

   é uma base ortonormal da imagem de A.

3. *Se $\{q_{r+1}, \ldots, q_n\}$ for uma base ortonormal do núcleo de A, então $\{q_1, \ldots, q_r, q_{r+1}, \ldots, q_n\}$ é uma base ortonormal de $\mathbf{C}^n$. Sendo*

$$\{\ p_1,\ \ldots,\ p_r,\ p_{r+1},\ \ldots,\ p_m\}$$

   uma base ortonormal de $\mathbf{C}^m$, então

$$A = P \Sigma Q^*$$

   onde

$$\begin{aligned} P &= [p_1,\ \ldots,\ p_m] \\ Q &= [q_1,\ \ldots,\ q_n] \\ \Sigma &= diag(\sigma_1,\ \ldots,\ \sigma_r) \end{aligned}$$

   onde apenas os primeiros r elementos da diagonal principal são diferentes de zero e iguais a $\sigma_1, \ldots, \sigma_r$.

Exemplo 11.63 *Obtenha a decomposição em valores singulares de* $A = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \end{pmatrix}$.

Exemplo 11.64 *Obtenha a decomposição em valores singulares de* $A = \begin{pmatrix} 1 & 2 & 0 & 1 \\ 1 & 0 & 2 & 1 \\ 2 & 2 & 2 & 2 \end{pmatrix}$.

Então $A^*A = \begin{pmatrix} 6 & 6 & 6 & 6 \\ 6 & 8 & 4 & 6 \\ 6 & 4 & 8 & 6 \\ 6 & 6 & 6 & 6 \end{pmatrix}$ *cujos autovalores são* $24,\ 4,\ 0$ *e* 0*.Os autovetores correspondentes a eles são*

$$\frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ -1 \\ 1 \\ 0 \end{bmatrix}, \quad \frac{1}{\sqrt{12}}\begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \frac{1}{\sqrt{6}}\begin{bmatrix} 0 \\ -1 \\ -1 \\ 2 \end{bmatrix}$$

Assim,

$$Q = \begin{pmatrix} 1/2 & 0 & -3/\sqrt{12} & 0 \\ 1/2 & -1/\sqrt{2} & 1/\sqrt{12} & -1/\sqrt{6} \\ 1/2 & 1/\sqrt{2} & 1/\sqrt{12} & -1/\sqrt{6} \\ 1/2 & 0 & 1/\sqrt{12} & 2/\sqrt{6} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sqrt{24} & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Para determinar P, calculamos $p_1 = (1/\sqrt{24})Aq_1$ e $p_2 = (1/2)Aq_2$ para obter

$$\frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \quad \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}.$$

Como o contradomínio de A tem dimensão 3, precisamos de mais um vetor para formar a base. Calculamos $p_3 = \begin{pmatrix} p_{13} & p_{23} & p_{33} \end{pmatrix}^T$ unitário e fazendo-o ortogonal a p_1 e a p_2. Obtemos então $p_3 = (1/\sqrt{3}) \begin{pmatrix} 1 & 1 & -1 \end{pmatrix}^T$. Com estes vetores montamos

$$P = \begin{pmatrix} 1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \\ 2/\sqrt{6} & 0 & -1/\sqrt{3} \end{pmatrix}.$$

Nota: Na decomposição em valores singulares de A, $AQ = P\Sigma$, de modo que $\|Aq_1\|_2 = \|\sigma_1 p_1\|_2 = \sigma_1$. Por outro lado, para qualquer x em $\mathbf{R}^n$ com $\|x\|_2 = 1$, temos $x = \sum_{i=1}^n \alpha_i q_i$ onde $\sum_{i=1}^n \alpha_i^2 = 1$. Daí,

$$\|Ax\|_2^2 = \left\| \sum_{i=1}^n \alpha_i A q_i \right\|_2^2 = \left\| \sum_{i=1}^r \alpha_i \sigma_i p_i \right\|_2^2 = \sum_{i=1}^r \sigma_i^2 \alpha_i^2 \|p_i\|_2^2$$

sendo a segunda igualdade verdadeira pois $Aq_i = 0$ para $i > r$ e a última igualdade se justifica pela ortogonalidade dos p_i. Como $\|p_i\|_2 = 1$, segue

$$\|Ax\|_2^2 = \sum_{i=1}^r \sigma_i^2 \alpha_i^2 \leq \sigma_1^2 \sum_{i=1}^r \alpha_i^2 \leq \sigma_1^2 \sum_{i=1}^n \alpha_i^2 = \sigma_1^2.$$

A primeira desigualdade se justifica pois σ_1 é o maior valor singular e a segunda por termos acrescentado parcelas positivas à soma. Provamos que $\|Ax\|_2^2 \leq \sigma_1^2$ e que $\|Aq_1\|_2 = \sigma_1$. Logo,

$$\|A\|_2 = \sup_{\substack{x \in \mathbf{R}^n \\ \|x\|_2 = 1}} \|Ax\|_2 = \sigma_1.$$

Se restringirmos A ao espaço $ger(q_1)^\perp$ podemos calcular σ_2, repetindo o procedimento acima

$$\sigma_2 = \sup_{\substack{x \in \langle q_1 \rangle^\perp \\ \|x\|_2 = 1}} \|Ax\|_2 .$$

Certamente a decomposição por valores singulares não é única pois a escolha de Q e de P não é única. Entretanto, temos o seguinte teorema:

Teorema 11.65 *Se*

$$A = UST^*$$

for outra decomposição por valores singulares de A, *então* $S = \Sigma$, *as colunas* $t_1, \ldots, t_n$ *de* T *formam um conjunto ortonormal de autovetores de* A^*A *correspondentes aos autovalores* $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0 = \cdots = 0$, *as colunas* $u_1, \ldots, u_m$ *de* U *são autovetores de* AA^* *correspondentes aos mesmos autovalores e, para* $i = 1, \ldots, r$,

$$u_i = \frac{1}{\sigma_i} A t_i.$$

Prova. Como $A^*A = TS^*U^*UST^* = T(S^*S)T^*$, vemos que a matriz diagonal S^*S é semelhante à matriz A^*A e, portanto, possuem os mesmos autovalores, o que prova a igualdade $S = \Sigma$.

Da decomposição acima, obtemos $A^*AT = TS^*S$ de onde segue que as colunas de T formam um conjunto ortonormal de autovetores de A^*A correspondentes aos autovalores $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0 = \cdots = 0$.

Da mesma decomposição, obtemos $AA^*U = USS^*$ de onde segue que as colunas de U formam um conjunto ortonormal de autovetores de AA^* correspondentes aos autovalores $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0 = \cdots = 0$.

Ainda da decomposição, segue $AT = US$, mostrando que $u_i = \frac{1}{\sigma_i} A t_i$ para $i = 1, \ldots, r$. $\square$

Capítulo 12

Forma canônica de Jordan

12.1 Operadores nilpotentes

Definição 12.1 *Um operador linear $L : V \to V$ é **nilpotente** se $L^k = 0$ para algum inteiro k positivo. O menor k para o qual $L^k = 0$ é chamado de **índice da nilpotência** de L.*

Se o índice da nilpotência de L for k, significa que existe v em V tal que $L^{k-1}v \neq 0$.

Observe que $L^k = 0$ quando $L^k v = 0$ para todo v em V. Se L for nilpotente com índice k, seu polinômio mínimo será t^k e assim o seu único autovalor é o zero.

Teorema 12.2 *Seja $L : V \to V$ linear e $v \in V$ não nulo tal que $L^{k-1}(v) \neq 0$ e $L^k(v) = 0$. Então:*

1. *O conjunto $S = \{ v, Lv, \ldots, L^{k-1}(v) \}$ é linearmente independente.*

2. *O subespaço gerado por S é invariante sob L.*

3. *A restrição de L ao subespaço gerado por S é nilpotente com índice k.*

4. *A matriz da restrição de L ao subespaço gerado por S em relação à base ordenada S é da forma*
$$\begin{pmatrix} 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{pmatrix}.$$

Prova. Vamos provar um item por vez.

1. Sejam $\beta_1, \ldots, \beta_k$ escalares tais que $\beta_1 v + \cdots + \beta_k L^{k-1}v = 0$. Aplicando L^{k-1} a esta igualdade, obtemos $\beta_1 L^{k-1}v = 0$. Como $L^{k-1}v \neq 0$, segue $\beta_1 = 0$. Aplicando sucessivamente L^i à igualdade inicial, com $i = k-2, k-3, \ldots, 1$, se prova que $\beta_2 = \beta_3 = \cdots = \beta_k = 0$, o que prova a primeira parte do teorema.

2. Seja $w = \alpha_1 v + \cdots + \alpha_k L^{k-1}v$, onde $\alpha_1, \ldots, \alpha_k$ são escalares, um vetor no subespaço gerado por S. Assim, $Lw = \alpha_1 Lv + \cdots + \alpha_{k-1}L^{k-1}v$ e, portanto, Lw pertence ao subespaço gerado por S.

3. Seja $T : [S] \to [S]$ a restrição de L ao subespaço $[S]$ gerado por S. Como $T^k w = 0$ para todo elemento de $[S]$ e $T^{k-1}v = L^{k-1}v \neq 0$, concluímos que T é nilpotente com índice k.

4. Como $T(v) = Lv$, $T(Lv) = L^2 v, \ldots, T(L^{k-2}v) = L^{k-1}v$, $T(L^{k-1}v) = 0$ e concluímos que a matriz de T na base S é exatamente aquela apresentada no enunciado.

□

Em particular, quando L for nilpotente, seu índice de nilpotência é menor ou igual à dimensão do espaço vetorial.

Teorema 12.3 *Seja $L : V \to V$ linear. Para todo inteiro $i \geq 0$,*

1. $\ker L^i \subset \ker\ L^{i+1}$

2. $L(\ker\ L^{i+1}) \subset \ker L^i$.

Prova. Se v pertence ao $\ker L^i$, então $L^i v = 0$ e $L^{i+1}v = L(L^i v) = L0 = 0$. Logo, v pertence ao $\ker L^{i+1}$, o que prova a primeira parte do teorema.

Seja w um elemento de $L(\ker L^{i+1}) = \{Lv : v \in \ker L^{i+1}\}$. Então $w = Lv$ para algum v no $\ker L^{i+1}$. Assim, $L^i w = L^{i+1}v = 0$, provando que w pertence ao $\ker L^i$. Provamos a segunda parte do teorema. □

Teorema 12.4 *Seja $L : V \to V$ linear e $i > 0$ um inteiro. Pelo teorema anterior,*

$$\ker L^{i-1} \subset \ker L^i \subset \ker L^{i+1}.$$

Suponhamos que

$$\begin{array}{l} \{\ u_1,\ \ldots,\ u_r\ \} \\ \{\ u_1,\ \ldots,\ u_r,\ v_1,\ \ldots,\ v_s\ \} \\ \{\ u_1,\ \ldots,\ u_r,\ v_1,\ \ldots,\ v_s,\ w_1,\ \ldots,\ w_t\ \} \end{array}$$

sejam bases de $\ker L^{i-1}$, $\ker L^i$ *e* $\ker L^{i+1}$. *Então o conjunto*

$$\{\ u_1,\ \ldots,\ u_r,\ Lw_1,\ \ldots,\ Lw_t\ \}$$

é linearmente independente e está contido no $\ker L^i$.

Prova. O teorema anterior assegura que os vetores $Lw_1, \ldots, Lw_t$ pertencem ao $\ker L^i$. Sejam $\alpha_1, \ldots, \alpha_r, \beta_1, \ldots, \beta_t$ escalares tais que

$$\alpha_1 u_1 + \cdots + \alpha_r \, u_r + \beta_1 L w_1 + \cdots + \beta_t L w_t = 0.$$

Aplicando L^{i-1} a esta igualdade, segue

$$L^i(\beta_1 w_1 + \cdots + \beta_t w_t) = 0,$$

mostrando que a combinação linear $\beta_1 w_1 + \cdots + \beta_t w_t$ pertence ao $\ker L^i$ e pode ser escrito como uma combinação linear de $u_1, \ldots, u_r, v_1, \ldots, v_s$

$$\beta_1 w_1 + \cdots + \beta_t w_t = c_1 u_1 + \cdots + c_r u_r + d_1 v_1 + \cdots + d_s v_s.$$

Sendo $\{u_1, \ldots, u_r, v_1, \ldots, v_s, w_1, \ldots, w_t\}$ uma base, concluímos que todos os escalares na igualdade acima são nulos e, em particular, $\beta_1 = \cdots = \beta_t = 0$. Sendo $\{u_1, \ldots, u_r\}$ uma base, deve-se ter também $\alpha_1 = \cdots = \alpha_r = 0$, o que prova a independência linear do conjunto de vetores $\{ u_1, \ldots, u_r, Lw_1, \ldots, Lw_t \}$. $\square$

Este resultado é interessante pois ele nos permite inferir que $s \geq t$. Sabemos que a dimensão do $\ker L^i$ cresce com i ou permanece inalterada. Além disso, o acréscimo na dimensão quando passamos do $\ker L^i$ para o $\ker L^{i+1}$ nunca é maior do que o acréscimo na dimensão quando passamos do $\ker L^{i-1}$ para o $\ker L^i$. ***

Teorema 12.5 *(Forma canônica de um operador nilpotente) Seja $L : V \to V$ um operador nilpotente com índice k. Existe uma base na qual a representação matricial de L tem a forma bloco diagonal*

$$N = \begin{pmatrix} N_1 & 0 & 0 & \cdots \\ 0 & N_2 & 0 & \cdots \\ 0 & 0 & N_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Cada bloco diagonal N_i é uma matriz quadrada que pode ser 1×1 e nula ou ter a forma

$$N_i = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

As ordens de todos os blocos são menores ou iguais a k. Pelo menos um bloco tem ordem k. O número de blocos é igual à nulidade de L. O número de blocos do tipo N é determinado de modo único por L.

Em lugar de demonstrar vamos dar exemplos.

Exemplo 12.6 *Seja $L : V \to V$ um operador linear nilpotente com índice k. Denotemos por W_i o núcleo de L^i. Relembre inicialmente que, se*

$$\begin{aligned}&\{u_1, \ldots, u_r\}\\&\{u_1, \ldots, u_r, v_1, \ldots, v_s\}\\&\{u_1, \ldots, u_r, v_1, \ldots, v_s, w_1, \ldots, w_t\}\end{aligned}$$

forem bases de W_{i-1}, W_i e W_{i+1}, respectivamente, então o conjunto

$$\{\ u_1,\ \ldots,\ u_r,\ L(w_1),\ \ldots,\ L(w_t)\ \}$$

é linearmente independente em W_i. Este fato nos fornece os fundamentos para obter uma base de V na qual a representação de um operador nilpotente se encontra na forma canônica preconizada.

1. *Imaginemos que V tem dimensão $n = 8$ e que $L : V \to V$ é nilpotente com índice $k = 4$. Assim o $\ker L^4 = V$. Denotemos o $\ker L^i$ por W_i. Seja $\{u_1,\ u_2,\ u_3,\ u_4,\ u_5,\ u_6,\ u_7,\ u_8\}$ uma base de $V = W_4 = \ker L^4$, de modo que*

$$\begin{aligned}&\{u_1, u_2, u_3\} \text{ é base do } \ker L.\\&\{u_1, u_2, u_3, u_4, u_5\} \text{ é base do } \ker L^2.\\&\{u_1, u_2, u_3, u_4, u_5, u_6, u_7\} \text{ é base do } \ker L^3.\\&\{u_1, u_2, u_3, u_4, u_5, u_6, u_7, u_8\} \text{ é base de } V = \ker L^4.\end{aligned}$$

   Podemos construir o seguinte quadro

W_1			W_2		W_3		W_4
u_1	u_2	u_3	u_4	u_5	u_6	u_7	u_8
L^3u_8	L^2u_7	u_{10}	L^2u_8	Lu_7	Lu_8	u_9	u_8
w_1	w_5	w_8	w_2	w_6	w_3	w_7	w_4

   A base ordenada $\{u_1,\ \ldots,\ u_8\}$ da segunda linha da tabela anterior é substituída pela base ordenada da terceira linha, que são renomeados na quarta linha para $\{w_1,\ \ldots,\ w_8\}$. Os vetores u_9 e u_{10} da terceira linha são construídos de modo que $\{Lu_8,\ u_9\}$ gere o mesmo subespaço que $\{u_6,\ u_7\}$ e $\{L^3u_8,\ L^2u_7,\ u_{10}\}$ gere o mesmo subespaço que $\{u_1,\ u_2,\ u_3\}$.

   Na base $\{w_1,\ w_2,\ w_3,\ w_4,\ w_5,\ w_6,\ w_7,\ w_8\}$ a matriz de L possui a forma

$$\left[\begin{array}{cccc|ccc|c} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ \hline 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

2. *Consideremos ainda que V tem dimensão $n = 8$ e que $L : V \to V$ é nilpotente com índice $k = 4$. Seja $\{u_1, u_2, u_3, u_4, u_5, u_6, u_7, u_8\}$ uma base de $V = W_4$, de modo que*

$$\begin{aligned}
&\{u_1, u_2, u_3, u_4\} \text{ é base de } W_1 \\
&\{u_1, u_2, u_3, u_4, u_5, u_6\} \text{ é base de } W_2 \\
&\{u_1, u_2, u_3, u_4, u_5, u_6, u_7\} \text{ é base de } W_3 \\
&\{u_1, u_2, u_3, u_4, u_5, u_6, u_7, u_8\} \text{ é base de } W_4
\end{aligned}$$

Podemos construir o seguinte quadro seguindo o esquema anterior.

W_1				W_2		W_3	W_4
u_1	u_2	u_3	u_4	u_5	u_6	u_7	u_8
L^3u_8	Lu_9	u_{10}	u_{11}	L^2u_8	u_9	Lu_8	u_8
w_1	w_5	w_7	w_8	w_2	w_6	w_3	w_4

Na base $\{w_1,\ w_2,\ w_3,\ w_4,\ w_5,\ w_6,\ w_7,\ w_8\}$ a matriz de L possui a forma

$$\left[\begin{array}{cccc|cc|c|c}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right]$$

12.2 Forma canônica de Jordan

Teorema 12.7 *Seja $L : V \to V$ linear com polinômio característico e mínimo iguais a*

$$\begin{aligned}
\Delta(t) &= (t - \lambda_1)^{n_1} \cdots (t - \lambda_r)^{n_r} \\
m(t) &= (t - \lambda_1)^{m_1} \cdots (t - \lambda_r)^{m_r}.
\end{aligned}$$

Então L possui uma representação matricial em bloco diagonal J, denominada de forma canônica de Jordan do operador L, cujos elementos diagonais têm a forma

$$J_{ij} = \begin{bmatrix}
\lambda_i & 1 & 0 & \cdots & 0 & 0 \\
0 & \lambda_i & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \cdots & \vdots & \vdots \\
0 & 0 & 0 & \ddots & 1 & 0 \\
0 & 0 & 0 & \cdots & \lambda_i & 1 \\
0 & 0 & 0 & \cdots & 0 & \lambda_i
\end{bmatrix}.$$

Para cada i fixado,

1. *Há pelo menos um J_{ij} de ordem m_i e todos os demais são de ordem menores ou iguais a m_i.*

2. *A soma das ordens dos J_{ij} é n_i.*

3. *O número de J_{ij} é igual à multiplicidade geométrica de λ_i, que é igual à nulidade de $N_i = (L_i - \lambda_i I)$.*

4. *O número de blocos J_{ij} de cada ordem possível é univocamente determinado por L.*

A matriz J_{ij} é denominada **bloco de Jordan** pertencente ao autovalor λ_i.
Observe que

$$J_{ij} = \lambda_i I + N$$

onde

$$N = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \ddots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

é um bloco nilpotente.

Exemplo 12.8 *Seja $L : \mathbf{R}^7 \to \mathbf{R}^7$ linear, cujos polinômios característico e mínimo são*

$$\begin{aligned} \Delta(t) &= (t-2)^4(t-3)^3 \\ m(t) &= (t-2)^2(t-3)^2 \end{aligned}$$

A forma canônica de Jordan de L é uma das seguintes

$$\begin{bmatrix} 2 & 1 & & & & & \\ 0 & 2 & & & & & \\ & & 2 & 1 & & & \\ & & 0 & 2 & & & \\ & & & & 3 & 1 & \\ & & & & 0 & 3 & \\ & & & & & & 3 \end{bmatrix} \quad \text{ou} \quad \begin{bmatrix} 2 & 1 & & & & & \\ 0 & 2 & & & & & \\ & & 2 & & & & \\ & & & 2 & & & \\ & & & & 3 & 1 & \\ & & & & 0 & 3 & \\ & & & & & & 3 \end{bmatrix}$$

A primeira matiz ocorre se L possui dois autovetores linearmente independentes pertencentes ao seu autovalor 2 e a segunda ocorre se L tem três autovetores linearmente independentes pertencentes ao seu autovalor 2.

Exemplo 12.9 *Vamos determinar a forma canônica J da matriz $A = \begin{bmatrix} 0 & 1 \\ -1 & 2 \end{bmatrix}$. A equação característica de A é $(\lambda - 1)^2 = 0$ e uma base do autoespaço correspondente ao* *autovalor* 1 *é* $v_1 = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$. *Calculamos então* $(A-\lambda I)^2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ *e vemos que o zero é autovalor desta matriz e que qualquer vetor é autovetor correspondente ao zero. Tomemos* $v_2 = \begin{bmatrix} 0 & 1 \end{bmatrix}^T$ *que, juntamente com* v_1*, forma uma base do espaço das matrizes* 2×1. *A matriz* $S = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$, *cujas colunas são* v_1 *e* v_2*, tem por inversa* $S^{-1} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$ *e é tal que* $J = S^{-1}AS = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Exemplo 12.10 *Vamos determinar a forma canônica* J *da matriz* $A = \begin{bmatrix} 3 & -2 & 5 \\ -1 & 2 & 1 \\ -1 & 1 & 0 \end{bmatrix}$. *A equação característica de* A *é* $(\lambda-1)(\lambda-2)^2 = 0$. *A multiplicidade algébrica do autovalor* 1 *é* 1 *e do autovalor* 2 *é* 2. *Uma base do autoespaço correspondente ao autovalor* 1 *é formada pelo vetor* $v_1 = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}^T$. *Uma base do autoespaço correspondente ao autovalor* 2 *é formada pelo vetor* $\begin{bmatrix} 1 & 3 & 1 \end{bmatrix}^T$. *Um conjunto gerador do autoespaço de* $(A-2I)^2 = \begin{bmatrix} -2 & 3 & -7 \\ -2 & 3 & -7 \\ 0 & 0 & 0 \end{bmatrix}$ *correspondente ao autovalor* 0 *é formado pelos vetores* $v_3 = \begin{bmatrix} 3 & 2 & 0 \end{bmatrix}^T$ *e* $v_4 = \begin{bmatrix} -7 & 0 & 2 \end{bmatrix}^T$. *Como nenhum deles é múltiplo de* $\begin{bmatrix} 1 & 3 & 1 \end{bmatrix}^T$ *podemos tomar* v_3 *para gerar a uma cadeia de Jordan de comprimento* 2, *correspondente ao autovalor* 2 *e calculamos* $v_2 = (A-2I)v_3 = \begin{bmatrix} -1 & -3 & -1 \end{bmatrix}^T$. *A matriz* $S = \begin{bmatrix} 1 & -1 & 3 \\ 1 & -3 & 2 \\ 0 & -1 & 0 \end{bmatrix}$, *cuja inversa é* $S^{-1} = \begin{bmatrix} -2 & 3 & -7 \\ 0 & 0 & -1 \\ 1 & -1 & 2 \end{bmatrix}$ *é tal que* $J = S^{-1}AS = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}$ *que é a forma canônica da matriz* A.

Exemplo 12.11 *****

12.3 Subespaços cíclicos

Seja $L : V \to V$ linear e $v \in V$ não nulo ($v \neq 0$). Consideremos a seqüência

$$v,\ Lv,\ L^2v,\ \ldots$$

Seja k o menor inteiro para o qual

$$L^kv \in [v,\ Lv,\ L^2v,\ \ldots,\ L^{k-1}v],$$

indicando com isto que o conjunto

$$\{v,\ Lv,\ L^2v,\ \ldots,\ L^{k-1}v\}$$

é linearmente independente.

O subespaço vetorial

$$Z(v, L) = [v,\ Lv,\ L^2 v,\ \ldots,\ L^{k-1} v]$$

é chamado de **subespaço cíclico de V gerado por** L e v. Sua dimensão é k.

Este subespaço é a interseção de todos os subespaços L invariantes que contêm v.

Denotemos por L_v a restrição de L a $Z(v, L)$. Se

$$L^k v = -a_0 v - a_1 Lv - a_2 L^2 v - \cdots - a_{k-1} L^{k-1} v$$

então

$$m_v(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_{k-1} t^{k-1} + t^k$$

é o polinômio mínimo de L_v e a representação de L_v na base

$$\{v,\ Lv,\ L^2 v,\ \ldots,\ L^{k-1} v\}$$

é

$$C = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & 0 & -a_1 \\ 0 & 1 & 0 & \cdots & 0 & 0 & -a_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & -a_{k-3} \\ 0 & 0 & 0 & \cdots & 1 & 0 & -a_{k-2} \\ 0 & 0 & 0 & \cdots & 0 & 1 & -a_{k-1} \end{bmatrix}$$

denominada de **matriz companheira** do polinômio $m_v(t)$. O polinômio $m_v(t)$ é denominado de L **anulador** de v e $Z(v, L)$.

12.4 Forma canônica racional

Lema 12.12 *11.13. Seja $L : V \to V$ linear, cujo polinômio mínimo é $f(t)^n$, onde $f(t)$ é irredutível. Então existem $v_1,\ v_2,\ \ldots,\ v_r$ tais que*

$$V = Z(v_1, L) \oplus \cdots \oplus Z(v_r, L).$$

O polinômio mínimo da restrição de L a $Z(v_i, L)$ é $f(t)^{n_i}$, onde n_i é um número inteiro menor ou iguais a n. Pode-se ordenar os expoentes n_i de modo que

$$n = n_1 \geq n_2 \geq \cdots \geq n_r.$$

Qualquer outra decomposição de V em subespaços L cíclicos tem o mesmo conjunto de polinômios mínimos, que é determinado de modo único por L. Assim L tem uma representação matricial

$$C = \begin{bmatrix} C^{(1)} & 0 & \cdots & 0 \\ 0 & C^{(2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & C^{(r)} \end{bmatrix}$$

onde $C^{(i)}$ é a matriz companheira do polinômio $f(t)^{n_i}$.

Teorema 12.13 *(Forma canônica racional) Seja $L : V \to V$ linear com polinômio mínimo*

$$m(t) = f_1(t)^{m_1} \dots f_s(t)^{m_s}$$

onde $f_i(t)$ são polinômios mônicos irredutíveis distintos. Então L possui uma única representação matricial em bloco

$$\left[\begin{bmatrix} C_1 & 0 & \cdots & 0 \\ 0 & C_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & C_s \end{bmatrix}\right]$$

onde cada C_i é uma matriz com o formato

$$C_i = \begin{bmatrix} C_i^{(1)} & 0 & \cdots & 0 \\ 0 & C_i^{(2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & C_i^{(r)} \end{bmatrix}$$

em que $C_i^{(j)}$ são matrizes companheiras de $f_i(t)^{n_{ij}}$ onde se pode ordenar os n_{ij} de modo que

$$\begin{aligned} m_1 &= n_{11} \geq n_{12} \geq \cdots \geq n_{1r_1} \\ & \cdots \\ m_s &= n_{s1} \geq n_{s2} \geq \cdots \geq n_{sr_s} \end{aligned}$$

Esta é a chamada forma canônica racional de L. Os polinômios $f_i(t)^{n_{ij}}$ são chamados de divisores elementares de L.

12.5 Forma triangular

Se um operador linear L possuir uma representação matricial triangular

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}$$

seu polinômio característico pode ser fatorado em polinômios do primeiro grau

$$\Delta(t) = \det(tI - A) = (t - a_{11})(t - a_{22}) \cdots (t - a_{nn}).$$

A recíproca também é verdadeira.

Teorema 12.14 *Seja n a dimensão de V e $L : V \to V$ um operador linear cujo polinômio característico $\Delta(t)$ pode ser fatorado num produto de fatores lineares*

$$\Delta(t) = (t - \lambda_1)^{n_1} \cdots (t - \lambda_r)^{n_r}$$

onde os números λ_i, $i = 1, \ldots, r$ são distintos e $n_1 + \cdots + n_r = n$. Então L possui uma representação matricial em forma triangular.

Prova. *** Como $\Delta(t)$ pode ser fatorado em polinômios do primeiro grau, L possui ao menos um autovalor. Denotemo-lo λ_1 e por v_1 o autovetor correspondente, de modo que $Lv_1 = \lambda_1 v_1$. Então $V = V_1 \oplus (V_1)^\perp$ onde $V_1 = ger(v_1)$. O espaço V_1 é invariantes sob L. Seja L_1 a restrição de L a V_1. $\{v_{12}, \ldots, v_{1n}\}$ uma base ortonormal de $(V_1)^\perp$. A matriz de L nesta base é da forma *** $\square$

12.6 Espaços quocientes

Esta é uma maneira inteligente de definir "projeções" em espaços vetoriais que não possuem produto interno.

Seja W um subespaço vetorial de V. Dado $v \in V$, definimos o conjunto

$$v + W = \{v + w : w \in W\},$$

denominado de **classe lateral** de W em V. Observe que $0 + W = W$.

Podemos definir duas operações no conjunto das classes laterais de modo a torná-lo um espaço vetorial.

Seja W um subespaço vetorial de V. Sejam u e v dois vetores em V e k um escalar pertencente ao corpo sobre o qual se define o espaço vetorial V. Definimos no conjunto das classes laterais de W as operações de **adição** de duas classes e **multiplicação** de uma classe **por um escalar** por

$$\begin{aligned} (u + W) + (v + W) &= (u + v) + W, \\ k(u + W) &= ku + W. \end{aligned}$$

O conjunto das classes laterais, com estas duas operações, é um espaço vetorial sobre o mesmo corpo sobre o qual se define V. Este espaço vetorial é denominado **espaço quociente** de V por W e é denotado por V/W. Se a dimensão de V for finita então $\dim(V/W) = \dim(V) - \dim(W)$.

Teorema 12.15 *Seja $L : V \to V$ linear e W um subespaço L invariante de V. Então L induz um operador linear $\bar{L}$ em V/W definido por*

$$\bar{L}(v + W) = L(v) + W.$$

Se L for um zero de um polinômio, então $\bar{L}$ também o é. Assim, o polinômio mínimo de $\bar{L}$ divide o polinômio mínimo de L.

Exemplo 12.16 *Vamos apresentar um exemplo que mostra como se pode obter uma representação matricial triangular de uma transformação linear. Seja $L : \mathbf{R}^3 \to \mathbf{R}^3$ definida por $L(x, y, z) = (4x + y - z, 2x + 5y - 2z, x + y + 2z)$. A matriz de L na base canônica do $\mathbf{R}^3$ é*

$$A = \begin{pmatrix} 4 & 1 & -1 \\ 2 & 5 & -2 \\ 1 & 1 & 2 \end{pmatrix}$$

Os vetores $v_1 = (-1,\ 1,\ 0)$, $v_2 = (0,\ 1,\ 0)$, $v_3 = (0,\ 0,\ 1)$ formam uma base do $\mathbf{R}^3$. Destacamos que v_1 é autovetor de L correspondente ao autovetor 3. *Como*

$$\begin{aligned} L(v_1) &= 3v_1 \\ L(v_2) &= -v_1 + 6v_2 + v_3 \\ L(v_3) &= v_1 - 3v_2 + 2v_3 \end{aligned}$$

a matriz de L na base $\{v_1,\ v_2,\ v_3\}$ é

$$B = \begin{pmatrix} 3 & -1 & 1 \\ 0 & 6 & -3 \\ 0 & 1 & 2 \end{pmatrix}.$$

O espaço vetorial W gerado por v_1 é invariante sob L. Observe que a matriz de L na base $\{v_1,\ v_2,\ v_3\}$ já possui a primeira coluna na forma desejada para se chegar à forma triangular.

Consideremos $\overline{V} = \{v + W : v \in V\}$ que é o espaço quociente V/W e a transformação linear induzida $\overline{L} : \overline{V} \to \overline{V}$ definida por $\overline{L}(\overline{v}) = L(v) + W$. Para esta transformação,

$$\begin{aligned} \overline{L}(\overline{v}_1) &= 3v_1 + W = W = \overline{0} \\ \overline{L}(\overline{v}_2) &= -v_1 + 6v_2 + v_3 + W = 6\overline{v}_2 + \overline{v}_3 \\ \overline{L}(\overline{v}_3) &= v_1 - 3v_2 + 2v_3 + W = -3\overline{v}_2 + 2\overline{v}_3 \end{aligned}$$

de modo que a matriz de $\overline{L}$ em relação à base $\{\overline{v}_2,\ \overline{v}_3\}$ de $\overline{V}$ é

$$C = \begin{pmatrix} 6 & -3 \\ 1 & 2 \end{pmatrix}.$$

Vamos omitir a barra e olhar para L no espaço gerado por v_2 e v_3. Sabemos que

$$\begin{aligned} L(v_2) &= 6v_2 + v_3 \\ L(v_3) &= -3v_2 + 2v_3 \end{aligned}$$

cuja matriz na base $\{v_1,\ v_2\}$ é C. Os autovalores de C são 5 *e* 3 *e o autovetor relativo ao autovalor* 5 *é $3v_2 + v_3$. Vamos então passar da base $\{v_1,\ v_2,\ v_3\}$ para a base $\{w_1,\ w_2,\ w_3\}$ onde*

$$w_1 = v_1, \qquad w_2 = 3v_2 + v_3, \qquad w_3 = v_3.$$

O w_3 foi escolhido de modo arbitrário, exigimos apenas que $\{w_1, w_2, w_3\}$ seja uma base de V. Podemos inverter as relações acima para obter

$$v_1 = w_1, \qquad v_2 = (w_2 - w_3)/3, \qquad v_3 = w_3.$$

Daí segue

$$\begin{aligned} L(w_1) &= 3w_1 \\ L(w_2) &= -2w_1 + 5w_2 \\ L(w_3) &= w_1 - w_2 + 3w_3 \end{aligned}$$

e, nesta base, a matriz de L é

$$D = \begin{pmatrix} 3 & -2 & 1 \\ 0 & 5 & -1 \\ 0 & 0 & 3 \end{pmatrix}$$

que está na forma triangular. Esta transformação linear pode ser representada por uma matriz diagonal pois ela possui três autovetores linearmente independente.

Capítulo 13

Aplicações

Aproximação por polinômios
- Cadeias de Markov
- Circuitos elétricos
- Diferenças finitas
- Elementos finitos
- Equação de Schröedinger
- Sistemas de equações diferenciais
- Exponencial de matriz
- Formas quadráticas
- Cônicas e quádricas
- Mínimos quadrados
- Modelo econômico de Leontief
- Método húngaro para alocação de tarefas
- Cifras de Hill
- Programação linear
- Séries de Fourier
- Sistemas de equações diferenciais
- Tensão nos meios contínuos
- Teoria dos grafos
- Teoria dos jogos

Apêndice A

Matrizes

Uma **matriz** é um arranjo retangular de números, denominados de **elementos** da matriz, dispostos em linhas e colunas. Quando uma matriz possuir m linhas e n colunas diremos que é uma matriz $m \times n$ ou matriz m por n ou matriz de **ordem** m por n. Matrizes reais são aquelas cujos elementos são números reais e matrizes complexas são aquelas cujos elementos são números complexos. Em nosso curso trabalharemos com matrizes reais ou complexas.

Uma matriz com uma única coluna é chamada de **vetor coluna** e uma matriz com uma única linha é chamada de **vetor linha**. Se o número de linhas for igual ao número de colunas se diz que a matriz é **quadrada**. Uma matriz quadrada com n linhas e n colunas é uma matriz n por n ou de ordem n. Neste caso, em lugar de dizermos que a ordem da matriz é m por m, diremos apenas que a matriz é de ordem m.

A menos que se especifique o contrário, $\mathbf{R}^n$ é o conjunto das matrizes coluna reais, que possuem n linhas e uma coluna. Denotaremos por $\mathbf{C}^n$ ao conjunto de matrizes coluna complexas, com n linhas e uma coluna. Designaremos o conjunto das matrizes reais m por n pelo símbolo $\mathbf{R}^{m\times n}$ e das matrizes complexas de ordem m por n pelo símbolo $\mathbf{C}^{m\times n}$. Também é usual escrever $A_{m\times n}$ para indicar que A possui m linhas e n colunas. Um número real ou complexo será denominado genericamente de **escalar**.

Usaremos a notação abreviada $A = (a_{ij})$ para denotar uma matriz

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

onde a_{ij} é o elemento da linha i e coluna j. No conjunto das matrizes m por n, se define a adição de duas matrizes e a multiplicação de uma matriz por um escalar através das fórmulas

$$\begin{aligned} (a_{ij}) + (b_{ij}) &= (a_{ij} + b_{ij}) \\ k\,(a_{ij}) &= (ka_{ij}) \end{aligned}$$

onde k é um escalar, (a_{ij}) e (b_{ij}) são matrizes de ordem m por n. Quando for conveniente, escreveremos $(a_{ij})k$ em lugar de $k(a_{ij})$.

A matriz em que todos os elementos são nulos é chamada de **matriz nula** e será denotada por 0.

Se $A = (a_{ij})$, então $-A = (-a_{ij})$ é chamada de **matriz oposta** de A. Definimos a diferença entre as matrizes A e B de mesma ordem por $A - B = A + (-B)$.

Propriedades

Nas propriedades enumeradas abaixo, A, B e C são matrizes de mesma ordem, incluindo a matriz nula e k_1, k_2 são escalares. O 1 indica a unidade escalar.

1. Associatividade: $A + (B + C) = (A + B) + C$.
2. Comutatividade: $A + B = B + A$.
3. Elemento neutro: $A + 0 = 0 + A = A$.
4. Elemento oposto: $A + (-A) = (-A) + A = 0$.
5. Associatividade: $(k_1k_2)A = k_1(k_2A)$.
6. Distributividade: $(k_1 + k_2)A = k_1A + k_2A$.
7. Distributividade: $k_1(A + B) = k_1A + k_1B$.
8. Unidade: $1A = A$

Estas propriedades indicam que o conjunto das matrizes $m \times n$ com as operações de adição e multiplicação por um escalar é um espaço vetorial sobre o corpo dos escalares que, em nosso caso, será o corpo dos números reais ou dos números complexos.

A.1 Matrizes especiais

Seja $A = (a_{ij})$ uma matriz m por n e $p = \min\{m, n\}$. Os elementos a_{11}, a_{22}, $\ldots$, a_{pp} formam a **diagonal principal** da matriz A. Uma matriz é **diagonal** se todos os elementos fora da diagonal principal forem nulos.

A matriz **identidade** I de ordem m é a matriz diagonal cujos elementos da diagonal principal são todos iguais a 1. O delta de Kronecker, definido para todo i e j inteiros por

$$\begin{aligned} \delta_{ij} &= 1 \quad \text{se} \quad i = j \\ \delta_{ij} &= 0 \quad \text{se} \quad i \neq j \end{aligned}$$

pode ser usado para representar os elementos da matriz identidade. Em termos deste símbolo, $I = (\delta_{ij})$.

Se os elementos abaixo da diagonal principal da matriz A forem nulos, a matriz é **triangular superior**. Se os elementos à direita da diagonal principal de A forem nulos, a matriz é **triangular inferior**.

Uma matriz A é **simétrica** se $A^T = A$, é **anti-simétrica** se $A^T = -A$ e **ortogonal** se $A^T = A^{-1}$.

Seja $A = (a_{ij})$ uma matriz complexa de ordem m por n. Vamos indicar por $\bar{a}_{ij}$ ao complexo conjugado de a_{ij}. A matriz $A^* = (b_{ij})$ de ordem n por m, onde

$$b_{ij} = \bar{a}_{ji}$$

é a **adjunta** de A. Se A for real, então $A^* = A^T$. Uma matriz complexa A é **hermitiana** se $A^* = A$, **anti-hermitiana** se $A^* = -A$ e **unitária** se $A^* = A^{-1}$. As matrizes reais simétricas são hermitianas, as matrizes reais anti-simétricas são anti-hermitianas e as matrizes reais ortogonais são unitárias.

Definição A.1 *Uma matriz m por n possui a* ***forma escalonada*** *se:*

1. *As linhas nulas, se existirem, se encontram na parte inferior da matriz.*
2. *Ao percorrer as linhas de cima para baixo, o primeiro elemento não nulo de cada linha vai se deslocando para a direita.*

O primeiro elemento não nulo em cada linha, quando percorrida da esquerda para a direira, é chamado de **pivô** da linha.

Definição A.2 *Uma matriz m por n possui a* ***forma escalonada reduzida*** *se:*

1. *As linhas nulas, se existirem, se encontram na parte inferior da matriz.*
2. *O primeiro elemento não nulo em cada linha, quando percorrida da esquerda para a direira, é igual a* 1. *Este é o pivô da linha.*
3. *São nulos todos os demais elementos da coluna que contém o pivô.*
4. *Ao percorrer as linhas de cima para baixo, o primeiro elemento não nulo de cada linha vai se deslocando para a direita.*

A.2 Multiplicação de matrizes

A multiplicação é a operação que leva duas matrizes $A = (a_{ij})_{m\times n}$ e $B = (b_{jk})_{n\times p}$ na matriz

$$AB = \left(\sum_{k=1}^{n} a_{ik} b_{kj} \right)$$

de ordemque é uma matriz m por p. Para efetuar o produto AB, o número de colunas de A deve ser igual ao número de linhas de B. Quando este for o caso, se diz que A e B são conformes para o produto.

A multiplicação de matrizes é uma operação associativa e distributiva mas não é comutativa. Assim,

1. $A_1(B_1C_1) = (A_1B_1)C_1$
2. $A_2(B_2 + C_2) = A_2B_2 + A_2C_2$
3. $(A_3 + B_3)C_3 = A_3C_3 + B_3C_3$

desde que as matrizes A_i, B_i e C_i sejam conformes para a adição e a multiplicação.

Se se o número de linhas for diferente do número de colunas em A e B, então o produto AB pode estar definido e o produto BA não.

A.3 Inversa

Uma matriz quadrada A de ordem m é **inversível** se existir uma matriz quadrada B de ordem m tal que $AB = BA = I$, onde I é a matriz identidade de ordem m. A matriz B é a **inversa** de A, sendo denotada por A^{-1}. Sendo $A = (a_{ij})$ e $B = (b_{ij})$, então as igualdades matriciais $AB = BA = I$ resultam nas seguintes igualdades entre os elementos de A, B e I

$$\sum_{k=1}^{n} a_{ik}b_{kj} = \delta_{ij} \quad \text{e} \quad \sum_{k=1}^{n} b_{ik}a_{kj} = \delta_{ij}.$$

Se a matriz não for inversível, diremos que é **singular**.

A inversa de uma matriz é única pois, se B e C forem inversas de A, então

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

Se A for inversível, então A^{-1} é inversível e $\left(A^{-1}\right)^{-1} = A$. Se k for um escalar não nulo e A for inversível, então kA é inversível e $(kA)^{-1} = k^{-1}A^{-1}$.

Teorema A.3 *Sejam A e B matrizes quadradas tais que $AB = I$. Isto é suficiente para garantir que $BA = I$.*

Prova. A prova deste fato se baseia em um teorema da Álgebra Linear que estabelece o seguinte: Se as matrizes envolvidas forem de ordem n, o posto de I é n e, consequentemente o posto de A é n, estabelecendo uma bijeção em $\mathbf{C}^n$. Então B é necessariamente a bijeção inversa e $BA = I$. $\square$

Se A e B forem inversíveis então AB é inversível e $(AB)^{-1} = A^{-1}B^{-1}$. Este resultado pode ser generalizado. Se $A_1, \ldots, A_n$ forem inversíveis, então o produto $A_1 \cdots A_n$ é inversível e

$$(A_1 \cdots A_n)^{-1} = A_n^{-1} \cdots A_1^{-1}.$$

Se A for uma matriz inversível, então as equações $AX = B$ e $YA = C$ possuem solução única dadas por $X = A^{-1}B$ e $Y = CA^{-1}$.

Se A for uma matriz quadrada, define-se as potências inteiras de A por

$$\begin{aligned} A^0 &= I, \\ A^k &= A^{k-1}A, \\ A^{-k} &= \left(A^{-1}\right)^k = \left(A^k\right)^{-1}. \end{aligned}$$

para todo $k \geq 1$ inteiro.

O posto de uma matriz é o número de suas colunas que são linearmente independentes. A nulidade de uma matriz é a dimensão do seu núcleo.

Teorema A.4 *Seja A uma matriz $m \times n$. O posto de A mais a nulidade de A é igual a n.*

Teorema A.5 *O posto de uma matriz não se modifica se ela for multiplicada por uma matriz inversível.*

Teorema A.6 *Seja A uma matriz $m \times n$ de posto k. Existe uma matriz P de ordem n, e uma matriz Q de ordem m, ambas inversíveis e tais que $D = Q^{-1}AP$ é uma matriz diagonal onde os k primeiros elementos da diagonal são iguais a 1 e os demais são todos nulos.*

Teorema A.7 *Seja A uma matriz $m \times n$ de posto k. Existe uma matriz inversível Q de ordem m, tal que $A' = Q^{-1}A$ é uma matriz escalonada reduzida.*

A **transposta** da matriz $A = (a_{ij})$ de ordem m por n é a matriz $A^T = (b_{ij})$, de ordem n por m, onde $b_{ij} = a_{ji}$. Vale a propriedade

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

Teorema A.8 *O número de linhas linearmente independentes de uma matriz é igual ao número de colunas linearmente independentes.*

Prova. Seja $A' = Q^{-1}A$ a matriz escalonada reduzida do teorema anterior. O número de linhas não nulas é o número de linhas linearmente independentes em A'. Em A', as colunas linearmente independentes são aquelas que contém os pivôs. Logo, o número de linhas linearmente independentes de A' é igual ao número de colunas linearmente independentes. Como Q e Q^T são inversíveis, o posto de $A = QA'$ e o de $A^T = (A')^T Q^T$ são idênticos, mostrando que o número de linhas e o número de colunas linearmente independentes de A são iguais. $\square$

A.4 Operações elementares e matrizes elementares

Operações elementares sobre linhas

1. Permutar duas linhas.
2. Multiplicar uma linha de A por um escalar não nulo.
3. Adicionar a uma linha um múltiplo de outra linha.

Operações elementares sobre colunas são definidas de modo análogo.

As operações elementares podem ser executadas mediante o produto de matrizes elementares. A matriz que troca a linha i pela linha j é aquela obtida a partir da matriz identidade, trocando a linha i com a linha j. A matriz que multiplica a linha i de A por um escalar $k \neq 0$ é obtida a partir da identidade, trocando o elemento diagonal da linha i por k. A matriz que adiciona um múltiplo k da linha i à linha j é obtida a partir da matriz identidade, trocando o zero da linha i coluna j por k.

Se E for uma matriz elementar, EA realiza operações elementares sobre as linhas de A e AE realiza operações elementares sobre as colunas de A, como mostram os exemplos que seguem.

Se

$$E = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

então a matriz EA é obtida de A trocando a primeira linha com a segunda; AE é uma matriz obtida de A trocando a primeira coluna com segunda.

Se

$$E = \begin{pmatrix} \beta & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

então a matriz EA é obtida de A multiplicando a primeira linha por β; a matriz AE é obtida de A multiplicando a primeira coluna por β.

Se

$$E = \begin{pmatrix} 1 & \beta & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

então EA é uma matriz obtida de A adicionando β vezes a segunda linha à primeira; AE é uma matriz obtida de A adicionando β vezes a primeira coluna à segunda.

As matrizes elementares são inversíveis. Se uma matriz A for inversível e E é uma matriz elementar, então AE e EA são inversíveis.

Se uma coluna ou uma linha de uma matriz for identicamente nula, ela é singular. Se uma coluna de uma matriz for uma combinação linear das outras, a matriz é singular.

Teorema A.9 *Uma matriz quadrada A é inversível se e só se puder ser escrita como um produto matrizes elementares.*

Prova. Se A for o produto de matrizes elementares, ela é inversível pois as matrizes elementares são inversíveis. Vamos provar a recíproca.

Como $A = (a_{ij})$ é inversível, nenhuma de suas colunas é identicamente nula. Pelo menos um elemento da primeira coluna é diferente de zero. Se a_{11} for igual a zero, podemos permutar a primeira linha de A com outra cujo elemento da primeira coluna é diferente de zero. Denotemos ainda por a_{11} o elemento da primeira linha e primeira coluna da matriz transformada. Podemos dividir a primeira linha por a_{11} de modo que o elemento da primeira linha primeira coluna fique igual a 1. Agora, podemos adicionar às demais linhas de A múltiplos da primeira de modo que todos os elementos da primeira coluna, exceto o primeiro, fiquem iguais a zero. Esta matriz obtida de A através de operações elementares é inversível e será denotada por A_1.

Se todos os elementos da segunda coluna de A_1 da diagonal principal para baixo forem nulos, a segunda coluna de A_1 seria um múltiplo da primeira e esta matriz seria singular. Como ela não é singular, pelo menos um elemento da segunda coluna da diagonal principal para baixo é diferente de zero. Se necessário, trocamos a segunda linha com outra abaixo dela que possui elemento não nulo na segunda coluna. O elemento da segunda linha segunda coluna desta matriz assim transformada é não nulo e podemos dividir agora a segunda linha por ele. O elemento $(2, 2)$ fica igual a 1. Podemos agora adicionar às outras linhas múltiplos da segunda de modo a anular todos os demais elementos da segunda coluna. Observe que a primeira coluna não é modificada neste processo pois o elemento da primeira coluna da segunda linha é zero. Denominemos esta nova matriz de A_2. Ela foi obtida de A_1 a partir de operações elementares e, portanto, é inversível.

Continuando com este processo, chegamos à matriz identidade, aplicando transformações elementares sobre as linhas de A. Sejam E_1, E_2, $\ldots$, E_k as matrizes elementares que realizam estas operações. Então $E_k \cdots E_1 A = I$ e $A = (E_k \cdots E_1)^{-1} I = E_1^{-1} \cdots E_k^{-1}$. Como a inversa de uma matriz elementar é elementar, A é um produto de matrizes elementares. $\square$

Apêndice B

Determinante

B.1 Permutação

Uma função bijetora $\sigma : \{1,\, 2,\, \ldots,\, n\} \to \{1,\, 2,\, \ldots,\, n\}$ é chamada de **permutação** do conjunto $\{1,\, 2,\, \ldots,\, n\}$. Basta apresentar a ênupla ordenada $(\sigma(1),\, \ldots,\, \sigma(n))$ para estabelecer σ sem ambiguidade. A identidade $(1,\, 2,\, \ldots,\, n)$ é uma permutação. Sendo bijetora, a permutação é inversível e, se $\sigma(i) = j$, sua inversa σ^{-1} leva j em i.

Sejam j e k dois inteiros distintos no conjunto $\{1,\, 2,\, \ldots,\, n\}$. Uma permutação que leva j em k e k em j, mantendo fixos os outros inteiros, é chamada de **transposição**. Se τ for uma transposição, basta informar que $\tau(j) = k$ para inferir que $\tau(k) = j$ e que $\tau(i) = i$ para todo i diferente de j e k.

Toda permutação é a composição de um número finito de transposições. De fato, sejam τ_i, $i = 1,\, \ldots,\, n$ permutações que tanto pode ser uma transposição quanto uma identidade, definidas por

$$\begin{aligned} \tau_1(1) &= \sigma(1), \\ \tau_2\tau_1(2) &= \sigma(2), \\ &\ldots, \\ \tau_n(\tau_{n-1}\cdots\tau_2\tau_1(n)) &= \sigma(n). \end{aligned}$$

Estas equações definem $\tau_1,\, \tau_2,\, \ldots,\, \tau_n$ sem ambiguidade. Observe que, se $\sigma(1) = 1$, então τ_1 é a identidade. Se $\sigma(2) = \tau_1(2)$, τ_2 é a identidade. Em geral, para $k \geq 2$, sendo $\sigma(k) = \tau_{k-1}\cdots\tau_2\tau_1(k)$, então τ_k é a permutação identidade. Em particular, τ_n é sempre a identidade e foi colocada na composição apenas para ficarmos com um número exato de n permutações, entre transposições e identidades. A permutação σ é igual à composição $\tau_n\, \cdots\, \tau_2\, \tau_1$.

Retirando as identidades desta composição, vemos que σ é uma composição de permutações que, entretanto, não é única. Todavia, duas decomposição de σ em permutações terá ou um número par de fatores ou um número ímpar de fatores. Provaremos esta afirmação logo adiante.

Seja σ uma permutação de $\{1, 2, \ldots, n\}$. Se $i < j$ e $\sigma(i) > \sigma(j)$ diremos que o par (i, j) é uma **inversão** de σ. Definimos o **sinal** de σ do seguinte modo: Se o número de inversões de σ for par, seu sinal será $+1$. Se o número de inversões de σ for ímpar, seu sinal será -1. O sinal de σ será denotado por $sign(\sigma)$.

A permutação identidade não apresenta nenhuma inversão. Portanto, seu sinal é $+1$.

Teorema B.1 *Sejam σ_1 e σ_2 duas permutações de $\{1, 2, \ldots, n\}$. Então*

$$sign(\sigma_2\sigma_1) = sign(\sigma_2)sign(\sigma_1).$$

Prova. Observe a tabela que vem em seguida, onde $i < j$.

		Inversões		
		σ_1	σ_2	$\sigma_2\sigma_1$
$\sigma_1(i) < \sigma_1(j)$	$\sigma_2\sigma_1(i) < \sigma_2\sigma_1(j)$	0	0	0
$\sigma_1(i) < \sigma_1(j)$	$\sigma_2\sigma_1(i) > \sigma_2\sigma_1(j)$	0	1	1
$\sigma_1(i) > \sigma_1(j)$	$\sigma_2\sigma_1(i) < \sigma_2\sigma_1(j)$	1	1	0
$\sigma_1(i) > \sigma_1(j)$	$\sigma_2\sigma_1(i) > \sigma_2\sigma_1(j)$	1	0	1

Ela mostra que quando há uma inversão em $\sigma_2\sigma_1$ ou há uma inversão em σ_1 ou há uma em σ_2 mas não em ambas ao mesmo tempo. Quando não há inversão em $\sigma_2\sigma_1$ então não há inversão nem em σ_1 nem em σ_2 ou ambas apresentam uma inversão. Isto significa que o número de inversões de $\sigma_2\sigma_1$ e a soma do número de inversões em σ_1 e σ_2 têm a mesma paridade. Isto implica na igualdade dos sinais

$$sign(\sigma_2\sigma_1) = sign(\sigma_2)sign(\sigma_1).$$

□

Se uma permutação σ mantém um número k fixo, isto é, se $\sigma(k) = k$, as inversões envolvendo este número não precisam ser contadas no cálculo do sinal. O número de inversões (i, k), com $i < k$ é igual ao número de inversões (k, j) com $k < j$. Logo, o número mantido fixo pela permutação sempre leva a um número par de inversões. Esta observação é útil na prova do próximo teorema.

Teorema B.2 *O sinal de uma transposição é -1.*

Prova. Se a transposição levar i em j e j em i, de acordo com a observação feita acima, podemos ignorar as inversões relativas aos números mantidos fixos. Sobram apenas i e j, para os quais há uma inversão. Logo, o sinal da transposição é -1. □

Teorema B.3 *Toda permutação é uma composição de transposições. Esta composição não é única. Entretanto, o número de transposições ou é sempre par ou é sempre ímpar.*

Prova. O sinal de toda transposição é -1. Quando $sign(\sigma) = +1$, qualquer decomposição de σ em transposições tem um número par de fatores. Quando $sign(\sigma) = -1$, o número de transposições que a compõem é ímpar. □

Teorema B.4 *O sinal de uma permutação é igual ao sinal de sua inversa.*

Prova. Como $\sigma^{-1}\sigma$ é a identidade cujo sinal é $+1$, segue $sign(\sigma^{-1})sign(\sigma) = sign(\sigma^{-1}\sigma) = 1$. Logo, $sign(\sigma^{-1})$ e $sign(\sigma)$ são ambos iguais a $+1$ ou ambos iguais a -1. □

B.2 Determinante

Seja $A = (a_{ij})$ uma matriz quadrada de ordem n. O **determinante** de A é definido por

$$\det(A) = \sum_{\sigma} sign(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}$$

onde σ pecorre o conjunto de todas as permutações de $\{1, 2, \ldots, n\}$.

Cada permutação σ de $\{1, 2, \ldots, n\}$ possui inversa τ. Se $\sigma(i) = j$, então $\tau(j) = i$ e $a_{i\sigma(i)} = a_{\tau(j)j}$. Consequentemente, o produto $a_{1\sigma(1)}\ a_{2\sigma(2)} \cdots a_{n\sigma(n)}$ é uma reordenação $a_{\tau(1)1}\ a_{\tau(2)2} \cdots a_{\tau(n)n}$ e, portanto, são iguais. Como $sign(\sigma) = sign(\tau)$, segue

$$\det(A) = \sum_{\tau} sign(\tau) a_{\tau(1)1} a_{\tau(2)2} \cdots a_{\tau(n)n}$$

onde τ percorre o conjunto de todas as permutações de $\{1, 2, \ldots, n\}$.

Teorema B.5 *O determinante de uma matriz é igual ao determinante de sua transposta.*

Prova. Se $B = (b_{ij})$ for a transposta de $A = (a_{ij})$, então $b_{ij} = a_{ji}$. Assim,

$$\begin{aligned} \det(A) &= \sum_{\sigma} sign(\sigma) a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n} \\ &= \sum_{\sigma} sign(\sigma) b_{1\sigma(1)} b_{2\sigma(2)} \cdots b_{n\sigma(n)} = \det(B). \end{aligned}$$

□

Teorema B.6 *Se uma linha ou uma coluna de uma matriz quadrada for nula, seu determinante é zero.*

Prova. Quando a linha i for nula, $a_{i\sigma(i)} = 0$ para toda permutação σ e assim, $\det(A) = 0$. Uma coluna nula na matriz é uma linha nula na transposta. Assim, $\det(A^T) = 0$ e, portanto, $\det(A) = 0$. □

Teorema B.7 *Se permutarmos duas linhas de uma matriz, o determinante muda de sinal. Se permutarmos duas colunas de uma matriz, o determinante muda de sinal.*

Prova. Seja $B = (b_{ij})$ a matriz obtida de $A = (a_{ij})$ permutando-se as linhas r e s, de modo que $b_{rj} = a_{sj}$ e $b_{sj} = a_{rj}$. Assim,

$$\begin{aligned} \det(B) &= \sum_{\sigma} sign(\sigma) \cdots b_{r\sigma(r)} \cdots b_{s\sigma(s)} \cdots \\ &= \sum_{\sigma} sign(\sigma) \cdots a_{s\sigma(r)} \cdots a_{r\sigma(s)} \cdots \\ &= \sum_{\sigma} sign(\sigma) \cdots a_{r\sigma(s)} \cdots a_{s\sigma(r)} \cdots \\ &= -\sum_{\sigma\tau} sign(\sigma\tau) \cdots a_{r,\sigma\tau(r)} \cdots a_{s,\sigma\tau(s)} \cdots \end{aligned}$$

onde τ é a transposição que leva r em s e s em r. Como σ percorre todas as permutações possíveis, $\sigma\tau$ também as percorre e assim,

$$\det(B) = -\sum_{\sigma} sign(\sigma) \cdots a_{r\sigma(r)} \cdots a_{s\sigma(s)} \cdots = -\det(A).$$

□

Teorema B.8 *Se duas linhas ou duas colunas de uma matriz quadrada forem iguais, seu determinante é zero.*

Prova. Se duas linhas da matriz A são iguais, ao trocar uma linha pela outra, a matriz A permanece inalterada e seu determinante troca de sinal. Logo, $\det(A) = -\det(A)$, o que resulta em $\det(A) = 0$. □

Teorema B.9 *Seja* $A = [v_1, \ldots, v_j + w_j, \ldots, v_n]$, $B = [v_1, \ldots, v_j, \ldots, v_n]$, *e* $C = [v_1, \ldots, w_j, \ldots, v_n]$, *matrizes quadradas de ordem* n, *onde* $v_1, \ldots, v_n$ *e* w_j *são as colunas de* B *e* C. *A coluna* j *de* A *é* $v_j + w_j$. *Então*

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

Prova. Imediata, a partir da definição. □

Vale um teorema análogo se os elementos de uma linha de A forem decompostos em duas parcelas.

Teorema B.10 *Sejam* $v_1, \ldots, v_n$ *vetores coluna em* $\mathbf{C}^n$. *Então para todo escalar* β,

$$\det[v_1, \ldots, \beta v_i, \ldots, v_n] = \beta \det[v_1, \ldots, v_i, \ldots, v_n]$$

O mesmo resultado se aplica ao multiplicarmos uma linha de A *por um escalar* β.

Prova. Imediata a partir da definição. □

Corolário B.11 *Se A é uma matriz quadrada de ordem n e β um escalar,*

$$\det(\beta A) = \beta^n \det(A).$$

Teorema B.12 *Se uma linha de uma matriz quadrada A for um múltiplo de outra linha de A, então* $\det(A) = 0$.

Prova. Se $\beta \neq 0$, $\det[\ \ldots, v_i, \ldots, \beta v_i, \ldots\ ] = \beta \det[\ \ldots, v_i, \ldots, v_i, \ldots\ ] = 0$. Quando $\beta = 0$, uma linha da matriz é nula e $\det(A) = 0$. □

Teorema B.13 *O determinante de uma matriz não se altera se adicionarmos a uma de suas colunas um múltiplo de outra. O mesmo resultado se aplica se adicionarmos a uma de suas linhas um múltiplo de outra.*

Prova. Se $\beta \neq 0$, $\det[\ \ldots, v_i, \ldots, v_j + \beta v_i, \ldots\ ] = \det[\ \ldots, v_i, \ldots, v_j, \ldots\ ] + \beta \det[\ \ldots, v_i, \ldots, v_i, \ldots\ ] = \det[\ \ldots, v_i, \ldots, v_j, \ldots\ ]$. □

Teorema B.14 *Se $A = (a_{ij})$ for uma matriz quadrada triangular superior ou triangular inferior, então*

$$\det(A) = a_{11} a_{22} \cdots a_{nn}.$$

Prova. Se A for uma matriz quadrada de ordem n, triangular superior, então $a_{ij} = 0$ quando $i > j$. Sendo σ uma permutação de $\{1, 2, \ldots, n\}$ termo

$$a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}$$

será não nulo apenas quando $\sigma(1) \geq 1$, $\sigma(2) \geq 2, \ldots, \sigma(n) \geq n$. Isto só ocorre se $\sigma(n) = n$, $\ldots$, $\sigma(2) = 2$, $\sigma(1) = 1$. Daí, o único termo não nulo do determinante de A é $a_{11}\ a_{22} \cdots a_{nn}$. □

Corolário B.15 *O determinante da matriz identidade é igual a* 1.

Teorema B.16 *Seja A uma matriz quadrada. O* $\det(A) \neq 0$ *se e só se A for inversível.*

Prova. Se A for inversível, seja B a sua inversa. Como $AB = I$, $\det(A)\det(B) = 1$, provando que $\det(A) \neq 0$.

Quando A é inversível, suas colunas formam uma base da imagem indicando que suas colunas são vetores linearmente independentes.

Se A for singular, uma de suas linhas é combinação linear das outras e $\det(A) = 0$. □

Teorema B.17 *Se E for uma matriz elementar e A uma matriz quadrada, todas de ordem n, então* $\det(EA) = \det(E)\ \det(A)$.

Prova. Se E for uma matriz elementar que permuta a linha i com a linha j, então $\det(E) = -1$ e $\det(EA) = -\det(A)$, provando que $\det(EA) = \det(E)\ \det(A)$ para este caso.

Se E for uma matriz elementar que multiplica uma linha por r, então $\det(E) = r$ e $\det(EA) = r\det(A)$, provando que neste caso também vale o teorema.

Se E for uma matriz elementar que multiplica à linha i um múltiplo r da linha j, então $\det(E) = 1$ e $\det(EA) = \det(A)$, provando que o teorema também vale neste último caso, o que completa a prova do teorema. □

Corolário B.18 *Se $E_1,\ E_2,\ \ldots,\ E_k$ forem matrizes elementares e A for uma matriz quadrada, todas de ordem n, então* $\det(E_1\ E_2\ \cdots\ E_k\ A) = \det(E_1)\ \det(E_2)\ \cdots\ \det(E_k)\ \det(A)$.

Teorema B.19 *Se A e B forem matrizes quadradas de mesma ordem,*

$$\det(AB) = \det(A)\det(B).$$

Prova. Se A for inversível, $A = E_1\ E_2\ \cdots\ E_k$, onde $E_1,\ E_2,\ \ldots,\ E_k$ são matrizes elementares. Assim, $\det(AB) = \det(E_1\ E_2\ \cdots\ E_k\ B) = \det(E_1)\ \det(E_2)\ \cdots\ \det(E_k)\ \det(B) = \det(A)\ \det(B)$.

Se A ou B for singular, AB é singular e $\det(AB) = 0$ e $\det(A)\ \det(B) = 0$. □

Corolário B.20 *Se A for uma matriz quadrada inversível,* $\det(A^{-1}) = 1/\det(A)$.

Teorema B.21 *Matrizes quadradas semelhantes possuem o mesmo determinante.*

Prova. Se A e B forem matrizes quadradas semelhantes, então existe uma matriz inversível P de modo que $B = PAP^{-1}$ e $\det(B) = \det(P)\ \det(A)\ \det(P^{-1}) = \det(A)$. □

B.3 Cofator

Seja $A = (a_{ij})$ uma matriz quadrada de ordem n. Seu determinante é

$$\det(A) = \sum_{\sigma} sign(\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)}$$

onde o somatório $\sum_\sigma$ percorre todas as permutações do conjunto $S = \{1,\ 2,\ \ldots,\ n\}$. Podemos agrupar esta soma do seguinte modo: tomemos todas as permutações que levam 1 em 1, depois aquelas que levam 1 em 2 e assim por diante, até aquelas que levam 1 em

n. Nas permutações que levam 1 em 1 podemos colocar a_{11} em evidência; nas que levam 1 em 2, o a_{12} pode ser colocado em evidência e, naquelas que levam 1 em n podemos colocar o a_{1n} em evidência e escrever

$$\det(A) = a_{11}c_{11} + a_{12}c_{12} + a_{13}c_{13} + \cdots + a_{1n}c_{1n}.$$

O escalar c_{1j} é chamado de **cofator** de a_{1j}.

Vemos que $c_{11} = \sum_{\sigma(1)=1} sign(\sigma) a_{2\sigma(2)} \cdots a_{n\sigma(n)}$ onde a soma percorre todas as permutações que levam 1 em 1. A cada permutação σ em $\{1, 2, \ldots, n\}$ que mantém fixo o 1, corresponde a uma permutação π em $S' = \{2, 3, \ldots, n\}$, onde $\pi(i) = \sigma(i)$ para $i = 2, \ldots, n$. Ambas possuem o mesmo número de inversões e, portanto, possuem o mesmo sinal. Para estabelecer o sinal de uma permutação, as inversões de um ponto fixo não precisam ser contadas, uma vez que o número dessas inversões é um número par. Logo, $c_{11} = \sum_{\pi} sign(\pi) a_{2\pi(2)} \cdots a_{n\pi(n)}$ é o determinante de uma matriz que se obtém de A excluindo a primeira linha e a primeira coluna. Denotamos este determinante por A_{11}.

Em geral, vamos denotar por A_{ij} o determinante da matriz obtida quando se elimina a linha i e a coluna j de A.

Para determinar o termo c_{12} faz-se o seguinte raciocínio: Permutando a primeira coluna de A com a segunda, obtemos uma matriz $B = (b_{ij})$ onde $b_{11} = a_{12}$ e $\det(B) = -\det(A)$. Desta igualdade segue $b_{11}B_{11} + \cdots = -a_{12}c_{12} + \cdots$ e, como $a_{12} = b_{11}$, se conclui que $c_{12} = -B_{11}$. O escalar B_{11} é o determinante da matriz obtida de B ao excluir sua primeira linha e sua primeira coluna, que são a primeira linha e a segunda coluna de A. Este determinante foi denotado por A_{12}. Desta forma, $c_{12} = -A_{12}$.

O termo c_{13} pode ser obtido trazendo a terceira coluna para o lugar da primeira, fazendo duas permutações: basta trocar esta coluna sucessivamente com as que estão à sua esquerda até conduzi-la para a posição da primeira. Neste processo, o sinal do determinante da matriz se modifica duas vezes. O determinante da matriz final é igual ao de A. Por um raciocínio análogo ao anterior, conclui-se que $c_{13} = A_{13}$, onde A_{13} é o determinante da matriz obtida ao eliminar a primeira linha e a terceira coluna de A.

Prosseguindo com o raciocínio anterior, chega-se ao desenvolvimento

$$\det(A) = a_{11}c_{11} + a_{12}c_{12} + a_{13}c_{13} + \cdots + a_{1n}c_{1n}$$

onde $c_{1j} = (-1)^{1j} A_{1j}$ é o cofator de a_{1j} e A_{1j} é o determinante da matriz obtida ao eliminar a linha 1 e a coluna j de A. Esta fórmula desenvolve o determinante pela primeira linha e é conhecida por **desenvolvimento do determinante pela primeira linha**.

De modo semelhante, podemos desenvolver o determinante pela linha i, usando o argumento seguinte. O determinante de A é a soma de diversas parcelas, cada uma com n fatores. Dentre os fatores de uma parcela do determinante há um único elemento da linha i. Aquelas parcelas que possuem como fator um elemento da linha i coluna j, não contém como fator outro elemento da linha i nem outro elemento da coluna j. Nas parcelas que possuem o fator a_{ij}, vamos colocá-lo em evidência. Denotemos por c_{ij} o termo que fica multiplicado por a_{ij} e vamos chamá-lo de **cofator** de a_{ij}. Assim,

$$\det(A) = a_{ij}c_{ij} + \cdots ,$$

onde os três pontos se referem às parcelas que contém elementos da linha i e colunas distintas da j. Mediante transposição de linhas e colunas, podemos transformar a matriz A numa matriz B, onde o elemento a_{ij} fique posicionado na linha 1 coluna 1 de B. Basta transpor $i-1$ vezes a linha i com as que estão acima, até posicioná-la no topo da matriz. Em seguida, mediante $j-1$ transposições da coluna j com as que estão à sua esquerda, coloca-se o elemento a_{ij} na primeira posição da matriz. A cada transposição, o determinante muda de sinal. Como há um total de $(i-1)+(j-1)$ transposições, $\det(A) = (-1)^{(i-1)+(j-1)}\det(B) = (-1)^{i+j}\det(B)$. O determinante de B possuirá parcelas onde um dos fatores é o a_{ij}. Como a_{ij} ocupa a primeira linha primeira coluna de B, sabemos de antemão que $\det(B) = a_{ij}c_{ij} + \cdots$ onde c_{ij} é o determinante de uma matriz obtida de B pela eliminação de sua linha 1 coluna 1. Ora, a matriz obtida ao eliminar a linha 1 coluna 1 de B é igual à matriz obtida ao eliminar a linha i coluna j de A. Assim, $\det(A) = (-1)^{i+j}a_{ij}\det A_{ij}$, onde A_{ij} é a matriz obtida de A retirando sua linha i e sua coluna j, chamada de **menor** $(i,\, j)$ de A. Provamos que o cofator de a_{ij} no desenvolvimento do $\det(A)$ é $c_{ij} = (-1)^{i+j}\det A_{ij}$.

Na soma que define o determinante de A, podemos colocar em evidência os elementos $a_{i1},\, a_{i2},\, \ldots,\, a_{in}$. A parcela que contém um desses fatores não conterá os demais. Cada um deles será multiplicado pelo seu cofator e assim

$$\det(A) = a_{i1}c_{i1} + a_{i2}c_{i2} + a_{i3}c_{i3} + \cdots + a_{in}c_{in}$$

onde $c_{ij} = (-1)^{i+j}\det A(i,j)$ é o **cofator** de a_{ij}. Como os elementos $a_{i1},\, a_{i2},\, \ldots,\, a_{in}$ são todos da linha i, a fórmula acima é conhecida por desenvolvimento do determinante pela linha i.

Um argumento semelhante nos permite desenvolver o determinante pela coluna j. Obtemos então o desenvolvimento do determinante pela coluna j

$$\det(A) = a_{1j}c_{1j} + a_{2j}c_{2j} + a_{3j}c_{3j} + \cdots + a_{nj}c_{nj}$$

Um modo prático de utilizar estas fórmulas consiste em aplicar transformações elementares sobre a matriz zerando o maior número de elementos de uma linha ou de uma coluna e usar as fórmulas acima para reduzir o determinante original a uma soma de outros envolvendo matrizes de ordem $n-1$. Este processo pode ser utilizado mais de uma vez reduzindo sucessivamente a ordem das matrizes cujos determinantes precisam ser calculados.

Definição B.22 *Seja A uma matriz quadrada de ordem n. A matriz cujo elemento da linha i coluna j é c_{ji} (observe a ordem dos índices em que primeiro vem o j e depois o i) é chamada de matriz* ***adjunta clássica*** *de A e é denotada por $adj(A)$.*

Teorema B.23 *Se A for inversível, então $A \cdot adj(A) = adj(A) \cdot A = \det(A) \cdot I$.*

Prova. Provamos que

$$\sum_j a_{ij}c_{ij} = \det(A).$$

Se i for diferente de k, $\sum_j a_{kj}c_{ij}$ corresponderia ao determinante de uma matriz em que a linha i foi substituída pela linha k. As linhas i e k desta matriz seriam iguais e seu determinante seria igual a zero. Portanto, para $i \neq k$,

$$\sum_j a_{kj}c_{ij} = 0.$$

Podemos usar o delta de Kronecker para unificar estas duas expressões em destaque

$$\sum_j a_{kj}c_{ij} = \delta_{ik}\det(A).$$

Ora, o lado esquerdo desta expressão é o elemento da linha k coluna i da matriz $A \cdot adj(A)$ e o lado direito é o elemento da linha k coluna i da matriz $\det(A) \cdot I$, provando que $A \cdot adj(A) = \det(A) \cdot I$.

O lado esquerdo da expressão também é o elemento da linha i coluna k da matriz $adj(A) \cdot A$ e o lado direito é o elemento da linha i coluna k da matriz $\det(A) \cdot I$, provando que $adj(A) \cdot A = \det(A) \cdot I$. $\square$

B.4 Regra de Cramer

Consideremos um sistema de n equações com n incógnitas

$$\sum_{j=1}^{n} a_{ij}x_j = b_i,$$

para $i = 1, 2, \ldots, n$. Se a matriz $A = (a_{ij})$ for inversível, o sistema possui solução única. O método de Cramer fornece um meio de resolver o sistema usando determinantes. Ele é pouco eficiente e é usado apenas para sistemas pequenos.

Sendo c_{ij} os cofatores de a_{ij} então $\sum_i a_{ij}c_{ik} = \delta_{jk}\det(A)$

$$\sum_{i=1}^{n} c_{ik}b_i = \sum_{i=1}^{n} c_{ik}\left(\sum_{j=1}^{n} a_{ij}x_j\right) = \sum_{j=1}^{n}\left(\sum_{i=1}^{n} a_{ij}c_{ik}\right)x_j = \sum_{j=1}^{n}\det(A)\delta_{jk}x_j = \det(A)x_k$$

Dividindo pelo $\det(A)$ segue

$$x_k = \frac{\sum_{i=1}^{n} c_{ik}b_i}{\det(A)} = \frac{\Delta_k}{\det(A)},$$

onde $\Delta_k = \sum_{i=1}^{n} c_{ik}b_i$ é o determinante de uma matriz obtida de A, trocando-se sua coluna k por b

$$\Delta_k = \det\begin{bmatrix} a_{11} & \cdots & a_{1,k-1} & b_1 & a_{1,k+1} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2,k-1} & b_2 & a_{2,k+1} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{n,k-1} & b_n & a_{n,k+1} & \cdots & a_{nn} \end{bmatrix}$$

B.5 Determinante de Vandermonde

Sejam x_1, ..., x_n números reais. O número real

$$V_n(x_1, x_2, \ldots, x_n) = \det \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{bmatrix}$$

é chamado de **determinante de Vandermonde**. Vamos mostrar que

$$V_n(x_1, \ldots, x_n) = \alpha(x_n - x_1)(x_n - x_2) \cdots (x_n - x_{n-1}).$$

Desenvolvendo $V_n(x_1, \ldots, x_n)$ pela última linha, vemos que α é o cofator de x_n^{n-1} que é igual ao determinante de Vandermonde de ordem inferior

$$\alpha = V_{n-1}(x_1, \ldots, x_{n-1}).$$

Calculando os determinantes de Vandermonde para $n = 2$ e $n = 3$, obtemos $V_2(x_1, x_2) = x_2 - x_1$ e $V_3(x_1, x_2, x_3) = V_2(x_1, x_2)(x_3 - x_1)(x_3 - x_2) = (x_2 - x_1)\ (x_3 - x_1)(x_3 - x_2)$.

Vamos provar por indução que

$$V_n(x_1, \ldots, x_n) = \prod_{i<j} (x_j - x_i),$$

adotando como hipótese de indução a validade de

$$V_{n-1}(x_1, \ldots, x_{n-1}) = \prod_{i<j<n} (x_j - x_i).$$

Daí,

$$\begin{aligned} V_n(x_1, \ldots, x_n) &= V_{n-1}(x_1, \ldots, x_{n-1}) \prod_{i<n} (x_n - x_i) \\ &= \prod_{i<j<n} (x_j - x_i) \prod_{i<n} (x_n - x_i) \\ &= \prod_{i<j} (x_j - x_i) \end{aligned}$$

Esta igualdade vale mesmo quando os x_i não forem distintos dois a dois, caso em que o determinante de Vandermonde é nulo.

Trocando o número x_n pela variável x, $V_n(x_1, \ldots, x_{n-1}, x)$ se torna um polinômio em x de grau $n - 1$ que possui $x_1, \ldots, x_{n-1}$ como raízes. Assim,

$$V_n(x_1, \ldots, x_{n-1}, x) = \alpha(x - x_1)(x - x_2) \cdots (x - x_{n-1}),$$

onde α é um número real que não depende de x.

B.6 Determinante, uma definição alternativa

Seja $A = (a_{ij})$ uma matriz $m \times n$ e $v_1 = (a_{1j})$, $v_2 = (a_{2j})$ …, $v_n = (a_{nj})$ suas colunas. Usaremos a notação $[v_1, \ldots, v_n]$ para designar a matriz A. O **determinante** de A é o número real denotado por $\det(A)$ ou $\det A$ com as propriedades abaixo.

1. Multilinearidade

$$\det[\ldots, \alpha v_i + \beta w_i, \ldots] = \alpha \det[\ldots, v_i, \ldots] + \beta \det[\ldots, w_i, , \ldots]$$

2. Alternatividade

$$\det[\ldots, v_i, \ldots, v_j, \ldots] = -\det[\ldots, v_j, \ldots, v_i, \ldots].$$

3. Normalização

$$\det I = \det[e_1, \ldots, e_n] = 1,$$

onde I é a matriz identidade de ordem n e $e_j = (0, \ldots, 0, 1, 0, \ldots, 0)^T$ é a matriz coluna $n \times 1$ cujo único elemento não nulo é o da linha j, que é igual a 1.

Pela alternatividade, se uma coluna da matriz for nula ou se duas colunas forem iguais, seu determinante é nulo. Agora, usando a multilinearidade e a observação acima, se α for um escalar,

$$\det(\ldots, v_i, \ldots, v_j + \alpha v_i, \ldots) = \det(\ldots, v_i, \ldots, v_j, \ldots).$$

Daí fica evidente que, se uma coluna for uma combinação linear das demais, então seu determinante é nulo. Se σ for uma permutação de $\{1, 2, \ldots, n\}$ então a alternatividade garante que

$$\det[e_{\sigma(1)}, e_{\sigma(2)}, \ldots, e_{\sigma(n)}] = sign(\sigma)$$

pois a cada inversão de colunas há uma troca de sinal. Observando que $v_j = \sum_{i_j=1}^{n} a_{i_j j} e_{i_j}$ obtemos

$$\det(A) = \sum_{i_1} \sum_{i_2} \cdots \sum_{i_n} a_{i_1 1} a_{i_2 2} \cdots a_{i_n n} \det[e_{i_1}, e_{i_2}, \ldots, e_{i_n}].$$

Se dois índices em $\det[e_{i_1}, e_{i_2}, \ldots, e_{i_n}]$ forem iguais, o determinante é nulo. Logo, *as* parcelas não nulas são aquelas em que $\{i_1, i_2, \ldots, i_n\}$ for uma permutação σ de $\{1, 2, \ldots, n\}$ e assim,

$$\begin{aligned} \det(A) &= \sum_{i_1} \sum_{i_2} \cdots \sum_{i_n} sign(i_1, i_2, \ldots, i_n) a_{i_1 1} a_{i_2 2} \cdots a_{i_n n} \\ &= \sum_{\sigma} sign(\sigma) a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n} \end{aligned}$$

onde o somatório percorre todas as permutação σ de $\{1, 2, \ldots, n\}$. Esta é exatamente a definição anterior. A fórmula acima ainda indica que as três propriedades enumeradas na definição de determinante são suficientes para garantir a existência e unicidade do determinante.

Referências Bibliográficas

[CaDoCo] Callioli, C.A., Domingues, H.H., e Costa R.C.F., Álgebra Linear e Aplicações. Editora Atual.

[Franklin] Joel N. Franklin, Matrix Theory, Dover publications, Inc., 1993.

[Hoffman] Hoffmann & Kunze, Álgebra Linear. Editora da USP com Editora Polígono.

[Kolman] Bernard Kolman, Introdução à Álgebra Linear com Aplicações, sexta edição. Editora LTC, 1998.

[Lang] Serge Lang, Álgebra Linear. Editora Edgard Blücher.

[Lawson] Terry Lawson, Álgebra Linear. Editora Edgard Blücher, 1997. Acompanham este livro: Matlab Labs for Linear Algebra, Mathematica Labs for Linear Algebra, Maple Labs for Linear Algebra.

[Lipschutz] Seymour Lipschutz, Álgebra Linear. Coleção Schaum. Makron Books.

[Nering] Evar D. Nering, Linear Algebra and Matrix Theory. John Wiley & Sons, 1970.

[Nicholson] W. Keith Nicholson, Álgebra Linear, segunda ed. McGraw-Hill, 2004.

[NobDan] Ben Noble & James W. Daniel, Álgebra Linear Aplicada. Guanabara Koogan.

[Poole] David Poole (Foi traduzido), Linear Algebra: A Modern Introduction (with CD-ROM) 2ed. Brooks Cole 2005.

[RoHo] Chris Rorres e Howard Anton, Álgebra Linear com Aplicações, oitava edição. Editora Bookman, 2001.

[TrefBau] Lloyd N. Trefethen and David Bau III, Numerical Linear Algebra. SIAM, Society for Industrial and Applied Mathematics, 1997.