Ferramentas de Análise Matricial para Aplicações em Engenharia Elétrica

Vítor H. Nascimento

6 de março de 2015

Capítulo 1

Introdução

O número de aplicações diretas de resultados de cálculo matricial e álgebra linear em engenharia elétrica é muito maior do que normalmente acredita um aluno de graduação de um curso de engenharia. A maior parte dos alunos crê que esse é "apenas mais um curso de matemática", e não vê a hora de começar a estudar algo "realmente importante" para a sua futura carreira profissional. Era assim quando comecei a graduação, e continua assim até hoje, pelas respostas que recebo sempre que converso sobre isso com meus alunos.

Essa impressão equivocada é um tanto prejudicial, pois faz com que os alunos não dêem a atenção devida ao curso de álgebra linear. Isso acaba fazendo falta nos anos seguintes, já que algumas técnicas modernas usadas em engenharia elétrica são essencialmente aplicações de resultados relativamente simples de álgebra linear. Conceitos como autovalores, espaço-imagem e espaço-nulo, pseudo-inversas, valores singu-

lares, subespaços, são utilizados no dia-a-dia de pesquisadores de áreas como processamento de sinais, controle e automação e telecomunicações.

O objetivo deste curso é duplo: de um lado, apresentar os conceitos e resultados mais importantes de álgebra linear que são usados em engenharia elétrica ao nível de pós-graduação, desde uma rápida recordação das definições básicas até resultados mais avançados, com exemplos de aplicação sempre que possível. O segundo objetivo é que os alunos aprendam a fazer demonstrações matemáticas.

A motivação por trás do primeiro objetivo é que um aluno de pós-graduação e futuro pesquisador deve ser capaz de ler um artigo sobre o seu assunto de trabalho sem tropeçar na parte matemática mais básica: com um pouco de reflexão, não é difícil entender, por exemplo, que toda matriz de autocorrelação $R \in \mathbb{C}^{M\times M}$ é tal que $\boldsymbol{x}^H R\boldsymbol{x} \geq 0$ para todo $\boldsymbol{x} \in \mathbb{C}^M$, e que isso implica que um sistema dinâmico $\boldsymbol{z}(n+1) = R\boldsymbol{z}(n)$ não pode ser um oscilador. Conclusões como essa decorrem facilmente de propriedades simples de matrizes positivas-definidas. O desconhecimento dessas propriedades pode fazer com que um texto que poderia ser facilmente compreendido em 15 minutos vire um profundo mistério.

Para o segundo objetivo a motivação é facilitar que o aluno desenvolva seus próprios trabalhos de pesquisa: demonstrações são úteis para analisar um método ou algoritmo que se propõe a resolver um determinado problema; ou para convencer outros pesquisadores que um novo método realmente funciona. Muitos alunos parecem ter um certo medo de demonstrações, pois não há em geral "receitas de bolo" para demonstrar teoremas. Mas há diversas técnicas que podem ser aplicadas, e que tendem a funcionar com certos tipos de problemas. Para aprendê-las, é necessário ver alguns exemplos, e testá-las em vários exercícios, para se começar a ter uma idéia de quando usar cada técnica. Logicamente não há técnica que funcione sempre, por isso a demonstração de teore-

mas é uma das oportunidades que o pesquisador tem de ser criativo e se divertir com o seu trabalho[1].

Cálculo matricial $\times$ Álgebra linear Às vezes um pesquisador pode usar resultados sobre matrizes em seu dia-a-dia e esquecer que teoria de matrizes é álgebra linear para espaços de dimensão finita (para não falar das matrizes infinitas). Como diversos programas e algoritmos de engenharia são baseados em propriedades de matrizes, como autovalores, espaço-nulo ou espaço-imagem, normas, é necessário que o pesquisador tenha uma boa compreensão desses conceitos para que entenda as limitações dos algoritmos existentes, e assim possa criar alternativas melhores, ou alterar esses algoritmos para resolver novos problemas.

Apesar de várias dessas propriedades poderem ser definidas e entendidas diretamente a partir de operações sobre os elementos de uma dada matriz, em muitas situações é mais fácil lembrar da relação entre matrizes e transformações lineares, e pensar nas propriedades (autovalores, etc) em termos de transformações lineares (veremos isso logo ao tratar de condições para testar a invertibilidade de matrizes).

Por esse motivo, vamos começar o estudo recordando a relação estreita entre matrizes e transformações lineares em espaços vetoriais, primeiramente relembrando a definição de espaços vetoriais e suas principais propriedades (capítulo 2). Depois faremos uma recordação de transformações lineares em geral (capítulo 3). A maior parte do material nestes dois primeiros capítulos provavelmente já terá sido vista alguma vez, nos primeiros anos do curso de graduação.

[1] Sim, isso é possível! Apesar de normalmente o pesquisador gastar 99% do tempo arrancando os cabelos tentando entender por que raios X ou Y não funciona como devia.

Capítulo 2

Espaços vetoriais

Neste capítulo vamos rever conceitos básicos de álgebra linear: vetores, espaços e subespaços vetoriais, independência linear, bases, ortogonalidade, mudança de base. O material coberto aqui é tratado mais vagarosamente em inúmeros livros sobre álgebra linear, como [9, 13], para citar apenas alguns bons exemplos.

O aluno deve procurar entender bem os conceitos apresentados neste capítulo, e, caso necessário, consultar as referências citadas ou mesmo o livro que usou quando estudou a matéria pela primeira vez.

2.1 Espaços vetoriais

Vamos começar definindo o que pode ser considerado um escalar, e o que é um espaço vetorial. A partir dessas definições, veremos nos exemplos

abaixo que uma variedade bem grande de conjuntos pode ser estudada a partir dos conceitos de álgebra linear.

Os escalares podem ser retirados de qualquer conjunto de elementos que satisfaça a seguinte definição:

Definição 2.1. Um *corpo* é um conjunto $\mathbb{F}$ de elementos (*números*) com uma operação de *adição* $(+)$ e outra de *multiplicação* $(\cdot)$ que satisfaçam as seguintes propriedades (note a semelhança com a definição de espaço vetorial abaixo):

Se a, b e c pertencem a $\mathbb{F}$, então

- Propriedade comutativa: $a + b = b + a$, $a \cdot b = b \cdot a$,
- Propriedade associativa: $a+(b+c) = (a+b)+c$, $a\cdot(b\cdot c) = (a\cdot b)\cdot c$,
- Identidade da adição (elemento 0): existe um elemento $0 \in \mathbb{F}$ tal que $a + 0 = 0 + a = a$ para todo $a \in \mathbb{F}$,
- Identidade da multiplicação (elemento 1): existe um elemento $1 \in \mathbb{F}$ tal que $a \cdot 1 = 1 \cdot a = a$ para todo $a \in \mathbb{F}$,
- Inversa da adição: para todo $a \in \mathbb{F}$ existe um elemento $-a$ tal que $a + (-a) = 0$,
- Inversa da multiplicação: para todo $a \in \mathbb{F}$, $a \neq 0$, existe um elemento a^{-1} tal que $a \cdot a^{-1} = 1$,
- Propriedade distributiva: $(a + b) \cdot c = (a \cdot c) + (b \cdot c)$.

Os exemplos mais comuns de corpos são:

1. Os números racionais $\mathbb{Q}$,

2. Os números reais $\mathbb{R}$,

3. Os números complexos $\mathbb{C}$,

4. Os números inteiros módulo um número primo (por exemplo, os inteiros módulo 3 são os números 0, 1 e 2, e soma e multiplicação são obtidas da soma e multiplicação normais, mas tirando o resto do resultado dividido por 3 — assim $1 + 2 \mod 3 = 0$, $2 \cdot 2 \mod 3 = 1$),

5. O conjunto {0,1} com a álgebra booleana, com a operação *OU EXCLUSIVO* fazendo o papel de $+$, e a operação *E* fazendo o papel de $\cdot$ (este exemplo é equivalente ao dos inteiros módulo 2). Esse conjunto e o anterior são muito utilizados para o estudo de codificação.

Observação 2.1. *Repare que o conjunto dos números inteiros,* $\mathbb{Z}$, não *forma um corpo, pois os únicos números inteiros que possuem inversa multiplicativa também inteira são* $+1$ *e* -1. *No entanto, os inteiros módulo um número primo formam um corpo: No conjunto dos inteiros módulo 3, por exemplo, como vimos acima* $2^{-1} = 2$, *pois* $2 \cdot 2 \mod 3 = 1$.

Com a definição de corpo, podemos agora definir um espaço vetorial:

Definição 2.2. Um *espaço vetorial* V sobre um corpo $\mathbb{F}$ é composto por

- um conjunto $\mathcal{V}$, com elementos chamados vetores, um dos quais é o elemento $\mathbf{0}$ (vetor nulo),

- uma operação binária $+$, chamada *adição vetorial*,

- uma operação chamada *multiplicação escalar* de vetores por elementos de $\mathbb{F}$,

que satisfazem as seguintes condições:

1. Para todos $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$

$$\begin{aligned} \boldsymbol{u} + \boldsymbol{v} &\in V, & \mathbf{0} + \boldsymbol{v} &= \boldsymbol{v}, \\ \boldsymbol{u} + \boldsymbol{v} &= \boldsymbol{v} + \boldsymbol{u}, & \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) &= (\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w}; \end{aligned}$$

2. Para todos $\boldsymbol{u}, \boldsymbol{v} \in V$ e todos $r, s \in \mathbb{F}$,

$$\begin{aligned} r\boldsymbol{v} &\in V, & 1\boldsymbol{v} &= \boldsymbol{v}, \\ 0\boldsymbol{v} &= \mathbf{0}, & r(\boldsymbol{u} + \boldsymbol{v}) &= r\boldsymbol{u} + r\boldsymbol{v}, \\ (r + s)\boldsymbol{v} &= r\boldsymbol{v} + s\boldsymbol{v}, & (rs)\boldsymbol{v} &= r(s\boldsymbol{v}). \end{aligned}$$

Vetores (ou seja, elementos de $\mathcal{V}$) serão sempre denotados por letras minúsculas em negrito, como $\boldsymbol{v}$, e escalares (ou seja, elementos de $\mathbb{F}$) serão denotados por letras minúsculas romanas ou gregas.

O conceito de espaço vetorial é útil porque vários entes matemáticos diferentes satisfazem as propriedades da Definição 2.2, por exemplo:

1. O conjunto $V_1 = \mathbb{R}^3$ das seqüências ordenadas de números reais (r_1, r_2, r_3), usadas em geometria analítica. Aqui, um vetor é um conjunto ordenado de três números reais, sendo que o vetor nulo é o elemento $\mathbf{0} = (0, 0, 0)$. As operações de adição e multiplicação por escalar são definidas da maneira usual:

$$\begin{aligned} (a_1, a_2, a_3) + (b_1, b_2, b_3) &= (a_1 + b_1, a_2 + b_2, a_3 + b_3), \\ \alpha(a_1, a_2, a_3) &= (\alpha a_1, \alpha a_2, \alpha a_3). \end{aligned}$$

   Observe que todas as propriedades listadas na Definição 2.2 são satisfeitas, e portanto V_1 é um espaço vetorial definido sobre o corpo dos números reais.

2. Note que o conjunto dos escalares usados é parte integrante da definição do espaço vetorial: todos os elementos $\alpha(1,2,3)$ pertencem também a V_1 se $\alpha \in \mathbb{R}$, mas $(1+j)(1,2,3) \notin V_1$. Para usar os números complexos como escalares, o conjunto de vetores deveria ser $V_2' = \mathbb{C}^3$ para que todas as propriedades fossem satisfeitas[1].

3. O conjunto V_3 das funções reais contínuas definidas no intervalo $(0,1)$, juntamente com os números reais.

4. O conjunto V_4 das variáveis aleatórias com valores em $\mathbb{R}$, média nula e variância finita, com os reais como escalares, forma também um espaço vetorial.

5. O conjunto de seqüências infinitas $s_n \in \mathbb{R}$, $n = 0, 1, \ldots,$ juntamente com os números reais como escalares, forma outro espaço vetorial, V_5. Um outro espaço vetorial é formado pelo conjunto das seqüências com valores complexos e os números complexos, V_5'.

6. O conjunto de polinômios de grau finito com coeficientes reais e os números reais formam um outro exemplo de espaço vetorial, V_6.

7. O conjunto de polinômios de grau menor ou igual a 5 e coeficientes reais, junto com $\mathbb{R}$, também forma um espaço vetorial V_7.

8. O conjunto das soluções da equação de diferenças homogênea

$$y(n) + a_1 y(n-1) + a_2 y(n-2) = 0,\ n \geq 0,$$

[1] As propriedades gerais de espaços vetoriais em geral não dependem do conjunto de escalares utilizado, desde que as condições da Definição 2.2 sejam satisfeitas. Por isso usamos um símbolo genérico $\mathbb{F}$ para o conjunto dos escalares na Definição 2.2.

forma, juntamente com os números complexos, um espaço vetorial V_8.

Pelos exemplos acima você pode notar que coisas bem diferentes podem ser vistas como "vetores" em um espaço vetorial, e portanto podem estudadas do ponto de vista de álgebra linear. Desta observação decorre a grande utilidade da álgebra linear: você pode estudar um ente bastante abstrato, "espaço vetorial", definido apenas pelas propriedades da Def. 2.2, sem pensar explicitamente em nenhum dos exemplos acima. Propriedades (teoremas) que você encontrar para um espaço vetorial genérico valem também para todos os casos específicos dos exemplos acima, ou outros que você provavelmente encontrará pela frente[2].

Pouco atrás eu disse que o poder da álgebra linear é a capacidade de abstrair de um pequeno conjunto de regras, que valem para um grande número de situações bem diferentes, uma série de resultados muito úteis (espero que você termine este curso convencido dessa última afirmação!). Isso não significa que não é necessário usar as propriedades específicas de cada caso: cada espaço vetorial tem características próprias, além daquelas que decorrem da estrutura de espaço vetorial.

É importante que, ao longo do estudo, você se acostume a interpretar o que significa cada teorema ou propriedade em termos de algum exemplo mais concreto. Pensar em alguns exemplos específicos é muito útil para obter intuição; é muito mais fácil entender o conceito de ortogonalidade pensando no $\mathbb{R}^3$ do que no espaço V_5, por exemplo.

[2]Algumas propriedades dependem de uma quantidade conhecida como *dimensão* de um espaço vetorial. A dimensão é uma propriedade de um espaço vetorial específico (por exemplo, V_1, V_2 e V_8 têm dimensão 3, V_7 tem dimensão 6, e V_3, V_4, V_5, V_5' e V_6 têm dimensão infinita). Quase sempre importa apenas saber se a dimensão é finita ou não — alguns teoremas só valem para espaços de dimensão finita. O conceito de dimensão será visto na Seção 2.5.

Esporadicamente pode acontecer de você precisar verificar se um conjunto (juntamente com um conjunto de escalares) realmente forma um espaço vetorial. Para isto, você precisa verificar se todas as propriedades da Def. 2.2 são satisfeitas. A rigor é necessário verificar *todas* as propriedades, mas na maioria dos casos práticos a existência do elemento $\mathbf{0}$, a identidade de multiplicação por escalar e as propriedades comutativas e associativas decorrem facilmente das propriedades dos números reais e complexos e da definição de adição de vetores, isto é, se $\boldsymbol{u}$, $\boldsymbol{v} \in V$, r, $s \in \mathbb{F}$, é usualmente fácil verificar que

$$\begin{aligned} &\boldsymbol{u}+\boldsymbol{v}=\boldsymbol{v}+\boldsymbol{u}, && \boldsymbol{u}+(\boldsymbol{v}+\boldsymbol{w})=(\boldsymbol{u}+\boldsymbol{v})+\boldsymbol{w}, && (r+s)\boldsymbol{v}=r\boldsymbol{v}+s\boldsymbol{v}, \\ &\mathbf{0}+\boldsymbol{v}=\boldsymbol{v}, && r(\boldsymbol{u}+\boldsymbol{v})=r\boldsymbol{u}+r\boldsymbol{v}, && (rs)\boldsymbol{v}=r(s\boldsymbol{v}), \\ &1\boldsymbol{v}=\boldsymbol{v}, && 0\boldsymbol{v}=\mathbf{0}. \end{aligned}$$

O trabalho mais difícil quase sempre é provar que, para quaisquer $\boldsymbol{u}$, $\boldsymbol{v} \in V$, $r \in \mathbb{F}$,

$$\boldsymbol{u}+\boldsymbol{v} \in V, \qquad\qquad r\boldsymbol{v} \in V.$$

É melhor ver alguns exemplos:

Exemplo 2.1. Prove que V_6 é um espaço vetorial.

Para resolver um problema deste tipo, a primeira coisa a fazer é recordar a definição de soma de vetores e multiplicação por escalar para o conjunto de vetores em questão. Neste caso, os vetores são todos os polinômios de grau finito e com coeficientes reais, e a operação de adição de vetores é a adição usual de dois polinômios, ou seja, a soma de

$$p_1 = a_{1,0} + a_{1,1}x + a_{1,2}x^2 + \cdots + a_{1,G_1}x^{G_1} \in V_6$$

e

$$p_2 = a_{2,0} + a_{2,1}x + a_{2,2}x^2 + \cdots + a_{2,G_2}x^{G_2} \in V_6$$

é um polinômio

$$p_3 = (a_{1,0} + a_{2,0}) + (a_{1,1} + a_{2,1})x + \cdots \in V_6.$$

A multiplicação de p_1 por um escalar $r \in \mathbb{R}$ é o polinômio

$$p_4 = (ra_{1,0}) + (ra_{1,1})x + (ra_{1,2})x^2 + \cdots + (ra_{1,G_1})x^{G_1}.$$

Vamos agora verificar todas as propriedades, uma por uma:

1. Propriedade comutativa na soma vetorial ($\boldsymbol{u}+\boldsymbol{v} = \boldsymbol{v}+\boldsymbol{u}$): decorre do fato de $a_{1,n}$ e $a_{2,n}$ serem números reais, e portanto $a_{1,n}+a_{2,n} = a_{2,n} + a_{1,n}$. Aplicando esta propriedade para todos os termos de p_3, concluímos que $p_1 + p_2 = p_2 + p_1$, para quaisquer $p_1,\ p_2 \in V_6$.

2. Propriedade associativa na soma vetorial ($(\boldsymbol{u}+\boldsymbol{v})+\boldsymbol{w} = \boldsymbol{u}+(\boldsymbol{v}+\boldsymbol{w})$): novamente decorre da propriedade equivalente dos números reais. Definindo $p_5 = a_{5,0} + \cdots + a_{5,G_5}x^{G_5}$, temos (para facilitar, vamos assumir que o grau de p_1 é maior ou igual ao grau de p_2, e que o grau de p_2 é maior ou igual ao grau de p_5. Vamos chamar também, para simplificar a notação, $a_{2,G_2+1} = a_{2,G_2+2} = \cdots =$

$a_{2,G_1} = a_{5,G_5+1} = \cdots = a_{5,G_1} = 0)$[3]:

$$
\begin{aligned}
(p_1 + p_2) + p_5 = &\Big((a_{1,0} + a_{2,0}) + a_{5,0}\Big) \\
&+ \Big((a_{1,1} + a_{2,1})x + a_{5,1}x\Big) + \cdots + \\
&+ \Big((a_{1,G_1} + a_{2,G_1})x^{G_1} + a_{5,G_1}x^{G_1}\Big) = \\
&= \Big(a_{1,0} + (a_{2,0} + a_{5,0})\Big) \\
&+ \Big(a_{1,1}x + (a_{2,1} + a_{5,1})x\Big) + \cdots + \\
&+ \Big(a_{1,G_1}x^{G_1} + (a_{2,G_1} + a_{5,G_1})x^{G_1}\Big) = p_1 + (p_2 + p_5).
\end{aligned}
$$

3. O vetor $\mathbf{0}$ é o polinômio identicamente nulo, $p_0(x) \equiv 0$, pois $p_0 + p_1 = p_1$ para qualquer $p_1 \in V_6$.

4. Propriedade aditiva ($\boldsymbol{u} + \boldsymbol{v} \in V_6$ para todos $\boldsymbol{u}, \boldsymbol{v} \in V_6$): Considere

[3]Note que esta hipótese serve apenas para facilitar a explicação, para não termos de considerar vários casos diferentes — se o grau de p_3 fosse o maior de todos, a explicação não mudaria de maneira significativa. Quando se faz alguma simplificação deste tipo, diz-se algo como "assumindo, sem perda de generalidade, que ..." Cuidado ao usar essa expressão: só a use se todos os casos forem realmente equivalentes, ou seja, se nenhum passo do seu raciocínio precisasse ser alterado por uma escolha diferente!

$p_3 = p_1 + p_2$, como acima. O último termo de p_3 é

$$\begin{array}{rcl} a_{1,G_1} x^{G_1} & \text{se} & G_1 > G_2, \\ a_{2,G_2} x^{G_2} & \text{se} & G_1 < G_2, \\ (a_{1,G_1} + a_{2,G_1})\, x^{G_1} & \text{se} & G_1 = G_2. \end{array}$$

Em todos os casos, p_3 é um polinômio de coeficientes reais e grau igual a $\max\{G_1, G_2\} < \infty$. Portanto, $p_3 \in V_6$, ou seja, provamos que

$$p_1 \in V_6,\ p_2 \in V_6 \Rightarrow p_1 + p_2 \in V_6.$$

As outras propriedades são demonstradas de maneira similar. $\diamondsuit$

Exercício 2.1. Complete a prova do exemplo 2.1.

Exemplo 2.2. Prove que V_4 é um espaço vetorial.

É necessário novamente verificar todas as propriedades da Def. 2.2. Nós vamos verificar aqui apenas a propriedade da soma, as outras ficam como exercício. Sejam então $\boldsymbol{x}$ e $\boldsymbol{y}$ duas variáveis aleatórias em V_4, ou seja, com média nula e variância finita. Então $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{y}$ também tem média nula, pois

$$\mathrm{E}\,\boldsymbol{z} = \mathrm{E}(\boldsymbol{x} + \boldsymbol{y}) = \mathrm{E}\,\boldsymbol{x} + \mathrm{E}\,\boldsymbol{y} = 0.$$

Como a média de $\boldsymbol{z}$ é zero, sua variância é $\mathrm{E}\,\boldsymbol{z}^2$, e portanto

$$\sigma_z^2 = \mathrm{E}\,\boldsymbol{z}^2 = \mathrm{E}\,\boldsymbol{x}^2 + 2\,\mathrm{E}(\boldsymbol{xy}) + \mathrm{E}\,\boldsymbol{y}^2 = \sigma_x^2 + \sigma_y^2 + 2\,\mathrm{E}(\boldsymbol{xy}).$$

Como $\boldsymbol{x}$, $\boldsymbol{y} \in V_4$, σ_x^2 e σ_y^2 são finitas. Para provar que σ_z^2 é finita, basta portanto mostrar que $\mathrm{E}(\boldsymbol{xy})$ é finita.

Vamos provar esse resultado usando um resultado auxiliar: considere dois números reais quaisquer a e b. Temos então

$$\begin{aligned} 0 \leq (a+b)^2 = a^2 + 2ab + b^2 \qquad &\Rightarrow \qquad -\frac{a^2}{2} - \frac{b^2}{2} \leq ab, \\ 0 \leq (a-b)^2 = a^2 - 2ab + b^2 \qquad &\Rightarrow \qquad ab \leq \frac{a^2}{2} + \frac{b^2}{2} \end{aligned}$$

e portanto

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

Como esse resultado vale para quaisquer números a e b reais, vale também para os resultados das variáveis aleatórias $\boldsymbol{x}$ e $\boldsymbol{y}$, então

$$-\frac{\sigma_x^2}{2} - \frac{\sigma_y^2}{2} = -\operatorname{E}\left(\frac{\boldsymbol{x}^2}{2} + \frac{\boldsymbol{y}^2}{2}\right) \leq \operatorname{E}(\boldsymbol{xy}) \leq \operatorname{E}\left(\frac{\boldsymbol{x}^2}{2} + \frac{\boldsymbol{y}^2}{2}\right) = \frac{\sigma_x^2}{2} + \frac{\sigma_y^2}{2},$$

pois é fácil ver que, se a variável aleatória $\boldsymbol{w}$ é sempre maior ou igual a $\boldsymbol{t}$, então $\operatorname{E}\boldsymbol{w} \geq \operatorname{E}\boldsymbol{t}$, pela definição de esperança. Como σ_x^2 e σ_y^2 são finitas, $\operatorname{E}(\boldsymbol{xy})$ é também finito, e assim concluímos que a variância de $\boldsymbol{z}$ é finita, e finalmente que $\boldsymbol{z} \in V_4$, que era o nosso objetivo. $\diamondsuit$

Exercício 2.2. Complete a prova de que V_4 é um espaço vetorial.

2.2 Subespaços

Um subespaço S de um espaço linear é um subconjunto de V que continua satisfazendo todas as propriedades da Def. 2.2. Ou seja, um subespaço de um espaço vetorial *também é um espaço vetorial*. Como várias

das propriedades da Def. 2.2 são naturalmente "herdadas" por qualquer subconjunto de vetores de V, para verificar se um subconjunto S realmente é um subespaço basta verificar se:

1. $\boldsymbol{x}$, $\boldsymbol{y} \in S$ implica $\boldsymbol{x} + \boldsymbol{y} \in S$,

2. $\boldsymbol{x} \in S$, $\alpha \in \mathbb{F}$ implica $(\alpha \boldsymbol{x}) \in S$.

Exercício 2.3. Prove que se $S \subset V$, então S necessariamente satisfaz todas as propriedades da Def. 2.2, exceto possivelmente as duas acima.

Exercício 2.4. Prove que a definição de subespaço implica que se S é um subespaço de um espaço vetorial, então necessariamente $\mathbf{0} \in S$.

O conceito de subespaço é muito utilizado em processamento de sinais, para problemas de identificação de sinais em ruído. Também é usado em controle, na separação de modos lentos e rápidos de um sistema linear. Veremos alguns exemplos de aplicações mais à frente. Vejamos agora alguns exemplos simples de subespaços de espaços vetoriais.

Exemplo 2.3. Considere o espaço vetorial $V_1 = \mathbb{R}^3$. O conjunto dos vetores

$$S_1 = \left\{ \boldsymbol{x} \in \mathbb{R}^3,\ \boldsymbol{x} = (x_1, x_2, x_1) \right\}$$

é um subespaço de V_1 correspondente ao plano que contém o eixo dos (0,x_2,0) e tem inclinação de 45º com relação ao plano (x_1,x_2,0).

Exemplo 2.4. V_7 (os polinômios com grau menor ou igual a 5 e coeficientes reais) é um subespaço de V_6 (polinômios de grau finito e coeficientes reais).

Exercício 2.5. Verifique que os conjuntos dos exemplos acima são realmente subespaços vetoriais.

Exercício 2.6. Prove que o plano em $\mathbb{R}^3$ definido pela equação

$$x_1 - x_3 = 1$$

não é um subespaço de $\mathbb{R}^3$.

2.3 Operações com subespaços vetoriais

Nesta seção vamos definir as operações de *soma*, *soma direta* e *intersecção* de subespaços vetoriais. Essas operações são usadas para descrever um espaço vetorial, separando-o em partes com propriedades diferentes. Veremos as definições, e em seguida alguns exemplos.

Definição 2.3. Seja V um espaço vetorial, e R,S subespaços de V. Então definem-se

1. A *soma* de R e S como[4]

$$R + S \triangleq \big\{\boldsymbol{r} + \boldsymbol{s} : \ \boldsymbol{r} \in R, \ \boldsymbol{s} \in S\big\};$$

2. A *intersecção* de R e S como

$$R \cap S \triangleq \big\{\boldsymbol{v} : \ \boldsymbol{v} \in R, \ \boldsymbol{v} \in S\big\};$$

O símbolo $\triangleq$ significa "igual por definição".

[4]Está sendo empregada a notação usual para definição de conjuntos:

$$A = \big\{x \in X : \ x \text{ tem a propriedade } Y\big\}$$

significa que o conjunto A é composto por todos os elementos de X que satisfazem a propriedade Y. Também é comum se usar a barra vertical "|" em lugar dos dois pontos ":".

3. T é a *soma direta* de R e S ($T = R \oplus S$) se

   (a) $R \cap S = \mathbf{0}$;

   (b) $R + S = T$.

A soma $R + S$ denota o subespaço vetorial formado por todos os elementos de V que podem ser escritos como $\boldsymbol{v} = \boldsymbol{r} + \boldsymbol{s}$, com $\boldsymbol{r} \in R$ e $\boldsymbol{s} \in S$. Por exemplo:

Exemplo 2.5. Considere o espaço vetorial $V_1 = \mathbb{R}^3$, e como subespaços as retas $R_1 = (x_1,0,0)$ (o eixo x_1) e $S_1 = (0, 0, x_3)$ (o eixo x_3). O subespaço $R_1 + S_1$ é formado por todos os vetores

$$\boldsymbol{v} = \alpha(x_1, 0, 0) + \beta(0, 0, x_3) = (\alpha x_1, 0, \beta x_3),$$

ou seja, o plano vertical P que contém os eixos x_1 e x_3. A intersecção $R_1 \cap S_1$ contém apenas o vetor $\mathbf{0}$, ou seja, $R_1 \cap S_1 = \mathbf{0}$. Como a intersecção contém apenas o vetor nulo, podemos dizer que P é a soma direta de R_1 e S_1, e escrever $P = R_1 \oplus S_1$.

Repare que $R_1 + S_1$ é diferente da união $R_1 \cup S_1$, que é apenas a união do eixo x_1 com o eixo x_3. O vetor $(1, 0, 2)$ pertence a $R_1 + S_1$, mas não pertence a $R_1 \cup S_1$. Veja a figura 2.1.

A soma e a interseção de dois subespaços é sempre um subespaço. Em particular, a interseção de dois subespaços sempre contém ao menos um vetor, o vetor $\mathbf{0}$. Vamos verificar esses fatos.

Teorema 2.1. *Seja V um espaço vetorial e R e S dois subespaços de V. Então $R + S$ e $R \cap S$ são também subespaços de V.*

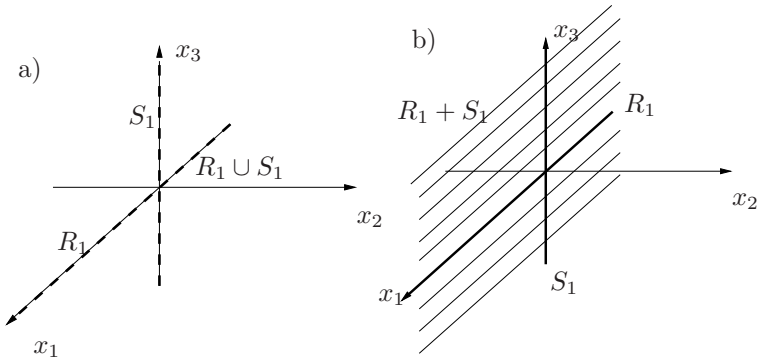

Figura 2.1: União (a) e soma (b) de subespaços vetoriais.

Demonstração. Para o caso de $R + S$: considere $\boldsymbol{v}_1$,$\boldsymbol{v}_2 \in R + S$. Por definição, existem $\boldsymbol{r}_1$,$\boldsymbol{r}_2 \in R$, $\boldsymbol{s}_1$,$\boldsymbol{s}_2 \in S$ tais que $\boldsymbol{v}_1 = \boldsymbol{r}_1 + \boldsymbol{s}_1$, $\boldsymbol{v}_2 = \boldsymbol{r}_2 + \boldsymbol{s}_2$. Portanto, $\boldsymbol{v}_1 + \boldsymbol{v}_2 = (\boldsymbol{r}_1 + \boldsymbol{s}_1) + (\boldsymbol{r}_2 + \boldsymbol{s}_2) = (\boldsymbol{r}_1 + \boldsymbol{r}_2) + (\boldsymbol{s}_1 + \boldsymbol{s}_2)$. Como $\boldsymbol{r}_0 \stackrel{\Delta}{=} \boldsymbol{r}_1 + \boldsymbol{r}_2 \in R$ (pois R é subespaço) e $\boldsymbol{s}_0 \stackrel{\Delta}{=} \boldsymbol{s}_1 + \boldsymbol{s}_2 \in S$ (pois S é subespaço), concluímos que $\boldsymbol{v}_1 + \boldsymbol{v}_2 = \boldsymbol{r}_0 + \boldsymbol{s}_0$, com $\boldsymbol{r}_0 \in R$, $\boldsymbol{s}_0 \in S$, e portanto (por definição) $\boldsymbol{v}_1 + \boldsymbol{v}_2 \in R + S$.

As provas para $\alpha \boldsymbol{v}_1 \in R + S$ e para $R \cap S$ são semelhantes. ■

Exercício 2.7. Complete a demonstração do Teorema 2.1.

Exercício 2.8. Prove que a união de dois subespaços vetoriais R e S de um espaço vetorial V *não é* necessariamente um subespaço vetorial. Há alguma situação em que a união $R \cup S$ também será um subespaço? Dê um exemplo, se houver.

Exemplo 2.6. Considere agora o espaço V_7, dos polinômios com coeficientes reais e grau menor ou igual a 5, e seja R_7 o subespaço dos polinômios de grau menor ou igual a 3, e S_7 o subespaço dos polinômios em que o coeficiente de x^5 é igual ao coeficiente de x^2, e os coeficientes de x^0 e x^1 são nulos, ou seja:

$$R_7 = \big\{p(x) \in V_7 : \ \text{grau}\big(p(x)\big) \leq 3\big\},$$
$$S_7 = \big\{p(x) \in V_7 : \ p(x) = \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4 + \beta_2 x^5\big\}.$$

Vamos determinar a soma, a intersecção e a união dos dois subespaços.

1. Soma: Neste caso, $R_7 + S_7$ corresponde a todos os polinômios $p(x)$ que podem ser escritos como a soma de um polinômio de R_7 com um de S_7, ou seja,

   $$R_7 + S_7 = \big\{p(x) \in V_7 : \ p(x) = r(x) + s(x), r(x) \in R_7, s(x) \in S_7\big\}.$$

   Um polinômio genérico de R_7 tem a forma $r(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3$, e um polinômio de S_7 tem a forma $s(x) = \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4 + \beta_2 x^5$, de maneira que os polinômios de $R_7 + S_7$ têm a forma

   $$r(x) + s(x) = \alpha_0 + \alpha_1 x + (\alpha_2 + \beta_2)x^2 + (\alpha_3 + \beta_3)x^3 + \beta_4 x^4 + \beta_2 x^5.$$

   É fácil de verificar que podemos escolher os α_i e β_j de maneira a obter qualquer polinômio de V_7. Portanto, $R_7 + S_7 = V_7$.

2. União: a união é mais fácil, consiste apenas de todos os polinômios de R_7 e de todos os de S_7. Repare que um polinômio como $1 + x^2 + x^5$ *não* pertence a $R_7 \cup S_7$.

3. Intersecção: os únicos polinômios que pertencem tanto a R_7 quanto a S_7 são da forma γx^3, e portanto

$$R_7 \cap S_7 = \big\{p(x) \in V_7 : \; p(x) = \gamma x^3; \; \gamma \in \mathbb{R}\big\}.$$

   Finalmente, note que, como $R_7 \cap S_7 \neq \big\{z(x) \equiv 0\big\}$, apesar de valer $V_7 = R_7 + S_7$, $V_7 \neq R_7 \oplus S_7$.

Você pode estar se perguntando a razão de existir um símbolo especial para a soma direta. É que se um subespaço vetorial puder ser descrito como a soma direta de dois subespaços R e S, cada vetor de $R \oplus S$ pode ser escrito *de maneira única* como uma soma de um vetor de R com um vetor de S (ou seja, se você tomar um vetor qualquer não nulo de R e outro de S, esses vetores serão necessariamente linearmente independentes — veja a próxima seção). Isso não acontece se a soma não for direta. Isso pode ser visto nos exemplos 2.5 e 2.6: no primeiro exemplo, qualquer vetor $(\alpha,0,\beta)$ do plano $R_1 \oplus S_1$ tem uma e apenas uma decomposição

$$(\alpha, 0, \beta) = (\alpha, 0, 0) + (0, 0, \beta).$$

Por outro lado, um polinômio de $V_7 = R_7 + S_7$ do exemplo 2.6 tem sempre mais de uma decomposição, por exemplo

$$1+x^2 = (1+x^2+x^3 \in R_7)+(-x^3 \in S_7) = (1+x^2-2x^3 \in R_7)+(2x^3 \in S_7).$$

O fato da decomposição ser única em uma soma direta facilita em muitos casos a análise teórica, e pode ser útil em aplicações também, como veremos em outros capítulos. Vamos demonstrar este fato.

Teorema 2.2. *Seja V um espaço vetorial e R e S dois subespaços de V. Caso $R \cap S = \{\mathbf{0}\}$, então cada vetor de $R + S (= R \oplus S)$ pode ser escrito* de maneira única *como uma soma de vetores de R e S. Caso $R \cap S \neq \{\mathbf{0}\}$, para cada $\boldsymbol{v} \in R + S$, há mais de um par $\boldsymbol{r} \in R$, $\boldsymbol{s} \in S$ tal que $\boldsymbol{v} = \boldsymbol{r} + \boldsymbol{s}$.*

Demonstração. Tome um vetor $\boldsymbol{v} \in R + S$ qualquer. Por definição, sabemos que existem $\boldsymbol{r} \in R$,$\boldsymbol{s} \in S$ tais que $\boldsymbol{v} = \boldsymbol{r} + \boldsymbol{s}$. Suponha que existam outros vetores $\boldsymbol{r}_1 \in R$,$\boldsymbol{s}_1 \in S$ tais que também valha $\boldsymbol{r}_1 + \boldsymbol{s}_1 = \boldsymbol{v}$. Nesse caso,

$$\boldsymbol{v} = \boldsymbol{r} + \boldsymbol{s} = \boldsymbol{r}_1 + \boldsymbol{s}_1 \Rightarrow \boldsymbol{r} - \boldsymbol{r}_1 = \boldsymbol{s}_1 - \boldsymbol{s}.$$

Como $\boldsymbol{r} - \boldsymbol{r}_1 \in R$ (por definição de subespaço) e $\boldsymbol{s}_1 - \boldsymbol{s} \in S$, necessariamente deve valer que $\boldsymbol{r} - \boldsymbol{r}_1 = \boldsymbol{s}_1 - \boldsymbol{s} \in R \cap S$. Portanto, se $R \cap S = \{\mathbf{0}\}$, concluímos que $\boldsymbol{r} = \boldsymbol{r}_1$ e $\boldsymbol{s} = \boldsymbol{s}_1$, ou seja, a decomposição de $\boldsymbol{v}$ é única, e a primeira parte do teorema está demonstrada.

Caso $R \cap S \neq \{\mathbf{0}\}$, tome um vetor qualquer $\boldsymbol{u} \in R \cap S$, $\boldsymbol{u} \neq \mathbf{0}$. Então se $\boldsymbol{r} - \boldsymbol{r}_1 = \boldsymbol{u} = \boldsymbol{s}_1 - \boldsymbol{s}$, teremos $\boldsymbol{r} \neq \boldsymbol{r}_1$, $\boldsymbol{s} \neq \boldsymbol{s}_1$, e $\boldsymbol{v} = \boldsymbol{r} + \boldsymbol{s} = \boldsymbol{r}_1 + \boldsymbol{s}_1$ (ou seja, a decomposição não é única). ■

2.4 Independência Linear

A operação básica da álgebra linear é a combinação linear de vetores, ou seja, relações do tipo

$$\boldsymbol{v} = \alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \cdots + \alpha_k \boldsymbol{v}_k.$$

Usualmente temos um conjunto de vetores, e precisamos saber:

- Que conjunto pode ser alcançado a partir de combinações lineares de um conjunto dado de vetores $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$?

- O mesmo conjunto pode ser alcançado retirando-se alguns dos vetores do conjunto inicial? É possível trocar-se alguns dos vetores do conjunto inicial e ainda alcançar o mesmo conjunto final?
- Qual é a forma mais compacta de descrever o conjunto de vetores alcançado?

Para responder a essas perguntas é que se usa o conceito de independência linear.

Definição 2.4. Um conjunto de vetores $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ é dito *linearmente independente* (ou simplesmente *l.i.*) se a relação

$$r_1\boldsymbol{v}_1 + r_2\boldsymbol{v}_2 + \cdots + r_k\boldsymbol{v}_k = \mathbf{0}$$

somente for satisfeita se $\boxed{r_1 = r_2 = \cdots = r_k = 0}$. Caso a relação acima possa ser satisfeita escolhendo-se ao menos um dos $r_i \neq 0$, diz-se que $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ é *linearmente dependente* (ou *l.d.*).

Exemplo 2.7. Um conjunto de um único vetor é l.i. se e somente se o vetor não for nulo.

Exemplo 2.8. No $\mathbb{R}^3$, dois vetores são l.i. se e somente se não estiverem em uma mesma reta passando pela origem, ou seja, se um não for múltiplo do outro. Assim,

$$\{(1,0,1);(2,0,2)\}$$

é l.d., enquanto

$$\{(1,0,1);(2,0,1)\}$$

é l.i..

Exemplo 2.9. Novamente no $\mathbb{R}^3$, três vetores são l.i. se e somente se não estiverem os três contidos em um mesmo plano passando pela origem. Quatro vetores no $\mathbb{R}^3$ são sempre l.d..

Observação 2.2. *Dizer que um conjunto de vetores é l.d. é equivalente a dizer que um dos vetores é uma combinação linear dos outros. Para verificar essa afirmação, assuma que um conjunto de vetores* $C = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ *seja l.d.. Pela Def. 2.4, deve existir pelo menos um* $r_i \neq 0$ *tal que*

$$r_1\boldsymbol{v}_1 + \cdots + r_i\boldsymbol{v}_i + \cdots + r_k\boldsymbol{v}_k = 0,$$

e portanto

$$\boldsymbol{v}_i = -\frac{r_1}{r_i}\boldsymbol{v}_1 - \cdots - \frac{r_{i-1}}{r_i}\boldsymbol{v}_{i-1} - \frac{r_{i+1}}{r_i}\boldsymbol{v}_{i+1} - \cdots - \frac{r_k}{r_i}\boldsymbol{v}_k.$$

Por outro lado, se um vetor qualquer de C *for combinação linear dos outros, uma relação como a anterior é válida, e claramente os vetores serão l.d..*

Como C *é l.d., e portanto um vetor* $\boldsymbol{v}_i$ *é combinação linear dos outros, o conjunto de vetores alcançado por combinações lineares de* C *não se altera se retirarmos* $\boldsymbol{v}_i$*.*

Exemplo 2.10. Seja V um espaço vetorial e R,S dois subespaços de V cuja intersecção contém apenas o vetor nulo. Então dois vetores não nulos de R e S são necessariamente l.i..

Prova: Considere $\boldsymbol{r} \in R$, $\boldsymbol{s} \in S$ dois vetores não nulos. Para satisfazer

$$\alpha\boldsymbol{r} + \beta\boldsymbol{s} = \boldsymbol{0},$$

necessariamente deve valer $\alpha\boldsymbol{r}(\in R) = -\beta\boldsymbol{s}(\in S)$, e portanto $\alpha\boldsymbol{r} = -\beta\boldsymbol{s} \in R \cap S$. Como $R \cap S = \{\boldsymbol{0}\}$ por hipótese, segue que $\alpha\boldsymbol{r} = -\beta\boldsymbol{s} = \boldsymbol{0}$.

Como assumimos que $\boldsymbol{r} \neq \mathbf{0}$, $\boldsymbol{s} \neq \mathbf{0}$, a única solução é $\alpha = \beta = 0$, e portanto os vetores são l.i..

Exercício 2.9. O conjunto de funções $\{\operatorname{sen}(x), \cos(x), \cos(x+\phi)\}$ é l.i. ou l.d. no espaço das funções contínuas definidas em $[0,1]$?

Exercício 2.10. Prove que, se de um conjunto l.i. $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ for retirado um vetor $\boldsymbol{v}_i$ qualquer, o conjunto permanece l.i..

Exercício 2.11. Ao conjunto do exercício anterior é acrescentado um novo vetor $\boldsymbol{v}_{k+1}$, tal que $\{\boldsymbol{v}_i, \boldsymbol{v}_{k+1}\}$ é l.i. para todo $1 \leq i \leq k$. Pode-se afirmar que o conjunto resultante é l.i.? Caso sim, prove; caso contrário, forneça um contra-exemplo (isto é, um exemplo de um caso em que o conjunto resultante é l.d., mesmo com todas as condições desse exercício sendo atendidas).

Exercício 2.12. Prove que, se no exercício 2.10 trocarmos $\boldsymbol{v}_1$ por $\boldsymbol{v}_1 + \alpha_2\boldsymbol{v}_2 + \cdots + \alpha_k\boldsymbol{v}_k$ ($\alpha_2, \ldots \alpha_k$ são escalares quaisquer), o conjunto resultante permanece l.i..

2.5 Bases e Dimensão

Alguns espaços vetoriais podem ter conjuntos l.i. com qualquer número finito de vetores (por exemplo, o espaço V_5 das seqüências infinitas da página 11). Outros espaços, como o $\mathbb{R}^3$, têm um número máximo de vetores l.i.. A esse número máximo dá-se o nome de *dimensão* do espaço vetorial.

Definição 2.5. O número máximo de elementos em um conjunto l.i. de vetores de um espaço vetorial V é chamado *dimensão* de V ($\dim V$). Se o número máximo não existir, diz-se que o espaço tem dimensão infinita.

Exemplo 2.11. A dimensão do $\mathbb{R}^2$ é 2, pois há conjuntos de 2 vetores l.i. em $\mathbb{R}^2$, mas quaisquer três vetores do $\mathbb{R}^2$ são sempre l.d..

Como veremos nos capítulos seguintes, muitas técnicas de análise matricial correspondem a escolher uma maneira adequada para descrever um espaço vetorial (um conjunto de coordenadas conveniente), fazendo com que propriedades úteis fiquem mais evidentes. Vamos agora definir uma *base* de um espaço vetorial, que é simplesmente um sistema de coordenadas que representa de maneira biunívoca os elementos do espaço vetorial. Vamos também precisar da seguinte definição:

Definição 2.6. O *espaço alcançado* (*range* ou *span*) de um conjunto de vetores $C = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ de um espaço vetorial V são todos os vetores de V que são combinações lineares de C:

$$\mathcal{R}(C) = \{\boldsymbol{v} \in V : \ \boldsymbol{v} = \alpha_1 \boldsymbol{v}_1 + \cdots + \alpha_n \boldsymbol{v}_n, \ \alpha_1, \ldots, \alpha_n \in \mathbb{F}\}.$$

Usa-se também a notação $\mathrm{Sp}(C)$ para denotar o espaço alcançado.

Exercício 2.13. Prove que $\mathcal{R}\big(\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}\big)$ é um subespaço vetorial de V.

Agora podemos ver a definição de base:

Definição 2.7. Um conjunto de vetores (não necessariamente finito) $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots\}$ de um espaço vetorial V é uma *base* $\mathcal{B}$ de V se

1. $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots$ forem l.i., 2. $\mathcal{R}\big(\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots\}\big) = V$.

Exemplo 2.12. A *base canônica* do $\mathbb{R}^n$ é o conjunto de vetores

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

Note que o conjunto acima é l.i., e que todo vetor do $\mathbb{R}^n$ pode ser escrito como uma combinação linear da base canônica.

Exercício 2.14. Mostre que os polinômios $1+x$ e $1-x$ formam uma base do espaço dos polinômios de coeficientes reais e grau menor ou igual a 1. Mostre também que os polinômios $1+x$, $1-x$ e x *não formam* uma base desse espaço.

Um resultado importante é que duas bases diferentes de um mesmo espaço vetorial têm sempre o mesmo número de vetores, exatamente a dimensão do espaço vetorial. Vamos demonstrar esse resultado no teorema a seguir.

Teorema 2.3. *Sejam V um espaço vetorial com dimensão finita n. Então o número de elementos de qualquer base $\mathcal{B}$ de V é exatamente n (e sempre existe uma base para V).*

Demonstração. O número de elementos de uma base de V não pode ser maior do que n, pois um espaço de dimensão n não pode ter mais de n vetores l.i.. Por outro lado, se uma base tivesse $m < n$ vetores $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$, como o espaço tem dimensão n, deveríamos poder escrever n vetores l.i. $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ como combinações lineares dos elementos da base:

$$\boldsymbol{u}_i = \alpha_{i,1}\boldsymbol{v}_1 + \alpha_{i,2}\boldsymbol{v}_2 + \cdots + \alpha_{i,m}\boldsymbol{v}_m, \ i = 1 \ldots n.$$

Para $i = 1$, deve haver ao menos um $\alpha_{i,1} \neq 0$ (caso contrário, o conjunto $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ não seria l.i.). Podemos considerar, sem perda de generalidade, que $\alpha_{1,1} \neq 0$ (caso contrário, bastaria trocar a ordem dos $\boldsymbol{v}_j$). Agora, vamos calcular $\boldsymbol{u}_i - (\alpha_{i,1}/\alpha_{1,1})\boldsymbol{u}_1$ para $i = 2 \ldots n$, resultando

$$\boldsymbol{u}_i - \frac{\alpha_{i,1}}{\alpha_{1,1}}\boldsymbol{u}_1 = \alpha_{i,2}^{(1)}\boldsymbol{v}_2 + \ldots \alpha_{i,m}^{(1)}\boldsymbol{v}_m, i = 2 \ldots n,$$

em que $\alpha_{i,j}^{(1)} = \alpha_{i,j} - \frac{\alpha_{i,1}}{\alpha_{1,1}}\alpha_{1,j}$ para i, $j = 2 \ldots n$.

Tome agora a equação para $i = 2$. Sabemos que ao menos um $\alpha_{2,j}^{(1)} \neq 0$, caso contrário, $\boldsymbol{u}_2$ seria proporcional a $\boldsymbol{u}_1$, contrariando a hipótese de $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ ser l.i.. Novamente sem perda de generalidade, vamos supor que $\alpha_{2,2}^{(1)} \neq 0$. Podemos então repetir o procedimento anterior, calculando

$$\boldsymbol{u}_i - \frac{\alpha_{i,2}^{(1)}}{\alpha_{2,2}^{(1)}}\left(\boldsymbol{u}_2 - \frac{\alpha_{2,1}}{\alpha_{1,1}}\boldsymbol{u}_1\right) - \frac{\alpha_{i,1}}{\alpha_{1,1}}\boldsymbol{u}_1 = \alpha_{i,3}^{(2)}\boldsymbol{v}_3 + \cdots + \alpha_{i,m}^{(2)}\boldsymbol{v}_m,\ i = 3 \ldots n,$$

para zerar os coeficientes de $\boldsymbol{u}_2$, e assim por diante. Novamente, caso $\alpha_{3,j}^{(2)} = 0$ para $j = 3 \ldots m$, o conjunto $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ seria l.d. (pois o coeficiente multiplicando $\boldsymbol{u}_3$ é $1 \neq 0$), o que não é possível por hipótese.

Podemos portanto continuar repetindo o procedimento, até chegarmos à equação

$$\boldsymbol{u}_{m+1} - \sum_{i=1}^{m} \beta_i \boldsymbol{u}_i = \boldsymbol{0},$$

em que os β_i são constantes dependentes dos $\alpha_{i,j}$. Como o coeficiente de $\boldsymbol{u}_{m+1}$ na combinação linear acima é não-nulo, concluímos que o conjunto $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{m+1}\}$ é l.d., contrariando nossa hipótese.

Com esse argumento, provamos que toda base deve ter exatamente n elementos. Para provar que sempre existe uma base, lembre que, por definição, se o espaço V tem dimensão n, necessariamente existe um conjunto $\mathcal{B} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ de n vetores l.i.. Como $\mathcal{B} \subset V$, $\mathcal{R}(\mathcal{B}) \subset V$, pois V é espaço vetorial. Além disso, se houvesse $\boldsymbol{v} \in V$ tal que $\boldsymbol{v} \notin \mathcal{R}(\mathcal{B})$, então

$$\beta\boldsymbol{v} + \alpha_1\boldsymbol{u}_1 + \cdots + \alpha_2\boldsymbol{u}_n = \boldsymbol{0} \tag{2.1}$$

implica $\beta = 0$ (pois se $\beta \neq 0$, $\boldsymbol{v}$ pertenceria a $\mathcal{R}(\mathcal{B})$). Como $\mathcal{B}$ é l.i., a equação resultante só tem solução para $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$.

Portanto, se $\boldsymbol{v} \notin \mathcal{R}(\mathcal{B})$ teríamos achado um conjunto de $n+1$ vetores l.i. em V, contrariando a hipótese da dimensão de V ser n.

■

A partir do teorema acima é fácil verificar o seguinte colorário:

Corolário 2.1. *A dimensão de $\mathcal{R}\big(\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}\big)$, em que os $\boldsymbol{v}_i$ pertencem a um espaço vetorial V de dimensão n, é menor ou igual ao mínimo entre k e n.*

Exercício 2.15. Demonstre o colorário acima.

2.5.1 Sistemas de coordenadas e mudança de base

Uma base de um espaço vetorial define um sistema de coordenadas para representar os elementos do espaço vetorial. Vamos ver agora como, através de uma base, podemos usar o espaço $\mathbb{R}^n$ ou $\mathbb{C}^n$ para representar os vetores de um espaço de dimensão n finita, e como achar as coordenadas de um dado vetor em uma nova base, conhecendo-se as coordenadas em uma base original.

O próximo teorema diz que a representação de um vetor qualquer em termos de uma base é única. Isso permite estabelecer uma relação biunívoca entre um espaço V de dimensão n e o $\mathbb{R}^n$ (se os escalares de V forem os reais) ou o $\mathbb{C}^n$ (se os escalares forem os complexos).

Teorema 2.4. *Dado um vetor $\boldsymbol{x}$ de um espaço vetorial V, e uma base $\mathcal{B}$ de V, composta pelos vetores $\boldsymbol{v}_i$, a representação de $\boldsymbol{x}$ como uma combinação linear dos $\boldsymbol{v}_i$ é única.*

Demonstração. Sejam

$$\boldsymbol{x} = \alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \dots$$

e

$$\boldsymbol{x} = \beta_1 \boldsymbol{v}_1 + \beta_2 \boldsymbol{v}_2 + \dots$$

duas representações de $\boldsymbol{x}$. Então, subtraindo as duas igualdades temos

$$\mathbf{0} = (\alpha_1 - \beta_1)\boldsymbol{v}_1 + (\alpha_2 - \beta_2)\boldsymbol{v}_2 + \dots$$

Como os $\boldsymbol{v}_i$ são l.i., necessariamente $\alpha_i = \beta_i$. ■

Dada uma base $\mathcal{B} = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ de um espaço V de dimensão n finita, um vetor $\boldsymbol{x} \in V$ é escrito em função dos $\boldsymbol{v}_i$ como

$$\boldsymbol{x} = \sum_{i=1}^{n} \alpha_i \boldsymbol{v}_i. \tag{2.2}$$

Como esta representação é biunívoca (dado um $\boldsymbol{x} \in V$ o conjunto ordenado dos α_i correspondentes existe e é único, e dados os α_i, o vetor correspondente $\boldsymbol{x}$ também é único), podemos representar cada vetor $\boldsymbol{x} \in V$ por um vetor do $\mathbb{R}^n$ (ou do $\mathbb{C}^n$, ou em geral do $\mathbb{F}^n$)

$$\boldsymbol{x} \text{ pode ser representado como } \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}_{\mathcal{B}} \triangleq [\boldsymbol{x}]_{\mathcal{B}}. \tag{2.3}$$

Uma relação biunívoca como esta é também chamada de *isomorfismo* entre o espaço vetorial V de dimensão n, e $\mathbb{F}^n$.

Devido a esta equivalência entre $\mathbb{F}^n$ e um espaço vetorial V de dimensão n definido sobre $\mathbb{F}$, vamos nos concentrar de agora em diante nos espaços $\mathbb{F}^n$.

Exercício 2.16. Prove que se R e S são subespaços de V, então $\dim(R+S) \leq \dim V$.

Exercício 2.17. Prove que, se $R \oplus S = V$, então $\dim R + \dim S = \dim V$.

Exercício 2.18. Prove agora o resultado geral:

$$\dim(R+S) = \dim R + \dim S - \dim(R \cap S).$$

Vamos ver agora como trocar a representação de um vetor entre duas bases diferentes, ou seja, passar de um sistema de coordenadas para outro. Suponha que um espaço V tenha uma base $\mathcal{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$. Um vetor $\boldsymbol{x} \in V$ pode portanto ser representado por um vetor do $\mathbb{F}^n$, como em (2.3). Suponha também que $\mathcal{C} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ forme uma segunda base de V. Qual é a representação de $\boldsymbol{x}$ em termos de $\mathcal{C}$?

Como $\mathcal{B}$ é uma base de V, os vetores $\boldsymbol{u}_j$ da base $\mathcal{C}$ podem ser representados em termos dos elementos de $\mathcal{B}$, ou seja,

$$\boldsymbol{u}_j = a_{1,j}\boldsymbol{v}_1 + a_{2,j}\boldsymbol{v}_2 + \cdots + a_{n,j}\boldsymbol{v}_n.$$

Seja $\boldsymbol{x} = \beta_1\boldsymbol{u}_1 + \cdots + \beta_n\boldsymbol{u}_n$ a representação de $\boldsymbol{x}$ em termos de $\mathcal{C}$. Portanto,

$$\boldsymbol{x} = \sum_{j=1}^{n} \beta_j \boldsymbol{u}_j = \sum_{j=1}^{n} \beta_j \left(\sum_{i=1}^{n} a_{i,j}\boldsymbol{v}_i \right) = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} a_{i,j}\beta_j \right) \boldsymbol{v}_i.$$

Comparando este resultado com (2.2), concluímos que

$$\alpha_i = \sum_{j=1}^{n} a_{i,j}\beta_j.$$

Em notação matricial,

$$\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} \Rightarrow [\boldsymbol{x}]_{\mathcal{B}} = \boldsymbol{A}[\boldsymbol{x}]_{\mathcal{C}},$$

em que $\boldsymbol{A}$ é a matriz com elementos $a_{i,j}$. Para calcular $[\boldsymbol{x}]_{\mathcal{C}}$ em função de $[\boldsymbol{x}]_{\mathcal{B}}$, precisamos portanto resolver um sistema de equações lineares. Veremos nos próximos capítulos que esse sistema realmente tem sempre uma solução única, sendo possível escrever

$$[\boldsymbol{x}]_{\mathcal{C}} = \boldsymbol{A}^{-1}[\boldsymbol{x}]_{\mathcal{B}}. \tag{2.4}$$

A matriz $\boldsymbol{A}$ é denominada *matriz de mudança de base* — repare que, pela definição, a matriz de mudança de base leva a representação na base *nova* para a antiga! Cuidado para não se confundir.

2.6 Produto Interno e Ortogonalidade

A noção de ortogonalidade do espaço euclideano $\mathbb{R}^3$ pode ser facilmente estendida para espaços vetoriais quaisquer através do conceito de produto interno (produto escalar). A gama de aplicações do conceito de

ortogonalidade é imensa, principalmente a solução de problemas de mínimos quadrados, e critérios para uma boa escolha de bases para representação de vetores.

Nesta seção veremos a definição geral de produto interno, exemplos para alguns espaços vetoriais, a definição de ortogonalidade, e o conceito de subespaços ortogonais.

2.6.1 Produtos internos

Um produto interno é uma função que leva um par de vetores $\boldsymbol{u}, \boldsymbol{v}$ de um espaço vetorial U para o conjunto dos escalares $\mathbb{F}$, com algumas propriedades especiais (note que para um mesmo espaço vetorial pode haver mais de uma definição de produto interno). Essa função permite que se meça se um vetor está "próximo" de ser paralelo a outro (proporcional ao outro), ou seja, um produto interno permite a definição do conceito de ângulo entre vetores. Para simplificar o texto, a definição abaixo é dada para espaços sobre os números complexos. Para espaços reais, basta eliminar o conjugado.

Definição 2.8. Dado um espaço vetorial U sobre os complexos, uma relação[5]

$$< \cdot, \cdot >: U \times U \to \mathbb{C}$$

é um *produto interno* se, para quaisquer vetores $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{z} \in U$ e quaisquer escalares $\alpha, \beta \in \mathbb{C}$, valer[6]

1. $< \boldsymbol{u}, \boldsymbol{v} > = < \boldsymbol{v}, \boldsymbol{u} >^*$,

[5]Recorde que, dados conjuntos A e B, o conjunto $A \times B$ (A *cartesiano* B) é formado por todos os pares possíveis de elementos (a,b), com $a \in A$ e $b \in B$.

[6]Indicamos por α^* o conjugado de α.

2. $< \boldsymbol{u}, \alpha \boldsymbol{w} + \beta \boldsymbol{z} >= \alpha < \boldsymbol{u}, \boldsymbol{w} > + \beta < \boldsymbol{u}, \boldsymbol{z} >$,

3. $< \boldsymbol{x}, \boldsymbol{x} > \geq 0$ para todo $\boldsymbol{x} \in U$, e $< \boldsymbol{x}, \boldsymbol{x} >= 0$ se e somente se $\boldsymbol{x} = \mathbf{0}$.

Exemplo 2.13. O produto escalar usual no $\mathbb{R}^3$ é um produto interno: se

$$\boldsymbol{x} = (\alpha, \beta, \gamma), \qquad \boldsymbol{y} = (a, b, c),$$

então

$$< \boldsymbol{x}, \boldsymbol{y} >= \alpha a + \beta b + \gamma c,$$

e não é difícil verificar que todas as condições da Def. 2.8 são satisfeitas.

Exemplo 2.14. Um outro produto escalar no $\mathbb{R}^3$ é dado por (usando os mesmos $\boldsymbol{x}$ e $\boldsymbol{y}$ do exemplo anterior)

$$< \boldsymbol{x}, \boldsymbol{y} >_1 \stackrel{\Delta}{=} \alpha a + 1/2 \beta b + 1/4 \gamma c.$$

Exemplo 2.15. No $\mathbb{C}^n$, é necessário um cuidado adicional, por causa da condição 2 da Def. 2.8: se $\boldsymbol{x} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ e $\boldsymbol{y} = (\beta_1, \beta_2, \ldots, \beta_n) \in \mathbb{C}^n$, então

$$< \boldsymbol{x}, \boldsymbol{y} >= \sum_{k=1}^{n} \alpha_k^* \beta_k$$

é um produto escalar.

Exemplo 2.16. No conjunto V_3 das funções reais contínuas definidas no intervalo $(0, 1)$, um produto escalar é

$$< f, g >= \int_0^1 f(t) g(t) \, \mathrm{d}\, t.$$

Exemplo 2.17. Considere o espaço vetorial formado pelas variáveis aleatórias reais com média nula e variância finita. Então a correlação cruzada de duas variáveis X e Y é um produto interno:

$$< X,Y >= \mathrm{E}\{X \cdot Y\}.$$

Exercício 2.19. Prove que as funções definidas nos cinco exemplos acima são realmente produtos escalares.

Vamos tratar de algumas propriedades de produtos internos quando formos discutir normas de vetores. Por enquanto, vamos nos concentrar no conceito de ortogonalidade.

2.6.2 Ortogonalidade

A noção de ortogonalidade é usada por exemplo em problemas de otimização, para achar a distância mínima entre dois conjuntos, ou para escolher bases em que produtos internos podem ser calculados com facilidade. A definição é conhecida:

Definição 2.9. Dois vetores $\boldsymbol{x}$ e $\boldsymbol{y}$ de um espaço vetorial U são *ortogonais* com respeito ao produto interno $< \cdot,\cdot >$ se $< \boldsymbol{x},\boldsymbol{y} >= 0$.

Repare que a definição de ortogonalidade depende do produto escalar escolhido: Por exemplo, usando o produto escalar usual do $\mathbb{R}^3$ (exemplo 2.13, os vetores $\boldsymbol{x} = \begin{bmatrix}1 & 1 & 0\end{bmatrix}^T$ e $\boldsymbol{y} = \begin{bmatrix}1 & -1 & 0\end{bmatrix}^T$ são ortogonais, mas $\boldsymbol{x}$ e $\boldsymbol{z} = \begin{bmatrix}1 & -2 & 0\end{bmatrix}^T$ não são. No entanto, usando o produto interno do exemplo 2.14, $\boldsymbol{x}$ é ortogonal a $\boldsymbol{z}$, mas não a $\boldsymbol{y}$.

Dado um espaço vetorial U de dimensão n. Dada uma base $\mathcal{B}$, é sempre possível construir-se uma base $\mathcal{B}_1$ com vetores ortogonais entre si.

Vamos demonstrar este fato no próximo teorema usando o algoritmo de Gram-Schmidt. Esse algoritmo é mais usado teoricamente, para mostrar que existe uma base ortogonal. Na prática, devido a problemas com erros em aritmética de precisão finita, são usados outros métodos (ver algoritmo QR nos próximos capítulos).

Vamos aproveitar a demonstração do teorema para mostrar a técnica de *prova por indução finita*. Quando se deseja provar que uma propriedade vale para um conjunto com elementos $x_1, x_2, \ldots,$ às vezes o método mais fácil é provar que a propriedade vale para o primeiro elemento x_1. Depois, demonstra-se que, se a propriedade valer para todo elemento x_k para $k < N$, então vale também para x_N. Conclui-se que a propriedade vale para qualquer valor finito de n.

Definição 2.10. Uma base $\mathcal{B}$ é chamada *ortonormal* se seus vetores $\boldsymbol{u}_i$ obedecem

$$< \boldsymbol{u}_i, \boldsymbol{u}_j >= \delta_{ij},$$

em que $\delta_{ij} = 1$ se $i = j$ e 0 caso contrário.

Teorema 2.5 (Algoritmo de Gram-Schimdt)**.** *Seja U um espaço vetorial com dimensão n e uma base $\mathcal{B} = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$. A partir de $\mathcal{B}$ é possível construir-se uma base ortonormal $\mathcal{B}_1 = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$.*

Demonstração. A demonstração é construtiva (quer dizer, vamos apresentar um algoritmo para obter a base desejada). Começamos tomando um vetor qualquer de $\mathcal{B}$,

1. Defina $\boldsymbol{v}_1 \triangleq \boldsymbol{u}_1/\sqrt{< \boldsymbol{u}_1, \boldsymbol{u}_1 >}$.

2. Para $k = 2 \ldots n$, repita

   (a) $\boldsymbol{z}_k = \boldsymbol{u}_k - \sum_{m=1}^{k-1} < \boldsymbol{v}_m, \boldsymbol{u}_k > \boldsymbol{v}_m$

   (b) $\boldsymbol{v}_k = \boldsymbol{z}_k / \sqrt{< \boldsymbol{z}_k, \boldsymbol{z}_k >}$

Repare que (b) é bem definido, pois $\boldsymbol{z}_k \neq \boldsymbol{0}$ (de outra forma, $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ seria l.d.), e que $< \boldsymbol{v}_k, \boldsymbol{v}_k >= 1$ para todo k. Então vale

$$< \boldsymbol{v}_1, \boldsymbol{v}_2 >= \frac{< \boldsymbol{v}_1, \boldsymbol{u}_2 > - < \boldsymbol{v}_1, \boldsymbol{u}_2 >< \boldsymbol{v}_1, \boldsymbol{v}_1 >}{< \boldsymbol{z}_2, \boldsymbol{z}_2 >} = 0.$$

Vamos mostrar a ortogonalidade entre os demais $\boldsymbol{v}_k$ por indução. Assuma que $< \boldsymbol{v}_\ell, \boldsymbol{v}_m >= \delta_{\ell,m}$ para todo ℓ,$m \leq k$. Então vamos mostrar que $< \boldsymbol{v}_m, \boldsymbol{v}_{k+1} >= 0$ para $m < k+1$.

$$\begin{aligned} < \boldsymbol{v}_m, \boldsymbol{v}_{k+1} > &= \frac{< \boldsymbol{v}_m, \boldsymbol{u}_{k+1} > - \sum_{\ell=1}^{k} < \boldsymbol{v}_\ell, \boldsymbol{u}_{k+1} >< \boldsymbol{v}_m, \boldsymbol{v}_\ell >}{< \boldsymbol{z}_{k+1}, \boldsymbol{z}_{k+1} >} = \\ &= \frac{< \boldsymbol{v}_m, \boldsymbol{u}_{k+1} > - < \boldsymbol{v}_m, \boldsymbol{u}_{k+1} >< \boldsymbol{v}_m, \boldsymbol{v}_m >}{< \boldsymbol{z}_{k+1}, \boldsymbol{z}_{k+1} >} = 0, \end{aligned}$$

■

Repare que o nome da técnica é *indução finita*, e o adjetivo não está aí por acaso. O método demonstra que a propriedade vale para todo x_n, com n finito, mas não necessariamente vale para $\lim_{n\to\infty} x_n$. Para ver um contra-exemplo, considere a sequência $x_n = 2^n$. Vamos mostrar que x_n é finito para todo n. Primeiro, lembre que se x é finito, então $2x$ é também finito. Claramente $x_1 = 2$ é finito. Assuma agora que x_n seja finito para todo $n < N$, então $x_N = 2x_{N-1}$ é finito, pois x_{N-1} é finito por hipótese. No entanto, $\lim_{n\to\infty} x_n = \infty$.

Podemos estender o conceito de ortogonalidade para subespaços, conforme a definição abaixo.

Definição 2.11. Os subespaços S e T do espaço vetorial U são ortogonais entre si com relação ao produto interno $< \cdot,\cdot >$ (escreve-se $S \perp T$) se (e somente se) *todo* vetor de S for ortogonal a *todo* vetor de T, ou seja,

$$S \perp T \iff < \boldsymbol{s}, \boldsymbol{t} >= 0 \quad \forall \boldsymbol{s} \in S, \boldsymbol{t} \in T.$$

É fácil pensar em exemplos no $\mathbb{R}^2$ e no $\mathbb{R}^3$. O subespaço $S = \{\boldsymbol{s} \in \mathbb{R}^3 : \boldsymbol{s} = \alpha \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T, \alpha \in \mathbb{R}\}$ é ortogonal ao subespaço $T = \{\boldsymbol{t} \in \mathbb{R}^3 : \boldsymbol{t} = \beta \begin{bmatrix} 1 & -2 & 1 \end{bmatrix}^T\}$, $\beta \in \mathbb{R}$.

Dado um subespaço vetorial S de um espaço U e um produto interno, podemos achar o conjunto de *todos* os vetores ortogonais a S. Esse conjunto pode ser usado para formar bases de U com propriedades interessantes.

Definição 2.12. Seja S um subespaço de U, e $< \cdot,\cdot >$ um produto interno. O *complemento ortogonal* de S com relação ao produto interno $< \cdot,\cdot >$ é o conjunto de todos os vetores de U ortogonais a S, ou seja,

$$S^{\perp} = \{\boldsymbol{u} \in U :< \boldsymbol{u},\boldsymbol{s} >= 0, \forall \boldsymbol{s} \in S\}.$$

O complemento ortogonal também é um subespaço de U, como veremos logo a seguir no Teorema 2.6.

Exemplo 2.18. Considere a reta no $\mathbb{R}^3$

$$S = \{\boldsymbol{s} \in \mathbb{R}^3 : \boldsymbol{s} = \alpha \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T, \alpha \in \mathbb{R}\}.$$

O espaço ortogonal a S é o plano

$$S^{\perp} = \{\boldsymbol{u} \in \mathbb{R}^3 : [\boldsymbol{u}]_1 + [\boldsymbol{u}]_2 + [\boldsymbol{u}]_3 = 0\}.$$

Note que a reta T definida acima está contida em $S^{\perp}$.

Vejamos algumas propriedades de complementos ortogonais. Os primeiros resultados valem para espaços de quaisquer dimensões:

Teorema 2.6. *Seja U um espaço vetorial, e R, S subespaços de U. Então vale*

1. *$S^{\perp}$ é um subespaço de U, e $S \cap S^{\perp} = \{\mathbf{0}\}$.*
2. *$(R+S)^{\perp} = R^{\perp} \cap T^{\perp}$.*
3. *$R \subset S$ se e somente se $R^{\perp} \subset S^{\perp}$*

Demonstração. 1. Se $\boldsymbol{u}_1$, $\boldsymbol{u}_2 \in S^{\perp}$, então para um vetor qualquer $\boldsymbol{s} \in S$, vale

$$< \boldsymbol{s}, \alpha_1 \boldsymbol{u}_1 + \alpha_2 \boldsymbol{u}_2 >= \alpha_1 < \boldsymbol{s}, \boldsymbol{u}_1 > + \alpha_2 < \boldsymbol{s}, \boldsymbol{u}_2 >= 0,$$

pois $< \boldsymbol{s}, \boldsymbol{u}_1 >=< \boldsymbol{s}, \boldsymbol{u}_2 >= 0$ para todo $\boldsymbol{s} \in S$, por definição. Como isso vale para qualquer vetor de S, $\alpha_1 \boldsymbol{u}_1 + \alpha_2 \boldsymbol{u}_2 \in S^{\perp}$, e $S^{\perp}$ é um subespaço.

Para verificar a segunda afirmação, tome $\boldsymbol{s} \in S \cap S^{\perp}$. Por definição, $< \boldsymbol{s}, \boldsymbol{s} >= 0$, e portanto $\boldsymbol{s} = \mathbf{0}$.

2. Para provar que dois conjuntos $\mathcal{A}$ e $\mathcal{B}$ são iguais, um truque útil é mostrar que $\mathcal{A} \subset \mathcal{B}$ e que $\mathcal{B} \subset \mathcal{A}$. Vamos então primeiramente mostrar que $(R+S)^{\perp} \subset R^{\perp} \cap S^{\perp}$.

   Seja $\boldsymbol{u} \in (R+S)^{\perp}$. Então,

$$< \boldsymbol{u}, \boldsymbol{v} >= 0, \text{ para todo } \boldsymbol{v} = \boldsymbol{r} + \boldsymbol{s}, \text{ com } \boldsymbol{r} \in R, \boldsymbol{s} \in S. \quad (2.5)$$

Tome em particular $\boldsymbol{s} = \mathbf{0} \in S$. Nesse caso, (2.5) implica que $\boldsymbol{u}$ é ortogonal a todo $\boldsymbol{r} \in R$, ou seja, $\boldsymbol{u} \in R^{\perp}$. Analogamente, tomando $\boldsymbol{r} = \mathbf{0} \in R$ em (2.5), vemos que $\boldsymbol{u} \in S^{\perp}$. Como $\boldsymbol{u}$ pertence a $R^{\perp}$ e a $S^{\perp}$ simultaneamente, concluímos que $\boldsymbol{u} \in R^{\perp} \cap S^{\perp}$. Como o argumento vale para todo $\boldsymbol{u} \in (R+S)^{\perp}$, concluímos que $(R+S)^{\perp} \subset R^{\perp} \cap S^{\perp}$.

Para a outra direção, seja $\boldsymbol{v} \in R^{\perp} \cap S^{\perp}$. Então $< \boldsymbol{v}, \boldsymbol{r} >= 0$ para todo $\boldsymbol{r} \in R$ e $< \boldsymbol{v}, \boldsymbol{s} >= 0$ para todo $\boldsymbol{s} \in S$. Seja agora $\boldsymbol{u} \in R+S$. Por definição, existem $\boldsymbol{r} \in R$, $\boldsymbol{s} \in S$ tais que $\boldsymbol{u} = \boldsymbol{r} + \boldsymbol{s}$. Pela propriedade distributiva do produto interno, $< \boldsymbol{v}, \boldsymbol{u} >=< \boldsymbol{v}, \boldsymbol{r} > + < \boldsymbol{v}, \boldsymbol{s} >= 0$, e concluímos que $\boldsymbol{v} \in (R+S)^{\perp}$. O argumento vale para todo $\boldsymbol{v} \in R^{\perp} \cap S^{\perp}$, e portanto $R^{\perp} \cap S^{\perp} \subset (R+S)^{\perp}$, concluindo a demonstração deste item.

3. A demonstração desta parte fica como exercício.

■

Para espaços de dimensão finita, vale $(S^{\perp})^{\perp} = S$, ou seja, o subespaço ortogonal ao subespaço ortogonal a S é o próprio S. Isso no entanto não vale para espaços de dimensão infinita, vejamos um exemplo.

Exemplo 2.19. Considere U o espaço das sequências de energia finita, isto é, $\boldsymbol{x} = \{x_0, x_1, x_2, \ldots\}$ tais que $x_n \in \mathbb{R}$ e $\sum_{n=0}^{\infty} x_n^2 < \infty$. Podemos verificar que U é um espaço vetorial com produto interno $< \boldsymbol{x}, \boldsymbol{y} >= \sum_{n=0}^{\infty} x_n y_n$ usando a desigualdade de Schwarz (ver capítulo sobre normas). Considere agora o subespaço S das sequências com um número finito de elementos, ou seja,

$$S = \{\boldsymbol{s} \in U : \exists N < \infty \text{ tal que } s_n = 0 \ \forall n \geq N\}.$$

É fácil verificar que S é realmente um subespaço: a soma de duas sequências finitas também é uma sequência finita, e o produto de uma sequência finita por um escalar é uma sequência finita (também podemos verificar que toda sequência finita tem energia finita, e portanto pertence a U). Também podemos verificar que $S \neq U$: por exemplo, a sequência $\boldsymbol{u} = \{1,\ 1/2,\ 1/4,\ \dots\}$ ou seja, $u_n = 1/2^n$ pertence a U (pois $\sum_n = 0^\infty 1/(2^{2n}) = 1/(1-1/4) = 4/3 < \infty$), mas $\boldsymbol{u} \notin S$.

No entanto, quem é $S^\perp$? Suponha que $\boldsymbol{u} \in S^\perp$. Então, $\sum_{n=0}^\infty s_n u_n = 0$ para todo $\boldsymbol{s} \in S$. Ora, a sequência $\boldsymbol{s}_0$ dada por $s_{0,0} = 1$ e $s_{0,n} = 0$ para todo $n > 0$ é finita, e portanto pertence a S. Como $\boldsymbol{u}$ deve ser ortogonal a $\boldsymbol{s}$, concluímos que $u_0 = 0$. Mas a sequência $\boldsymbol{s}_1$ tal que $s_{1,1} = 1$, $s_{1,n} = 0$ para $n \neq 1$ também pertence a S. Como $\boldsymbol{u}$ também deve ser ortogonal a $\boldsymbol{s}_1$, concluímos que $u_1 = 0$. Continuando, vemos que $u_n = 0$ para todo n se $\boldsymbol{u} \in S^\perp$, portanto $S^\perp = \{\mathbf{0}\}$, em que $\mathbf{0}$ é a sequência $\{0,\ 0,\ \dots\}$.

Por fim, todos os vetores de U são ortogonais a $\mathbf{0}$, portanto $(S^\perp)^\perp = U \neq S$.

No entanto, para espaços de dimensão finita tudo fica mais fácil. Vejamos alguns resultados úteis. Dada a relação biunívoca entre um espaço de dimensão n e o $\mathbb{C}^n$ (ou o $\mathbb{R}^n$), vamos demonstrar o teorema abaixo para o $\mathbb{C}^n$.

Teorema 2.7. *Sejam S, T subespaços do $\mathbb{C}^n$. Então vale:*

1. $S \oplus S^\perp = \mathbb{C}^n$.
2. $(S^\perp)^\perp = S$
3. $(R \cap S)^\perp = R^\perp + S^\perp$

Demonstração. 1. Como S é um subespaço do $\mathbb{C}^n$, tem dimensão $k \leq n$. Se a dimensão de S for n, então $S = \mathbb{C}^n$ (verifique essa

afirmação), e $S^\perp = \{\mathbf{0}\}$, pois o único vetor que é ortogonal a todos os vetores do $\mathbb{C}^n$ é o vetor nulo, e nesse caso trivialmente vale $S \oplus S^\perp = \mathbb{C}^n$.

Considere então que $\dim(S) = k < n$, e seja $\mathcal{B}_S = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_k\}$ uma base ortonormal de S (que existe, de acordo com o Teorema 2.5). Vamos mostrar que existe um conjunto de vetores $\boldsymbol{u}_{k+1}, \ldots, \boldsymbol{u}_n$ tal que $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_k, \boldsymbol{u}_{k+1}, \ldots, \boldsymbol{u}_n\}$ seja uma base de $\mathbb{C}^n$. Sabemos que existe uma base $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ para o $\mathbb{C}^n$. Como $S \in \mathbb{C}^n$, o vetor $\boldsymbol{s}_1$ pode ser escrito como uma combinação linear dos $\boldsymbol{u}_k$, $k = 1 \ldots n$:

$$\boldsymbol{s}_1 = \alpha_1 \boldsymbol{u}_1 + \cdots + \alpha_n \boldsymbol{u}_n,$$

em que pelo menos um $\alpha_i \neq 0$ (pois o conjunto $\{\boldsymbol{s}_1, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ é l.d., e $\boldsymbol{s} \neq \mathbf{0}$. Vamos admitir (sem perda de generalidade) que $\alpha_1 \neq 0$, então podemos escrever também

$$\boldsymbol{u}_1 = \frac{1}{\alpha_1}(\boldsymbol{s}_1 - \alpha_2 \boldsymbol{u}_2 - \cdots - \alpha_n \boldsymbol{u}_n).$$

Os vetores $\boldsymbol{s}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ são portanto l.i., caso contrário seria possível escrever $\boldsymbol{u}_1$ como combinação linear dos $\boldsymbol{u}_i$, $i = 2 \ldots n$. Repetindo o argumento (e assumindo que os $\boldsymbol{u}_i$ estão ordenados convenientemente), podemos mostrar que $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_k, \boldsymbol{u}_{k+1}, \ldots, \boldsymbol{u}_n\}$ é l.i., e portanto forma uma base para o $\mathbb{C}^n$. Usando o algoritmo de Gram-Schmidt, podemos trocar os vetores $\boldsymbol{u}_{k+1}, \ldots, \boldsymbol{u}_n$ por vetores $\boldsymbol{v}_{k+1}, \ldots, \boldsymbol{v}_n$ de modo que $\mathcal{B} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_k, \boldsymbol{v}_{k+1}, \ldots, \boldsymbol{v}_n\}$ seja uma base ortonormal do $\mathbb{C}^n$.

Se valer $S^\perp = \mathrm{Sp}\{\boldsymbol{v}_{k+1}, \ldots, \boldsymbol{v}_n\}$, então o resultado estaria demonstrado, pois como $\mathcal{B}$ é uma base do $\mathbb{C}^n$, $S + S^\perp = \mathrm{Sp}\{\mathcal{B}\} = \mathbb{C}^n$, e pelo Teorema 2.6, $S \cap S^\perp = \{\mathbf{0}\}$.

Uma direção dessa igualdade é simples: se $\boldsymbol{u} \in \mathrm{Sp}\{\boldsymbol{v}_{k+1}, \ldots, \boldsymbol{v}_n\}$, e $\boldsymbol{s} \in S$, então $\boldsymbol{u} = \alpha_{k+1}\boldsymbol{v}_{k+1} + \cdots + \alpha_n \boldsymbol{v}_n$, e $\boldsymbol{s} = \beta_1 \boldsymbol{s}_1 + \cdots + \beta_k \boldsymbol{s}_k$, e portanto

$$< \boldsymbol{s}, \boldsymbol{u} >= \sum_{i=1}^{k} \sum_{j=k+1}^{n} < \beta_i \boldsymbol{s}_i, \alpha_j \boldsymbol{v}_j >= 0,$$

já que $< \boldsymbol{s}_i, \boldsymbol{v}_j >= 0$ sempre. Concluímos que $\mathrm{Sp}\{\boldsymbol{v}_{k+1}, \ldots, \boldsymbol{v}_n\} \subset S^{\perp}$.

Considere agora que $\boldsymbol{u} \in S^{\perp}$. Como $\mathcal{B}$ é uma base do $\mathbb{C}^n$, existem constantes α_i, $i = 1 \ldots n$ tais que

$$\boldsymbol{u} = \alpha_1 \boldsymbol{s}_1 + \cdots + \alpha_k \boldsymbol{s}_k + \alpha_{k+1}\boldsymbol{v}_{k+1} + \cdots + \alpha_n \boldsymbol{v}_n.$$

Vamos mostrar que $\alpha_i = 0$ para $1 \leq i \leq k$. Suponha por absurdo que $\alpha_i \neq 0$ para algum $i \leq k$. Então,

$$< \boldsymbol{s}_i, \boldsymbol{u} >= \alpha_i < \boldsymbol{s}_i, \boldsymbol{s}_i >= \alpha_i \neq 0,$$

e $\boldsymbol{u} \neq S^{\perp}$, uma contradição. Concluímos então que

$$S^{\perp} \subset \mathrm{Sp}\{\boldsymbol{v}_{k+1}, \ldots, \boldsymbol{v}_n\},$$

concluindo a demonstração.

2. Se $\boldsymbol{u} \in S$, então $< \boldsymbol{u}, \boldsymbol{v} >= 0$ para qualquer $\boldsymbol{v} \in S^{\perp}$, e portanto $\boldsymbol{u} \in (S^{\perp})^{\perp}$, ou seja, $S \subset (S^{\perp})^{\perp}$. Para provar que $(S^{\perp})^{\perp} \subset S$, vamos considerar dois casos. Se $S = \mathbb{C}^n$ ou $S = \{\mathbf{0}\}$, o resultado é direto. Para $S \neq \mathbb{C}^n$, $S \neq \{\mathbf{0}\}$, vamos usar a base $\mathcal{B}$ do item anterior. Vimos que $\{\boldsymbol{v}_{k+1}, \ldots, \boldsymbol{v}_n\}$ é uma base para $S^{\perp}$.

Seja agora $\boldsymbol{u} \in (S^{\perp})^{\perp}$. Suponha que $\boldsymbol{u} \notin S$. Como $S \oplus S^{\perp} = \mathbb{C}^n$, existem constantes α_i tais que

$$\boldsymbol{u} = \alpha_1 \boldsymbol{s}_1 + \cdots + \alpha_k \boldsymbol{s}_k + \alpha_{k+1} \boldsymbol{v}_{k+1} + \cdots + \alpha_n \boldsymbol{v}_n.$$

Se $\boldsymbol{u} \notin S$, então existe ao menos um $\alpha_m \neq 0$ para $k+1 \leq m \leq n$. Nesse caso,

$$< \boldsymbol{v}_m, \boldsymbol{u} >= \alpha_m \neq 0,$$

e $\boldsymbol{u} \notin \left(S^{\perp}\right)^{\perp}$, uma contradição. Portanto, $(S^{\perp})^{\perp} \subset S$, e concluímos a demonstração.

3. Para o terceiro resultado, basta usar o segundo resultado do Teorema 2.6 para $R^{\perp}$ e $S^{\perp}$, e depois o item anterior: $(R^{\perp} + S^{\perp})^{\perp} = (R^{\perp})^{\perp} \cap (S^{\perp})^{\perp} = R \cap S$. Usando o resultado do item anterior novamente, chegamos ao resultado desejado: $R^{\perp} + S^{\perp} = (R \cap S)^{\perp}$.

■

2.7 Aplicações

Vamos ver alguns exemplos de aplicações dos conceitos vistos até agora. Os exemplos iniciais serão necessariamente simples (ou incompletos), enquanto não tivermos visto ferramentas mais avançadas. Nos capítulos seguintes esses exemplos serão completados, e novos serão vistos.

2.7.1 Solução de equações de diferenças (ou diferenciais) lineares

Considere uma equação de diferenças do tipo

$$y(n) = -a_1 y(n-1) - a_2 y(n-2) - \cdots - a_N y(n-N) + b_0 x(n) + \cdots + b_M x(n-M), \tag{2.6}$$

para $n \geq 0$, com condições iniciais

$$y(-N) = y_{-N}, \quad y(-N+1) = y_{-N+1}, \quad \ldots \quad y(-1) = y_{-1} \tag{2.7}$$

dadas, e $x(n)$ sendo uma seqüência conhecida satisfazendo $x(n) = 0$ para $n < 0$.

A solução $y(n)$ de (2.6) pode ser calculada achando-se uma "solução particular" $x_{\mathrm{p}}(n)$ para (2.6), sem preocupação com as condições iniciais (2.7), e somando-se à solução particular uma solução conveniente da *equação homogênea*

$$y_{\mathrm{h}}(n) + a_1 y_{\mathrm{h}}(n-1) + a_2 y_{\mathrm{h}}(n-2) + \cdots + a_N y_{\mathrm{h}}(n-N) = 0, \quad \text{para } n \geq 0. \tag{2.8}$$

Vamos ver um exemplo simples: Considere o problema de segunda ordem

$$y(n) + y(n-1) + 0{,}24 y(n-2) = x(n), \; y(-1) = y(-2) = 0.$$

com $x(n) = \cos(\pi/3n)$.

Uma solução particular pode ser encontrada facilmente usando a resposta em freqüência desse sistema:

$$H(e^{j\omega}) = \frac{1}{1 + e^{-j\omega} + 0{,}24 e^{-2j\omega}}.$$

Para $\omega = \pi/3$, temos $H(e^{j\pi/3}) = 0{,}5719\ e^{j0{,}6613}$, e portanto

$$y_{\mathrm{p}} = 0{,}5719\ \cos(\pi/3n + 0{,}6613).$$

As soluções da equação homogênea são da forma $y_{\mathrm{h}}(n) = A\lambda^n$, em que A e λ são constantes. Para calcular as constantes, substituímos $y_{\mathrm{h}}(n)$ em (2.8):

$$A\lambda^n + A\lambda^{n-1} + 0{,}24A\lambda^{n-2} = 0 \Rightarrow A\lambda^n\big(1 + \lambda^{-1} + 0{,}24\lambda^{-2}\big) = 0.$$

Como a equação acima deve valer para todo n, o termo entre parênteses deve ser nulo. Resolvendo a equação de segundo grau resultante, obtemos $\lambda = 0{,}6$ ou $\lambda = 0{,}4$. Ou seja, encontramos duas soluções possíveis da equação homogênea: $y_{\mathrm{h}1}(n) = A\ 0{,}6^n$ e $y_{\mathrm{h}2}(n) = B\ 0{,}4^n$. Na verdade, se somarmos $y_{\mathrm{h}1}(n) + y_{\mathrm{h}2}(n)$, o resultado continua satisfazendo (2.8). Portanto, encontramos uma família de soluções de (2.8). Para resolver o problema completo então, fazemos

$$y(n) = y_{\mathrm{p}}(n) + y_{\mathrm{h}}(n) = 0{,}5719\ \cos(\pi/3n + 0{,}6613) + A\ 0{,}6^n + B\ 0{,}4^n.$$

Como queremos $y(-1) = y(-2) = 0$, chegamos a um sistema de equações

$$\begin{aligned} 0{,}5719\ \cos(-\pi/3 + 0{,}6613) + A\ 0{,}6^{-1} + B\ 0{,}4^{-1} &= 0 \\ 0{,}5719\ \cos(-2\pi/3 + 0{,}6613) + A\ 0{,}6^{-2} + B\ 0{,}4^{-2} &= 0, \end{aligned}$$

e resolvendo, obtemos $A = -0{,}8972$ e $B = 0{,}3862$.

Note que o conjunto das soluções da equação homogênea tem dois parâmetros livres neste exemplo. De fato, dentro do espaço vetorial V das seqüências reais $s(n)$ para $n \geq 0$, as soluções de (2.8) formam um

subespaço S: se $s_1(n)$ e $s_2(n)$ forem soluções de (2.8), então $\alpha s_1(n) + \beta s_2(n)$ também será uma solução, pois

$$\big(\alpha s_1(n) + \beta s_2(n)\big) + \sum_{i=1}^{N} a_i\big(\alpha s_1(n-i) + \beta s_2(n-i)\big) =$$

$$= \alpha\left[s_1(n) + \sum_{i=1}^{N} a_i s_1(n-i)\right] + \beta\left[s_2(n) + \sum_{i=1}^{N} a_i s_2(n-i)\right] = 0.$$

Esse subespaço das soluções da equação homogênea tem dimensão N em geral.

Uma solução y_{h} de (2.8) pode ser somada a uma solução particular y_{p} de (2.6), e o resultado ainda obedece (2.6), a menos das condições iniciais. Para resolver o problema completo (com as condições iniciais), deve-se encontrar uma solução de (2.8), que, somada à solução particular, permita que sejam obedecidas as condições (2.7). Isso é feito como no exemplo, calculando-se $y_{\mathrm{p}}(n) + y_{\mathrm{h}}(n)$ para $n = -1, \ldots, -N$. Como o subespaço das soluções da equação homogênea tem dimensão N, temos N parâmetros livres (incógnitas) e N equações (o sistema sempre tem solução, pois sempre pode-se encontrar exatamente N soluções l.i. da equação homogênea).

2.7.2 Detecção de senóides em ruído

Em várias aplicações, como determinação da direção de sinais em redes de antenas, processamento de sinais de voz ou sinais musicais, e em radares, o sinal recebido tem a forma

$$s(n) = \sum_{k=1}^{M} A_k e^{j\omega_k n + \phi_k} + r(n),$$

em que $r(n)$ é um ruído branco, e M, A, ω_k e ϕ_k são constantes que se deseja determinar. Vários dos métodos usados para resolver este problema são baseados na seguinte observação: se o ruído for nulo ($r(n) \equiv 0$), se montarmos um vetor $\boldsymbol{s}$, de dimensão N, formado de amostras de $s(n)$ a partir de um certo instante inicial n_0, isto é,

$$\boldsymbol{s} = \begin{bmatrix} s(n_0) & s(n_0+1) & \dots & s(n_0+N-1) \end{bmatrix}^T,$$

então o vetor $\boldsymbol{s}$ ficará restrito a um subespaço de dimensão M do $\mathbb{C}^N$. De fato, observe que

$$\boldsymbol{s} = \begin{bmatrix} A_1 e^{j(\omega_1 n_0+\phi_1)} \\ A_1 e^{j(\omega_1 n_0+\phi_1)} e^{j\omega_1} \\ A_1 e^{j(\omega_1 n_0+\phi_1)} e^{j2\omega_1} \\ \vdots \\ A_1 e^{j(\omega_1 n_0+\phi_1)} e^{j(N-1)\omega_1} \end{bmatrix} + \begin{bmatrix} A_2 e^{j(\omega_2 n_0+\phi_2)} \\ A_2 e^{j(\omega_2 n_0+\phi_2)} e^{j\omega_2} \\ A_2 e^{j(\omega_2 n_0+\phi_2)} e^{j2\omega_2} \\ \vdots \\ A_2 e^{j(\omega_2 n_0+\phi_2)} e^{j(N-1)\omega_2} \end{bmatrix} + \cdots +$$

$$+ \begin{bmatrix} A_M e^{j(\omega_M n_0+\phi_M)} \\ A_M e^{j(\omega_M n_0+\phi_M)} e^{j\omega_M} \\ A_M e^{j(\omega_M n_0+\phi_M)} e^{j2\omega_M} \\ \vdots \\ A_M e^{j(\omega_M n_0+\phi_M)} e^{j(N-1)\omega_M} \end{bmatrix} = A_1 e^{j(\omega_1 n_0+\phi_1)} \begin{bmatrix} 1 \\ e^{j\omega_1} \\ e^{j2\omega_1} \\ \vdots \\ e^{j(N-1)\omega_1} \end{bmatrix} + \cdots +$$

$$+ A_M e^{j(\omega_M n_0+\phi_M)} \begin{bmatrix} 1 \\ e^{j\omega_M} \\ e^{j2\omega_M} \\ \vdots \\ e^{j(N-1)\omega_M} \end{bmatrix}.$$

Note que, independentemente do instante inicial n_0 e das amplitudes A_k e fases ϕ_k, na ausência de ruído o vetor $\boldsymbol{s}$ é uma combinação linear de M vetores que dependem apenas das freqüências ω_k. Pode-se mostrar que esses vetores são sempre l.i. se as freqüências forem todas distintas entre si. Portanto, o espaço em que $\boldsymbol{s}$ pode estar *na ausência de ruído* é um subespaço de dimensão M do $\mathbb{C}^N$. Esta observação é a base de diversos algoritmos para se estimar M e os valores A_k, ϕ_k e ω_k, como veremos nos próximos capítulos.

Exercício 2.20. Prove que os vetores

$$\boldsymbol{s}_k = \begin{bmatrix} 1 & e^{j\omega_k} & e^{j2\omega_k} & \dots & e^{j(N-1)\omega_k} \end{bmatrix}^T, \; k = 1,2,\dots,M$$

formam um conjunto l.i. sempre que $\omega_i \neq \omega_j$ para todo $i \neq j$.

Capítulo 3

Transformações Lineares

Transformações em espaços vetoriais são funções que levam vetores de um espaço vetorial origem para outros vetores em um espaço vetorial destino. Recordando,

Definição 3.1. Uma *função* ou *transformação* f de um conjunto origem (também conhecido como *domínio* da função ou transformação) D para um conjunto destino (ou *contra-domínio* da transformação) E é uma relação entre elementos de D e E tal que, para todo elemento $x \in D$, existe um (e somente um) correspondente $y = f(x) \in E$.

Essa definição está colocada aqui para recordar que uma função é uma relação entre dois conjuntos (o domínio e o contra-domínio) que tem de satisfazer duas propriedades: 1) a função deve ser definida para *todos* os elementos do domínio; 2) o valor da função para cada elemento do domínio deve ser *único*. É importante lembrar dessas condições quando

formos tratar de inversas de transformações lineares e matrizes, principalmente quando definirmos pseudo-inversas.

Transformações *lineares* são funções que satisfazem uma propriedade simples de aditividade:

Definição 3.2. Uma *transformação linear* de um espaço vetorial U para um espaço vetorial V (ambos sobre o mesmo conjunto de escalares $\mathbb{F}$) é uma função $f : U \to V$ satisfazendo[1]

$$f(\alpha \boldsymbol{x} + \beta \boldsymbol{y}) = \alpha f(\boldsymbol{x}) + \beta f(\boldsymbol{y}),$$

para todos os vetores $\boldsymbol{x}, \boldsymbol{y} \in U$ e quaisquer $\alpha, \beta \in \mathbb{F}$.

Repare que a definição de transformação linear é bem restritiva: por exemplo, se T é uma transformação linear, então

$$T(\mathbf{0}_U) = T(\mathbf{0}_U + \mathbf{0}_U) = T(\mathbf{0}_U) + T(\mathbf{0}_U) \Rightarrow T(\mathbf{0}_U) = \mathbf{0}_V,$$

ou seja, uma transformação linear T de U em V sempre leva o vetor nulo de U, $\mathbf{0}_U$, para o vetor nulo de V, $\mathbf{0}_V$.

Se o conceito de transformação linear é assim "restritivo", por que então estudá-lo a fundo? As principais razões são:

1. A propriedade de linearidade e a estrutura de espaço vetorial permitem que o efeito de uma transformação linear seja muito bem

[1] A partir deste capítulo, dada a relação de equivalência (o *isomorfismo*) entre um espaço vetorial de dimensão n sobre os escalares $\mathbb{F}$ e o $\mathbb{F}^n$ (visto na Seção 2.5.1), quase sempre que formos tratar apenas de espaços de dimensão finita vamos trabalhar diretamente com o $\mathbb{F}^n$, ou, mais especificamente, com o $\mathbb{C}^n$ (já que os resultados equivalentes para o $\mathbb{R}^n$ são essencialmente iguais ou mais simples). Quando quisermos tratar de um conceito que vale tanto para espaços de dimensão finita quanto de dimensão infinita, daremos as definições em função de espaços genéricos U, V, etc.

compreendido, tanto do ponto de vista teórico, quanto, por meio de algoritmos bastante eficientes (rápidos e precisos), do ponto de vista prático. Isto significa que um problema real modelado por transformações lineares pode ser analisado de maneira eficiente e bastante profunda usando as ferramentas que veremos ao longo do curso.

2. Muitas operações de interesse prático são lineares ou aproximadamente lineares, como por exemplo, a distorção observada em uma imagem feita por uma lente desfocada, a mudança das coordenadas de um corpo rígido devida a uma rotação, a derivada de uma função (e, sob certas condições, também a integral), um atraso em uma seqüência, ou mais geralmente a transformação efetuada em uma seqüência ou função por um filtro linear, a transformada de Fourier (de tempo discreto ou contínuo), o cálculo de estimativas de sinais Gaussianos segundo o critério de máxima verossimilhança.

3. Vários problemas cujas soluções ótimas não envolvem transformações lineares são resolvidos por meio sequências de transformações lineares devido aos algoritmos eficientes e ao estado avançado da teoria disponíveis para estas últimas; por exemplo, vários algoritmos iterativos para cálculo de zeros de funções ou para otimização, métodos para integração numérica de equações diferenciais, o cálculo de estimativas lineares para sinais não-Gaussianos, o filtro de Kalman estendido, etc.

Vamos ver neste capítulo algumas propriedades mais simples de transformações lineares, principalmente a relação entre transformações lineares sobre espaços vetoriais de dimensão finita e matrizes. Logo será visto

que essa relação é muito útil para se entender corretamente algumas propriedades de matrizes e de transformações lineares.

3.1 Matrizes e Transformações Lineares

Dada uma transformação linear $f : \mathbb{C}^n \to \mathbb{C}^m$ e bases $\mathcal{B}_n = \{\boldsymbol{u}_1 \dots \boldsymbol{u}_n\}$ e $\mathcal{B}_m = \{\boldsymbol{v}_1 \dots \boldsymbol{v}_m\}$ para $\mathbb{C}^n$ e $\mathbb{C}^m$, podemos descrever a transformação linear por meio de uma matriz, da seguinte maneira. Para cada $\boldsymbol{x} \in \mathbb{C}^n$, temos

$$\boldsymbol{x} = \sum_{j=1}^{n} \alpha_j \boldsymbol{u}_j \Rightarrow \boldsymbol{y} = f(\boldsymbol{x}) = \sum_{j=1}^{n} \alpha_j f(\boldsymbol{u}_j).$$

Como $f(\boldsymbol{u}_j) \in \mathbb{C}^m$, e $\mathcal{B}_m$ é uma base de $\mathbb{C}^m$, vale $f(\boldsymbol{u}_j) = \sum_{i=1}^{m} \beta_{i,j} \boldsymbol{v}_i$. O resultado final é que $f(\boldsymbol{x})$ pode ser escrita em termos da base $\mathcal{B}_m$, como

$$f(\boldsymbol{x}) = \sum_{j=1}^{n} \alpha_j \left(\sum_{i=1}^{m} \beta_{i,j} \boldsymbol{v}_i \right) = \sum_{i=1}^{m} \boldsymbol{v}_i \left(\sum_{j=1}^{n} \beta_{i,j} \alpha_j \right).$$

Agora, juntando os α_j e $\beta_{i,j}$ em forma de vetor e matriz,

$$(\boldsymbol{x})_{\mathcal{B}_n} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \qquad \boldsymbol{A} = \begin{bmatrix} \beta_{1,1} & \beta_{1,2} & \dots & \beta_{1,n} \\ \beta_{2,1} & \beta_{2,2} & \dots & \beta_{2,n} \\ \vdots & & \ddots & \vdots \\ \beta_{m,1} & \beta_{m,2} & \dots & \beta_{m,n} \end{bmatrix},$$

$f(\boldsymbol{x})$, na base $\mathcal{B}_m$, se escreve como

$$\big(f(\boldsymbol{x})\big)_{\mathcal{B}_m} = \boldsymbol{A}.(\boldsymbol{x})_{\mathcal{B}_n}.$$

Além disso, dadas $\mathcal{B}_n$, $\mathcal{B}_m$ e f, a matriz $\boldsymbol{A}$ é única. Vamos denotar os elementos da matriz $\boldsymbol{A}$ por $[\boldsymbol{A}]_{ij}$ ou por a_{ij}, o primeiro índice correspondendo a linhas e o segundo a colunas.

Repare que as definições habituais de soma e produto entre matrizes são definidas calculando-se a matriz que representa as transformações lineares $(f+g)(\boldsymbol{x}) \triangleq f(\boldsymbol{x}) + g(\boldsymbol{x})$ e $(f \circ g)(\boldsymbol{x}) \triangleq f\big(g(\boldsymbol{x})\big)$.

Teorema 3.1 (Soma de matrizes)**.** *Sejam f e g duas transformações lineares de $\mathbb{C}^n$ para $\mathbb{C}^m$, e defina a* soma *de f e g por*

$$(f+g)(\boldsymbol{x}) = f(\boldsymbol{x}) + g(\boldsymbol{x}).$$

Então, dadas as bases $\mathcal{B}_n$ de $\mathbb{C}^n$ e $\mathcal{B}_m$ de $\mathbb{C}^m$, se a matriz correspondente a f for $\boldsymbol{A} \in \mathbb{C}^{m\times n}$, e a matriz correspondente a g for $\boldsymbol{B} \in \mathbb{C}^{m\times n}$, então a matriz $\boldsymbol{C}$ correspondente a $f+g$ será

$$\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B},$$

em que o elemento c_{ij} de $\boldsymbol{C}$ é dado por $c_{ij} = a_{ij} + b_{ij}$.

Teorema 3.2 (Multiplicação de matrizes)**.** *Seja f uma transformação linear de $\mathbb{C}^p$ para $\mathbb{C}^m$, e g uma transformação linear de $\mathbb{C}^n$ para $\mathbb{C}^p$. Sejam também $\boldsymbol{A}$ e $\boldsymbol{B}$ as matrizes correspondentes a f e g, respectivamente, definidas a partir das bases $\mathcal{B}_n$, $\mathcal{B}_p$ e $\mathcal{B}_m$ de $\mathbb{C}^n$, $\mathbb{C}^p$ e $\mathbb{C}^m$. Defina $(f\circ g)(\boldsymbol{x}) = f\big(g(\boldsymbol{x})\big)$. Então a matriz correspondente a $f\circ g$ com respeito às bases $\mathcal{B}_n$ e $\mathcal{B}_m$ é*

$$\boldsymbol{C} = \boldsymbol{A} \cdot \boldsymbol{B},$$

em que $c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj}$.

Exercício 3.1. Prove os teoremas 3.1 e 3.2.

3.2 Imagem, núcleo, e posto

O subspaço *imagem* (ou *range*) de uma transformação linear $f: U \to V$ é o conjunto

$$\mathcal{R}(f) = \{\boldsymbol{v} \in V : \ \boldsymbol{v} = f(\boldsymbol{u}), \, \boldsymbol{u} \in U\} \subseteq V.$$

$\mathcal{R}(f)$ é realmente um subespaço de V, pois se $\boldsymbol{v}_1$, $\boldsymbol{v}_2 \in \mathcal{R}(f)$, então existem $\boldsymbol{u}_1$ e $\boldsymbol{u}_2$ em U tais que $\boldsymbol{v}_1 = f(\boldsymbol{u}_1)$ e $\boldsymbol{v}_2 = f(\boldsymbol{u}_2)$ (por definição). Pela linearidade da f, temos então, para quaisquer α_1, $\alpha_2 \in \mathbb{F}$,

$$\boldsymbol{v} \triangleq \alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 = \alpha_1 f(\boldsymbol{u}_1) + \alpha_2 f(\boldsymbol{u}_2) = f(\alpha_1 \boldsymbol{u}_1 + \alpha_2 \boldsymbol{u}_2).$$

Como $\boldsymbol{u} \triangleq \alpha_1 \boldsymbol{u}_1 + \alpha_2 \boldsymbol{u}_2 \in U$, concluímos que $\boldsymbol{v} \in \mathcal{R}(f)$, e portanto $\mathcal{R}(f)$ é um subespaço.

Podemos estender esta definição para matrizes, lembrando a relação entre matrizes e transformações lineares. Assuma que U tenha dimensão n e V tenha dimensão m, e considere bases $\mathcal{B}_n$ para U e $\mathcal{B}_m$ para V. Assuma que, com as bases escolhidas, a matriz que representa f é $\boldsymbol{A} \in \mathbb{C}^{m \times n}$. Note que, como $\mathcal{B}_m \triangleq \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$ é uma base de V, há uma relação biunívoca entre os vetores $\boldsymbol{y} \in \mathbb{C}^m = \begin{bmatrix} y_1 & y_2 & \ldots & y_m \end{bmatrix}^T$ e os vetores $\boldsymbol{v} \in V$, de modo que para todo $\boldsymbol{v} \in V$ existe um único $\boldsymbol{y} \in \mathbb{C}^m$ tal que

$$\boldsymbol{v} = y_1 \boldsymbol{v}_1 + y_2 \boldsymbol{v}_2 + \cdots + y_m \boldsymbol{v}_m,$$

e vice-versa: para todo $\boldsymbol{y} \in \mathbb{C}^m$ existe um único $\boldsymbol{v} \in V$ correspondente.

Usando essa relação, o espaço-imagem de f, na base $\mathcal{B}_m$, é dado portanto por

$$\{\boldsymbol{y} \in \mathbb{C}^m : \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}, \, \boldsymbol{x} \in \mathbb{C}^n\}.$$

Dizemos então que o subespaço imagem da $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ é:

$$\mathcal{R}(\boldsymbol{A}) = \{\boldsymbol{y} \in \mathbb{C}^m : \ \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x},\, \boldsymbol{x} \in \mathbb{C}^n\} \subseteq \mathbb{C}^m.$$

Definindo $\boldsymbol{a}_i = i$-ésima coluna de $\boldsymbol{A}$, podemos escrever

$$\boldsymbol{A}\boldsymbol{x} = \begin{bmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2 & \dots & \boldsymbol{a}_n\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ \vdots\\ x_n\end{bmatrix} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \dots x_n\boldsymbol{a}_n,$$

ou seja, $\mathcal{R}(\boldsymbol{A})$ é o subspaço formado pelas combinações lineares das colunas de $\boldsymbol{A}$.

Exemplo 3.1. Considere a transformação linear em V_7 (polinômios reais de grau menor ou igual a 5):

$$\begin{aligned} D:\ & V_7 \to V_7 \\ & p \mapsto D(p) = \frac{\mathrm{d}\,p(x)}{\mathrm{d}\,x}, \end{aligned}$$

isto é, a derivada. O subespaço imagem de D é portanto o conjunto dos polinômios com grau menor ou igual a 4.

Usando a base $\mathcal{B} = \{1,\ x,\ x^2,\ x^3,\ x^4,\ x^5\}$ para V_7, a matriz que descreve a transformação D é

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

O espaço-imagem de $\boldsymbol{D}$ é o conjunto $\mathcal{R}(\boldsymbol{D}) = \{\boldsymbol{x} \in \mathbb{R}^6 : \ [\boldsymbol{x}]_6 = 0\}$.

Exemplo 3.2. Considere agora a transformação D_1 definida em V_6 (polinômios de grau finito):

$$\begin{aligned} D_1 : \ V_6 &\to V_6 \\ p &\mapsto D_1(p) = \frac{\mathrm{d}\, p(x)}{\mathrm{d}\, x}, \end{aligned}$$

O subespaço imagem de D_1 é novamente V_6, pois dado um polinômio qualquer contido em V_6,

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

podemos encontrar um polinômio q tal que $p = D_1(q)$, por exemplo,

$$q(x) = a_0 x + a_1 x^2 + \cdots + a_n x^{n+1}.$$

Exercício 3.2. Encontre o subespaço imagem da matriz

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

O *posto* (ou *rank*) da transformação linear f é a dimensão do espaço-imagem de f, ou seja

$$\mathrm{posto}(f) = \dim(\mathcal{R}(f)).$$

Por exemplo, o posto da transformação D acima é 5, pois o espaço dos polinômios com grau menor ou igual a 4 tem dimensão 5.

Analogamente, o *posto* de $\boldsymbol{A}$ é a dimensão de $\mathcal{R}(\boldsymbol{A})$. O posto de $\boldsymbol{A}$ corresponde portanto ao número de colunas l.i. de $\boldsymbol{A}$ (veremos que o posto de $\boldsymbol{A}$ é também igual ao número de *linhas* l.i. de $\boldsymbol{A}$, isto é, o posto de $\boldsymbol{A}$ é igual ao posto de $\boldsymbol{A}^T$). Assim, o posto da matriz $\boldsymbol{D}$ do exemplo acima é igual a 5. Note que, como a matriz $\boldsymbol{D}$ descreve a transformação D se for usada a base $\mathcal{B}$ para representar os vetores de V_7, claramente posto($\boldsymbol{D}$) = posto(D). Como sempre com matrizes, às vezes é interessante pensar na definição de posto em termos de número de colunas linearmente independentes da matriz, às vezes é mais conveniente pensar na dimensão do espaço-imagem de uma transformação linear descrita pela matriz.

A importância dos conceitos de espaço-imagem e posto de matrizes vem principalmente da seguinte questão, que aparece quando se quer resolver um sistema de equações lineares: Dados $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ e $\boldsymbol{b} \in \mathbb{C}^m$, quando existe uma solução para $\boldsymbol{Ax} = \boldsymbol{b}$?

Claramente, a solução existe apenas se $\boldsymbol{b} \in \mathcal{R}(\boldsymbol{A})$. Quando existe uma solução para *qualquer* $\boldsymbol{b} \in \mathbb{C}^m$? Quando $\mathcal{R}(\boldsymbol{A}) = \mathbb{C}^m$, ou seja, quando posto($\boldsymbol{A}$) $= m$. Como veremos mais à frente, essas perguntas reaparecem frequentemente em aplicações. Veremos a seguir algumas técnicas para avaliar o posto e o espaço-imagem de uma matriz qualquer. O primeiro resultado é simples, como veremos no teorema abaixo.

Teorema 3.3. *O posto de uma transformação linear f de $\mathbb{C}^n$ para $\mathbb{C}^m$ (ou o posto de uma matriz $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ é menor ou igual ao mínimo de m, n.*

Demonstração. Como $\mathcal{R}(\boldsymbol{A}) \subset \mathbb{C}^m$, é claro que posto($\boldsymbol{A}$) $\leq m$. Para verificar a outra desigualdade, considere um conjunto de $k > n$ vetores $\{\boldsymbol{y}_i \in \mathcal{R}(\boldsymbol{A})\}_{i=1}^k$. Para mostrar que esse conjunto é necessariamente l.d.,

vamos estudar as soluções possíveis de

$$\alpha_1 \boldsymbol{y}_1 + \cdots + \alpha_k \boldsymbol{y}_k = \boldsymbol{0}.$$

Como $\boldsymbol{y}_i \in \mathcal{R}(\boldsymbol{A})$, existem $\{\boldsymbol{x}_i \in \mathbb{C}^n\}_{i=1}^k$ tais que $\boldsymbol{y}_i = \boldsymbol{A}\boldsymbol{x}_i$, portanto a equação anterior fica

$$\alpha_1 \boldsymbol{A}\boldsymbol{x}_1 + \cdots + \alpha_k \boldsymbol{A}\boldsymbol{x}_k = \boldsymbol{0} \Leftrightarrow \boldsymbol{A}(\alpha_1 \boldsymbol{x}_1 + \cdots + \alpha_k \boldsymbol{x}_k) = \boldsymbol{0}. \tag{3.1}$$

O conjunto $\{\boldsymbol{x}_i\}$ é l.d., pois os $\boldsymbol{x}_i \in \mathbb{C}^n$, e $n < k$. Portanto, existe um conjunto de escalares α_i com ao menos um $\alpha_{i_0} \neq 0$, tal que

$$\alpha_1 \boldsymbol{x}_1 + \cdots + \alpha_k \boldsymbol{x}_k = \boldsymbol{0}.$$

Substituindo esse resultado em (3.1), vemos que os $\boldsymbol{y}_i$ são também l.d. — ou seja, não há nenhum conjunto de vetores $\{\boldsymbol{y}_i \in \mathcal{R}(\boldsymbol{A})\}_{i=1}^k$ com $k > n$ que seja l.i., o que significa que posto$(\boldsymbol{A}) \leq n$.

■

O espaço-imagem ajuda a responder à pergunta "quando existe uma solução para um sistema de equações lineares?" O *núcleo* (ou *kernel*) vai ajudar a responder à pergunta "quando a solução de um sistema de equações lineares é única?" Vejamos, imagine que $\boldsymbol{x} \in \mathbb{C}^n$ seja uma solução de $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, com $\boldsymbol{b} \in \mathbb{C}^m$, $\boldsymbol{A} \in \mathbb{C}^{m \times n}$. Imagine que exista uma outra solução, $\boldsymbol{x}_1$, para o mesmo sistema, isto é, $\boldsymbol{A}\boldsymbol{x}_1 = \boldsymbol{b}$. Então, subtraindo as duas equações, temos

$$\boldsymbol{A}\boldsymbol{x} - \boldsymbol{A}\boldsymbol{x}_1 = \boldsymbol{b} - \boldsymbol{b} = \boldsymbol{0} \Rightarrow \boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}_1) = \boldsymbol{0}.$$

Note que $\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{x}_1)=\boldsymbol{0}$ não significa necessariamente que $\boldsymbol{x}=\boldsymbol{x}_1$! Por exemplo, considere a matriz

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

Se $\boldsymbol{x}-\boldsymbol{x}_1=\begin{bmatrix}0\\ 1\end{bmatrix}^T$, então $\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{x}_1)=\boldsymbol{0}$, mas $\boldsymbol{x}-\boldsymbol{x}_1\neq\boldsymbol{0}$.

Vemos que a unicidade da solução de um sistema de equações lineares está relacionada ao conjunto

$$\mathcal{N}(\boldsymbol{A})\triangleq\{\boldsymbol{x}\in\mathbb{C}^n:\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}\}\subseteq\mathbb{C}^n,$$

denominado *núcleo* de $\boldsymbol{A}$.

A mesma ideia vale para uma transformação linear qualquer. Seja então f uma transformação linear de U para V. O *núcleo* de f é o conjunto

$$\mathcal{N}(f)=\{\boldsymbol{u}\in U:f(\boldsymbol{u})=\boldsymbol{0}\}\subseteq U.$$

Ou seja, $\mathcal{N}(f)$ é o subconjunto de U que contém todos os vetores que f leva para o vetor nulo em V.

O núcleo é um subespaço de U, pois, se $\boldsymbol{x}_1$, $\boldsymbol{x}_2\in\mathcal{N}(f)$, então para quaisquer α_1, $\alpha_2\in\mathbb{F}$,

$$f(\alpha_1\boldsymbol{u}_1+\alpha_2\boldsymbol{u}_2)=\alpha_1 f(\boldsymbol{u}_1)+\alpha_2 f(\boldsymbol{u}_2)=\boldsymbol{0},$$

e portanto $\boldsymbol{u}\triangleq\alpha_1\boldsymbol{u}_1+\alpha_2\boldsymbol{u}_2\in\mathcal{N}(f)$. A dimensão do núcleo de f é às vezes denominada *nulidade* de f (igualmente para a dimensão do núcleo de uma matriz $\boldsymbol{A}$). Indicaremos então $\mathrm{nul}(\boldsymbol{A})=\dim(\mathcal{N}(\boldsymbol{A}))$.

Com a definição de núcleo podemos responder à pergunta sobre a unicidade das soluções de um sistema de equações lineares:

Teorema 3.4. *Considere o sistema de equações lineares* $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$*, em que* $\boldsymbol{x} \in \mathbb{C}^n$*,* $\boldsymbol{b} \in \mathbb{C}^m$*, e* $\boldsymbol{A} \in \mathbb{C}^{m\times n}$*. Admitindo-se que exista uma solução, a solução será única se e somente se* $\mathcal{N}(\boldsymbol{A}) = \{\boldsymbol{0}\}$*, ou seja, a solução (se existir) é única se e somente se a dimensão do núcleo de* $\boldsymbol{A}$ *(a nulidade de* $\boldsymbol{A}$*) for zero.*

Demonstração. Já mostramos que se existirem duas soluções $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ e $\boldsymbol{A}\boldsymbol{x}_1 = \boldsymbol{b}$, então $\boldsymbol{x} - \boldsymbol{x}_1 \in \mathcal{N}(\boldsymbol{A})$. Portanto, se houver soluções distintas, haverá ao menos um vetor não nulo em $\mathcal{N}(\boldsymbol{A})$. Por outro lado, se a dimensão de $\mathcal{N}(\boldsymbol{A})$ maior que zero, haverá um vetor não nulo $\boldsymbol{w} \in \mathcal{N}(\boldsymbol{A})$. Se $\boldsymbol{x}$ for uma solução de $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, então $\boldsymbol{A}(\boldsymbol{x}+\boldsymbol{w}) = \boldsymbol{A}\boldsymbol{x}+\boldsymbol{A}\boldsymbol{w} = \boldsymbol{b}+\boldsymbol{0} = \boldsymbol{b}$, ou seja, $\boldsymbol{x} + \boldsymbol{w} \neq \boldsymbol{x}$ também é uma solução do sistema. ■

Exemplo 3.3. O núcleo da transformação D do exemplo 3.1 tem dimensão 1, pois apenas os polinômios da forma $p(x) = c$, em que c é uma constante, têm derivada nula. Assim, $\mathcal{N}(D) = \{p \in V_7 : p(x) = c\}$. Analogamente, o núcleo da matriz $\boldsymbol{D}$ é $\mathcal{N}(\boldsymbol{D}) = \{\boldsymbol{x} \in \mathbb{R}^6 : [\boldsymbol{x}]_2 = [\boldsymbol{x}]_3 = \cdots = [\boldsymbol{x}]_6 = 0\}$, ou seja, o núcleo de $\boldsymbol{D}$ são os vetores da forma $\alpha \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^T$.

3.2.1 Subespaços fundamentais de uma transformação linear

Nesta seção, vamos utilizar o produto interno usual do $\mathbb{C}^n$: $< \boldsymbol{x}, \boldsymbol{y} >= \boldsymbol{y}^H\boldsymbol{x}$, em que $\boldsymbol{y}^H$ é o vetor transposto e conjugado.

Considere a matriz $\boldsymbol{A} \in \mathbb{C}^{m\times n}$, e a transformação linear

$$\begin{aligned} f:\ & \mathbb{C}^n \to \mathbb{C}^m \\ & \boldsymbol{x} \longmapsto \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}. \end{aligned}$$

A matriz $\boldsymbol{A}^H = (\boldsymbol{A}^*)^T$ (transposta e com todos os elementos conjugados) define uma transformação linear de $\mathbb{C}^m$ para $\mathbb{C}^n$

$$\begin{aligned} f^H : \ & \mathbb{C}^m \to \mathbb{C}^n \\ & \boldsymbol{y} \longmapsto \boldsymbol{x} = \boldsymbol{A}^H \boldsymbol{y}. \end{aligned}$$

Considere os quatro subespaços $\mathcal{R}(\boldsymbol{A})$, $\mathcal{N}(\boldsymbol{A})$, $\mathcal{R}(\boldsymbol{A}^H)$ e $\mathcal{N}(\boldsymbol{A}^H)$. O seguinte teorema relaciona esses quatro subespaços:

Teorema 3.5 (Quatro subespaços fundamentais)**.** *Considere* $\boldsymbol{A} \in \mathbb{C}^{m\times n}$. *Então*

1. $\mathcal{N}(\boldsymbol{A}) = \mathcal{R}(\boldsymbol{A}^H)^\perp$.
2. $\mathcal{R}(\boldsymbol{A}) = \mathcal{N}(\boldsymbol{A}^H)^\perp$,

Demonstração. 1. Vamos provar inicialmente que $\mathcal{N}(\boldsymbol{A}) \subset \mathcal{R}(\boldsymbol{A})^\perp$. Para qualquer $\boldsymbol{x} \in \mathcal{N}(\boldsymbol{A})$ e qualquer $\boldsymbol{u} \in \mathcal{R}(\boldsymbol{A}^H)$ (por definição, existe $\boldsymbol{y} \in \mathbb{C}^m$ tal que $\boldsymbol{u} = \boldsymbol{A}^H\boldsymbol{y}$), vale

$$\boldsymbol{x}^H\boldsymbol{u} = \boldsymbol{x}^H\boldsymbol{A}^H\boldsymbol{y} = (\boldsymbol{A}\boldsymbol{x})^H\boldsymbol{y} = 0,$$

ou seja, $\boldsymbol{x} \in \mathcal{R}(\boldsymbol{A})^\perp$. Por outro lado, seja $\boldsymbol{v} \in \mathcal{R}(\boldsymbol{A}^H)^\perp$, e suponha que $\boldsymbol{A}\boldsymbol{v} \neq \boldsymbol{0}$. Então existiria um vetor $\boldsymbol{y} \in \mathbb{C}^m$ tal que

$$\boldsymbol{v}^H\boldsymbol{A}^H\boldsymbol{y} \neq 0.$$

No entanto, $\boldsymbol{A}^H\boldsymbol{y} \in \mathcal{R}(\boldsymbol{A}^H)$, e concluímos que $\boldsymbol{v} \notin \mathcal{R}(\boldsymbol{A}^H)^\perp$, uma contradição.

2. Para demonstrar esse segundo resultado, use o segundo resultado do Teorema 2.7 no resultado anterior para verificar que $\mathcal{R}(\boldsymbol{A}^H) = \mathcal{N}(\boldsymbol{A})^\perp$, e substitua $\boldsymbol{A}$ por $\boldsymbol{A}^H$ e vice-versa.

■

A divisão do $\mathbb{C}^n$ e do $\mathbb{C}^m$ em termos dos espaços-imagem e núcleos de $\boldsymbol{A}$ e $\boldsymbol{A}^H$ tem diversas consequências interessantes, que levam finalmente à decomposição em valores singulares, que veremos no Capítulo 8, e vão dar origem a algoritmos para solução de problemas de mínimos quadrados. Por hora vamos apenas ver, no próximo teorema, que $\text{posto}(\boldsymbol{A}) = \text{posto}(\boldsymbol{A}^H)$, ou seja, o número de colunas l.i. de $\boldsymbol{A}$ é igual ao número de colunas l.i. de $\boldsymbol{A}^H$ (que é igual ao número de *linhas* l.i. de $\boldsymbol{A}$).

Teorema 3.6. *Seja* $\boldsymbol{A} \in \mathbb{C}^{m\times n}$. *Então,* $\text{posto}(\boldsymbol{A}) = \text{posto}(\boldsymbol{A}^H)$.

Demonstração. Vamos demonstrar esse resultado mostrando primeiro que $\dim(\mathcal{R}(\boldsymbol{A})) \leq \dim(\mathcal{R}(\boldsymbol{A}^H))$, e depois que a desigualdade também vale no sentido contrário. Tome um vetor $\boldsymbol{y} \in \mathcal{R}(\boldsymbol{A})$. Por definição, existe $\boldsymbol{x} \in \mathbb{C}^n$ tal que $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$. Sabemos (pelo Teorema 2.7) que $\mathbb{C}^n = \mathcal{N}(\boldsymbol{A}) \oplus \mathcal{N}(\boldsymbol{A})^\perp$, e portanto podemos escrever $\boldsymbol{x}$ de maneira única como

$$\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2,$$

em que $\boldsymbol{x}_1 \in \mathcal{N}(\boldsymbol{A})^\perp$ e $\boldsymbol{x}_2 \in \mathcal{N}(\boldsymbol{A})$ (Teorema 2.2). Pelo Teorema 3.5, $\mathcal{N}(\boldsymbol{A})^\perp = \mathcal{R}(\boldsymbol{A}^H)$, portanto $\boldsymbol{x}_1 \in \mathcal{R}(\boldsymbol{A}^H)$. Como $\boldsymbol{A}\boldsymbol{x}_2 = \boldsymbol{0}$, temos

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x}_1,$$

ou seja, *todo vetor* $\boldsymbol{y} \in \mathcal{R}(\boldsymbol{A})$ pode ser escrito como $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_1$, com $\boldsymbol{x}_1 \in \mathcal{R}(\boldsymbol{A}^H)$. Usando os mesmos argumentos do Teorema 3.3, vemos que isso significa que $\dim(\mathcal{R}(\boldsymbol{A})) \leq \dim(\mathcal{R}(\boldsymbol{A}^H))$, ou seja, que $\text{posto}(\boldsymbol{A}) \leq \text{posto}(\boldsymbol{A}^H)$.

O argumento acima pode ser repetido trocando $\boldsymbol{A}$ com $\boldsymbol{A}^H$, o que leva à conclusão que $\text{posto}(\boldsymbol{A}^H) \leq \text{posto}(\boldsymbol{A})$, e portanto $\text{posto}(\boldsymbol{A}) = \text{posto}(\boldsymbol{A}^H)$.

■

Corolário 3.1. *Para* $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ *vale*

$$\operatorname{posto}(\boldsymbol{A}) + \operatorname{nul}(\boldsymbol{A}^H) = m, \qquad \operatorname{posto}(\boldsymbol{A}^H) + \operatorname{nul}(\boldsymbol{A}) = n.$$

Demonstração. Fica como exercício. ■

3.2.2 Inversas e pseudo-inversas

É comum ser necessário verificar quando uma determinada matriz $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ possui inversa. Por exemplo, se sabemos que $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$, podemos concluir que $\boldsymbol{x} = \boldsymbol{0}$? Para responder a essa pergunta, o primeiro impulso é verificar se o determinante de $\boldsymbol{A}$ é nulo ou não. No entanto, nem sempre essa é a melhor forma de verificar se uma matriz tem inversa — pode ser mais fácil determinar a dimensão do núcleo de $\boldsymbol{A}$, por exemplo. Vejamos um conjunto de condições que podem ser usadas para determinar se uma matriz possui inversa.

Teorema 3.7. *Seja* $\boldsymbol{A} \in \mathbb{C}^{n\times n}$. *As seguintes afirmações são equivalentes:*

(a) $\operatorname{posto}(\boldsymbol{A}) = n$,

(b) $\operatorname{nul}(\boldsymbol{A}) = 0$,

(c) $\boldsymbol{A}$ *tem linhas l.i.,*

(d) $\boldsymbol{A}$ *tem colunas l.i.,*

(e) Existe uma única matriz $\boldsymbol{A}^{-1} \in \mathbb{C}^{n\times n}$ *tal que*

$$\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}_n.$$

Demonstração. É necessário demonstrar que qualquer uma das afirmações acima implica todas as outras, ou (por exemplo), que (a) implica (b), (b) implica (c), etc, até (e) implica (a). Vamos demonstrar apenas que (d) implica (e) e (e) implica (a), as outras equivalências seguem diretamente de definições e teoremas anteriores.

Considere então que $\boldsymbol{A}$ tem colunas l.i.. Isso significa que $\mathcal{R}(\boldsymbol{A}) = \mathbb{C}^n$ e que $\mathcal{N}(\boldsymbol{A}) = \{\boldsymbol{0}\}$. Portanto, a transformação linear

$$\begin{aligned} f:\ \mathbb{C}^n &\to \mathbb{C}^n \\ \boldsymbol{x} &\mapsto \boldsymbol{A}\boldsymbol{x}, \end{aligned}$$

é sobrejetora (pois $\mathcal{R}(f) = \mathbb{C}^n$) e injetora (pelo Teorema 3.4). Uma função sobrejetora e injetora possui uma inversa f^{-1} tal que

$$\begin{aligned} f^{-1}:\ \mathbb{C}^n &\to \mathbb{C}^n \\ \boldsymbol{y} &\mapsto \boldsymbol{x}, \text{ tal que } \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}. \end{aligned}$$

A transformação f^{-1} é linear, pois para $\boldsymbol{y}_1$, $\boldsymbol{y}_2 \in \mathbb{C}^n$, existem $\boldsymbol{x}_1$, $\boldsymbol{x}_2 \in \mathbb{C}^n$ tais que $\boldsymbol{y}_1 = \boldsymbol{A}\boldsymbol{x}_1$, $\boldsymbol{y}_2 = \boldsymbol{A}\boldsymbol{x}_2$, e $\boldsymbol{y}_1 + \boldsymbol{y}_2 = \boldsymbol{A}(\boldsymbol{x}_1 + \boldsymbol{x}_2)$. Portanto, $f^{-1}(\boldsymbol{y}_1 + \boldsymbol{y}_2) = f^{-1}(\boldsymbol{x}_1) + f^{-1}(\boldsymbol{x}_2)$. Como f^{-1} é linear, f^{-1} pode ser descrita por uma matriz $\boldsymbol{A}^{-1} \in \mathbb{C}^{n\times n}$ (estamos sempre assumindo a base canônica para o $\mathbb{C}^n$), ou seja,

$$f^{-1}(\boldsymbol{y}) = \boldsymbol{A}^{-1}\boldsymbol{y}, \text{ para todo } \boldsymbol{y} \in \mathbb{C}^n.$$

Por definição, $f^{-1}(f(\boldsymbol{x})) = \boldsymbol{x}$, e portanto, $\boldsymbol{A}^{-1}(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{x}$ para todo $\boldsymbol{x} \in \mathbb{C}^n$, ou seja, $\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_n$. Analogamente, $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}_n$.

Para mostrar que (e) implica (a), suponha por absurdo que (e) vale, mas (a) não. Então existiria um vetor não nulo $\boldsymbol{x} \neq \boldsymbol{0}$ em $\mathcal{N}(\boldsymbol{A})$, ou seja, $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$. Mas nesse caso, teríamos $\boldsymbol{A}^{-1}(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{0} \neq \boldsymbol{x}$, uma contradição. ■

Quando $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ e $m \neq n$, claramente não pode haver inversa (pois, se $m < n$, a transformação linear correspondente não pode ser injetora, e se $m > n$, a transformação não pode ser sobrejetora). O que se pode fazer nesse caso, quando nos deparamos com um problema $\boldsymbol{Ax} = \boldsymbol{b}$? Nesse caso, pode não haver solução (caso $\boldsymbol{b}$ não pertença ao espaço-imagem de $\boldsymbol{A}$), ou, se houver solução, ela pode não ser única (caso $\mathrm{nul}(\boldsymbol{A}) > 0$). Uma solução possível (não a única) é a *pseudo-inversa*. Estudaremos nos próximos capítulos formas de calcular a pseudo-inversa, mas por hora podemos defini-la e verificar algumas de suas propriedades.

Considere a transformação linear

$$\begin{aligned} f:\ & \mathbb{C}^n \to \mathbb{C}^m \\ & \boldsymbol{x} \mapsto \boldsymbol{y} = \boldsymbol{Ax}. \end{aligned}$$

Mesmo que f não seja sobrejetora, podemos definir uma transformação $\tilde{f}$ sobrejetora restringindo o contra-domínio de $\tilde{f}$ a $\mathcal{R}(\boldsymbol{A})$, ou seja, definimos

$$\begin{aligned} \tilde{f}:\ & \mathbb{C}^n \to \mathcal{R}(\boldsymbol{A}) \\ & \boldsymbol{x} \mapsto \boldsymbol{y} = \boldsymbol{Ax}. \end{aligned}$$

Repare que $\tilde{f}$ leva o vetor $\boldsymbol{x} \in \mathbb{C}^n$ ao mesmo vetor $\boldsymbol{y} = \boldsymbol{Ax}$, mas agora estamos considerando que $\boldsymbol{y} \in \mathcal{R}(\boldsymbol{A}) \subset \mathbb{C}^m$. Nessas condições, $\tilde{f}$ é sobrejetora, mas não necessariamente injetora.

Podemos agora restringir o *domínio* da $\tilde{f}$, definindo uma transformação $\hat{f}$ com domínio apenas em $\mathcal{N}(\boldsymbol{A})^\perp$

$$\begin{aligned} \hat{f}:\ & \mathcal{N}(\boldsymbol{A})^\perp \to \mathcal{R}(\boldsymbol{A}) \\ & \boldsymbol{x} \mapsto \boldsymbol{y} = \boldsymbol{Ax}. \end{aligned}$$

A transformação $\hat{f}$ é agora injetora e sobrejetora (pois, como vimos, todo vetor $\boldsymbol{y} \in \mathcal{R}(\boldsymbol{A})$ pode ser escrito como $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, com $\boldsymbol{x} \in \mathcal{N}(\boldsymbol{A})^{\perp}$). Apesar da f poder não possuir inversa, a $\hat{f}$ sempre possui inversa, pois por construção, $\hat{f}$ é injetora e sobrejetora:

$$\begin{aligned} \hat{f}^{-1}: \ \mathcal{R}(\boldsymbol{A}) &\to \mathcal{N}(\boldsymbol{A})^{\perp} \\ \boldsymbol{y} &\mapsto \boldsymbol{x} \text{ tal que } \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}. \end{aligned}$$

A pseudo-inversa f^{+} é definida estendendo-se o domínio e o contra-domínio de $\hat{f}$ para $\mathbb{C}^n$ e para $\mathbb{C}^m$, da seguinte maneira:

$$\begin{aligned} f^{+}: \ \mathbb{C}^m &\to \mathbb{C}^n \\ \boldsymbol{y} &\mapsto \boldsymbol{x} \text{ tal que } \boldsymbol{y}_1 = \boldsymbol{A}\boldsymbol{x}, \end{aligned}$$

em que $\boldsymbol{y} = \boldsymbol{y}_1 + \boldsymbol{y}_2$, e $\boldsymbol{y}_1 \in \mathcal{R}(\boldsymbol{A})$, $\boldsymbol{y}_2 \in \mathcal{R}(\boldsymbol{A})^{\perp}$. A pseudo-inversa da matriz $\boldsymbol{A}$ é a matriz $\boldsymbol{A}^{+}$ que descreve a transformação f^{+}. Repare que $\mathcal{N}(\boldsymbol{A}^{+}) = \mathcal{N}(\boldsymbol{A}^{H})$, $\mathcal{R}(\boldsymbol{A}^{+}) = \mathcal{R}(\boldsymbol{A}^{H})$.

Exemplo 3.4. Considere a matriz

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

O espaço-imagem de $\boldsymbol{A}$ é $\{\boldsymbol{y} \in \mathbb{R}^2 : \ \boldsymbol{y} = \alpha \begin{bmatrix} 0 & 1 \end{bmatrix}^T\}$, e $\mathcal{N}(\boldsymbol{A})\perp = \{\boldsymbol{x} \in \mathbb{R}^3 : \ \boldsymbol{x} = \beta \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T$. A pseudo-inversa de $\boldsymbol{A}$ será a matriz (veja a Figura 3.1)

$$\boldsymbol{A}^{+} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1/2 \end{bmatrix}.$$

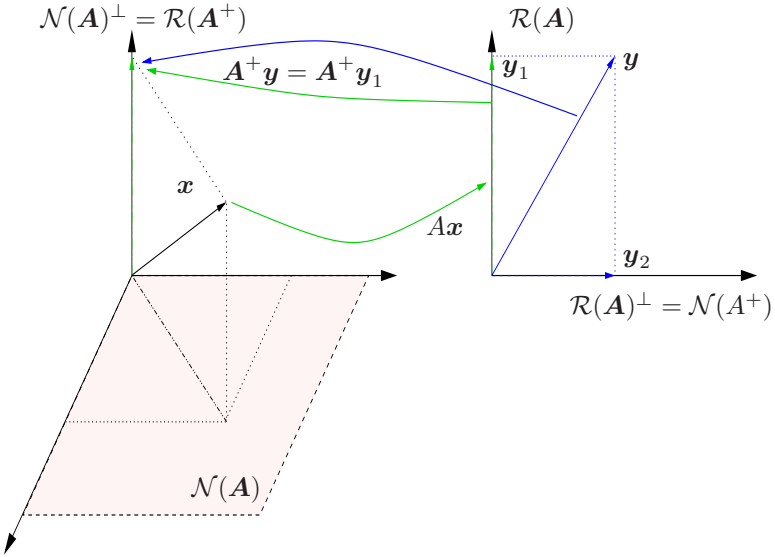

Figura 3.1: Exemplo de pseudo-inversa.

Capítulo 8

Decomposição em Valores Singulares (SVD)

8.1 Definição e Propriedades Principais

A decomposição em valores singulares (*singular-value decomposition, SVD*) é um dos resultados mais importantes em Álgebra Linear, tanto computacional quanto teórica. A SVD é a base dos métodos mais precisos para resolução de problemas de mínimos quadrados, para determinação do posto de matrizes, do espaço-imagem ($\mathcal{R}(\cdot)$) e do espaço nulo ($\mathcal{N}(\cdot)$) de matrizes, e da solução de vários problemas envolvendo normas euclideanas ($\|\cdot\|$).

O resultado básico é o seguinte teorema:

Teorema 8.1 (SVD). *Se* $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ *(*$\boldsymbol{A} \in \mathbb{R}^{m\times n}$*) então existem ma-*

trizes unitárias (ortogonais) $\boldsymbol{U} \in \mathbb{C}^{m\times m}$ $(\mathbb{R}^{m\times m})$ *e* $\boldsymbol{V} \in \mathbb{C}^{n\times n}$ $(\mathbb{R}^{n\times n})$ *tais que*

$$\boldsymbol{U}^H \boldsymbol{A}\boldsymbol{V} = \boldsymbol{S},$$

onde $\boldsymbol{S} \in \mathbb{R}^{m\times n}$ *é uma matriz* real *e "diagonal" com elementos* $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$, *com* $p = \min\{m,n\}$.

Por exemplo, para $m > n$, $\boldsymbol{S}$ *será da forma*

$$\boldsymbol{S} = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_p \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}$$

Os σ_i *são denominados* valores singulares *de* $\boldsymbol{A}$. *Note que há exatamente* $r = \text{posto}(\boldsymbol{A})$ *valores singulares não-nulos. As colunas de* $\boldsymbol{U}$ *e de* $\boldsymbol{V}$,

$$\boldsymbol{U} = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \dots & \boldsymbol{u}_m \end{bmatrix} \qquad e \qquad \boldsymbol{V} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \dots & \boldsymbol{v}_n \end{bmatrix}$$

são denominadas respectivamente vetores singulares à esquerda e à direita *de* $\boldsymbol{A}$.

Demonstração. Faça $\sigma_1 = \|\boldsymbol{A}\|_2$, e sejam $\boldsymbol{u}_1^{(1)} \in \mathbb{C}^m$, $\boldsymbol{v}_1^{(1)}\mathbb{C}^n$ tais que $\|\boldsymbol{u}_1^{(1)}\|_2 = \|\boldsymbol{v}_1^{(1)}\|_2 = 1$ e $\sigma_1\boldsymbol{u}_1^{(1)} = \boldsymbol{A}\boldsymbol{v}_1^{(1)}$ ($\boldsymbol{u}_1^{(1)}$ e $\boldsymbol{v}_1^{(1)}$ existem, pois $\sigma_1 = \max_{\|\boldsymbol{x}\|=1} \|\boldsymbol{A}\boldsymbol{x}\|_2$, e sabemos que o máximo é atingido).

Escolha agora vetores $\left\{\boldsymbol{u}_2^{(1)},\ldots,\boldsymbol{u}_m^{(1)}\right\}$ e $\left\{\boldsymbol{v}_2^{(1)},\ldots,\boldsymbol{v}_n^{(1)}\right\}$ tais que

$$\boldsymbol{U}^{(1)} = \begin{bmatrix}\boldsymbol{u}_1^{(1)} & \boldsymbol{u}_2^{(1)} & \ldots & \boldsymbol{u}_m^{(1)}\end{bmatrix} \quad \text{e} \quad \boldsymbol{V}^{(1)} = \begin{bmatrix}\boldsymbol{v}_1^{(1)} & \boldsymbol{v}_2^{(1)} & \ldots & \boldsymbol{v}_n^{(1)}\end{bmatrix}$$

sejam unitárias. Então,

$$\begin{aligned}
\boldsymbol{U}^{(1)\ H}\boldsymbol{A}\boldsymbol{V}^{(1)} &= \boldsymbol{U}^{(1)\ H}\begin{bmatrix}\boldsymbol{A}\boldsymbol{v}_1^{(1)} & \boldsymbol{A}\boldsymbol{v}_2^{(1)} & \ldots & \boldsymbol{A}\boldsymbol{v}_n^{(1)}\end{bmatrix} = \\
&= \begin{bmatrix}\boldsymbol{u}_1^{(1)\ H}\\ \boldsymbol{u}_2^{(1)\ H}\\ \vdots \\ \boldsymbol{u}_m^{(1)\ H}\end{bmatrix}\begin{bmatrix}\sigma_1\boldsymbol{u}_1^{(1)} & \boldsymbol{A}\boldsymbol{v}_2^{(1)} & \ldots & \boldsymbol{A}\boldsymbol{v}_n^{(1)}\end{bmatrix} \\
&= \begin{bmatrix}\sigma_1 & \boldsymbol{w}_1^H \\ \boldsymbol{0} & B_1\end{bmatrix} = \boldsymbol{A}_1.
\end{aligned}$$

Mas $\sigma_1 = \|\boldsymbol{A}\|_2 = \|\boldsymbol{A}_1\|_2$ (pois $\boldsymbol{U}^{(1)}$ e $\boldsymbol{V}^{(1)}$ são unitárias), e portanto

$$\begin{aligned}
\sigma_1 = \|\boldsymbol{A}_1\|_2 = \max_{\|\boldsymbol{x}\|_2=1}\|\boldsymbol{A}_1\boldsymbol{x}\|_2 &\geq \left\|\boldsymbol{A}_1\begin{bmatrix}\sigma_1\\ \boldsymbol{w}_1\end{bmatrix}\frac{1}{\sqrt{\sigma_1^2+\boldsymbol{w}_1^H\boldsymbol{w}_1}}\right\|_2 \\
&\geq \sqrt{\sigma_1^2+\boldsymbol{w}_1^H\boldsymbol{w}_1} \geq \sigma_1.
\end{aligned}$$

A desigualdade acima só pode ser satisfeita se $\boldsymbol{w} = \boldsymbol{0}$. A conclusão é que

$$\boldsymbol{U}^{(1)\ H}\boldsymbol{A}\boldsymbol{V}^{(1)} = \begin{bmatrix}\sigma_1 & \boldsymbol{0}^H \\ \boldsymbol{0} & \boldsymbol{B}_1\end{bmatrix}.$$

Repetindo agora o procedimento para $\boldsymbol{B}_1 \in \mathbb{C}^{(m-1)\times(n-1)}$, encontram-se $\boldsymbol{U}_1^{(2)}$ e $\boldsymbol{V}_1^{(2)}$ unitárias tais que

$$\boldsymbol{U}_1^{(2)\ H}\boldsymbol{B}_1\boldsymbol{V}_1^{(2)} = \begin{bmatrix} \sigma_2 & \boldsymbol{0}^H \\ \boldsymbol{0} & \boldsymbol{B}_2 \end{bmatrix}.$$

Definindo então

$$\boldsymbol{U}^{(2)} = \begin{bmatrix} 1 & \boldsymbol{0}^H \\ \boldsymbol{0} & \boldsymbol{U}_1^{(2)} \end{bmatrix} \qquad \text{e} \qquad \boldsymbol{V}^{(2)} = \begin{bmatrix} 1 & \boldsymbol{0}^H \\ \boldsymbol{0} & \boldsymbol{V}_1^{(2)} \end{bmatrix},$$

as matrizes $\boldsymbol{U}^{(i)}$, $\boldsymbol{V}^{(i)}$ e $\boldsymbol{A}$ satisfazem

$$\boldsymbol{U}^{(2)\ H}\boldsymbol{U}^{(1)\ H}\boldsymbol{A}\boldsymbol{V}^{(1)}\boldsymbol{V}^{(2)} = \begin{bmatrix} \sigma_1 & 0 & \boldsymbol{0}^H \\ 0 & \sigma_2 & \boldsymbol{0}^H \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{B}_2 \end{bmatrix}.$$

Continuando com o procedimento acima para $\boldsymbol{B}_2$, etc, chega-se à decomposição desejada.

■

Nota 1. É comum definir-se

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix},$$

com $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, e $r = \text{posto}(\boldsymbol{A})$. Com estas definições, escreve-se

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^H = \boldsymbol{U}\begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}\boldsymbol{V}^H.$$

Observe que alguns dos zeros acima podem não existir, dependendo do posto de $\boldsymbol{A}$. Assim, se posto$(\boldsymbol{A}) = n$, ter-se-á

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{\Sigma} \\ \boldsymbol{0} \end{bmatrix},$$

enquanto que, se posto$(\boldsymbol{A}) = m$, $\boldsymbol{S}$ será da forma

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{0} \end{bmatrix}.$$

Nota 2 (SVD compacta). Pode-se também escrever, de forma mais compacta,

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{U}_1 & \boldsymbol{U}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_1^H \\ \boldsymbol{V}_2^H \end{bmatrix} = \boldsymbol{U}_1 \boldsymbol{\Sigma} \boldsymbol{V}_1^H,$$

onde $\boldsymbol{\Sigma} \in \mathbb{R}^{r \times r}$ é como acima, e $\boldsymbol{U}_1 \in \mathbb{C}^{m \times r}$, $\boldsymbol{U}_2 \in \mathbb{C}^{m \times (m-r)}$, $\boldsymbol{V}_1 \in \mathbb{C}^{n \times r}$, $\boldsymbol{V}_2 \in \mathbb{C}^{n \times (n-r)}$.

Algumas propriedades da SVD estão listadas a seguir:

1. $\boldsymbol{V}_2$ é uma base ortonormal para $\mathcal{N}(\boldsymbol{A})$,
2. $\boldsymbol{V}_1$ é uma base ortonormal para $\mathcal{N}(\boldsymbol{A})^\perp$,
3. $\boldsymbol{U}_1$ é uma base ortonormal para $\mathcal{R}(\boldsymbol{A})$,
4. $\boldsymbol{U}_2$ é uma base ortonormal para $\mathcal{R}(\boldsymbol{A})^\perp$,
5. $\left\{\sigma_1^2, \ldots, \sigma_r^2\right\}$ são os autovalores não-nulos de $\boldsymbol{A}^H\boldsymbol{A}$ e de $\boldsymbol{A}\boldsymbol{A}^H$,

Demonstração. 1. A prova apresentada a seguir segue a seguinte linha: Como as colunas de $\boldsymbol{V}_2$ são ortonormais (já que $\boldsymbol{V}$ é unitária), basta

mostrar que $\mathcal{R}(\boldsymbol{V}_2) = \mathcal{N}(\boldsymbol{A})$. Isto será feito em duas etapas, primeiro provando que $\mathcal{R}(\boldsymbol{V}_2) \subset \mathcal{N}(\boldsymbol{A})$, e em seguida provando que $\mathcal{N}(\boldsymbol{A}) \subset \mathcal{R}(\boldsymbol{V}_2)$.

Começando: seja $\boldsymbol{y} \in \mathbb{C}^n$, $\boldsymbol{y} \in \mathcal{R}(\boldsymbol{V}_2)$ (ou seja: existe $\boldsymbol{z} \in \mathbb{C}^{n-r}$ tal que $\boldsymbol{y} = \boldsymbol{V}_2\boldsymbol{z}$). Como $\boldsymbol{V}$ é unitária, vale $\boldsymbol{V}_1^H\boldsymbol{V}_2 = \boldsymbol{0}$, e portanto $\boldsymbol{V}_1^H\boldsymbol{y} = \boldsymbol{V}_1^H\boldsymbol{V}_2\boldsymbol{z} = \boldsymbol{0}$. Desta forma,

$$\boldsymbol{A}\boldsymbol{y} = \boldsymbol{U}\begin{bmatrix}\boldsymbol{\Sigma} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{V}_1^H\\ \boldsymbol{V}_2^H\end{bmatrix}\boldsymbol{y} = \boldsymbol{U}\begin{bmatrix}\boldsymbol{\Sigma} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{V}_2^H\boldsymbol{y}\end{bmatrix} = \boldsymbol{0}.$$

Isto prova que $\boldsymbol{y} \in \mathcal{R}(\boldsymbol{V}_2) \Rightarrow \boldsymbol{y} \in \mathcal{N}(\boldsymbol{A})$, ou seja, $\mathcal{R}(\boldsymbol{V}_2) \subset \mathcal{N}(\boldsymbol{A})$.

A segunda inclusão será provada mostrando que se $\boldsymbol{y} \notin \mathcal{R}(\boldsymbol{V}_2)$, então $\boldsymbol{y} \notin \mathcal{N}(\boldsymbol{A})$ (convença-se que isto implica que $\mathcal{N}(\boldsymbol{A}) \subset \mathcal{R}(\boldsymbol{V}_2)$). Considere portanto um vetor $\boldsymbol{y} \in \mathbb{C}^m$. Por definição de subspaço ortogonal, existem vetores $\boldsymbol{y}_1$ e $\boldsymbol{y}_2$ tais que

$$\boldsymbol{y} = \boldsymbol{y}_1 + \boldsymbol{y}_2,$$

com $\boldsymbol{y}_1 \in \mathcal{R}(\boldsymbol{V}_2)^\perp$ e $\boldsymbol{y}_2 \in \mathcal{R}(\boldsymbol{V}_2)$. Se $\boldsymbol{y} \notin \mathcal{R}(\boldsymbol{V}_2)$, então necessariamente $\boldsymbol{y}_1 \neq \boldsymbol{0}$. Por outro lado, o fato de $\boldsymbol{V}$ ser unitária implica $\mathcal{R}(\boldsymbol{V}_2)^\perp = \mathcal{R}(\boldsymbol{V}_1)$, ou seja, $\boldsymbol{0} \neq \boldsymbol{y}_1 \in \mathcal{R}(\boldsymbol{V}_1)$. Verifique que isto significa que $\boldsymbol{V}_1^H\boldsymbol{y}_1 \neq \boldsymbol{0}$.

Isto prova que

$$\boldsymbol{V}_1^H\boldsymbol{y} = \boldsymbol{V}_1^H(\boldsymbol{y}_1 + \boldsymbol{y}_2) = \boldsymbol{V}_1^H\boldsymbol{y}_1 \neq \boldsymbol{0} \quad \text{se} \;\; \boldsymbol{y} \notin \mathcal{R}(\boldsymbol{V}_2),$$

e portanto,

$$\boldsymbol{A}\boldsymbol{y} = \boldsymbol{U}\begin{bmatrix}\boldsymbol{\Sigma} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{V}_1^H\boldsymbol{y}\\ \boldsymbol{V}_2^H\boldsymbol{y}\end{bmatrix} = \boldsymbol{U}\begin{bmatrix}\boldsymbol{\Sigma}\boldsymbol{V}_1^H\boldsymbol{y}_1\\ \boldsymbol{0}\end{bmatrix} \neq \boldsymbol{0}, \quad \text{se } \boldsymbol{y} \notin \mathcal{R}(\boldsymbol{V}_2),$$

ou seja, $\boldsymbol{y} \notin \mathcal{R}(\boldsymbol{V}_2) \Rightarrow \boldsymbol{y} \notin \mathcal{N}(\boldsymbol{A})$, o que completa a demonstração.

2. Demonstração similar à anterior.

3. Como $\boldsymbol{A} = \boldsymbol{U}_1\boldsymbol{\Sigma}\boldsymbol{V}_1^H$, é óbvio que $\mathcal{R}(\boldsymbol{A}) \subset \mathcal{R}(\boldsymbol{U}_1)$. Como posto$(\boldsymbol{U}_1) = r$, para provar que também vale $\mathcal{R}(\boldsymbol{U}_1) \subset \mathcal{R}(\boldsymbol{A})$, é necessário observar que tanto $\boldsymbol{\Sigma}$ quanto $\boldsymbol{V}_1^H$ têm posto máximo e igual a r (por que?).

Seja $\boldsymbol{z} \in \mathcal{R}(\boldsymbol{U}_1)$. Então existe $\boldsymbol{y} \in \mathbb{C}^r$ tal que $\boldsymbol{z} = \boldsymbol{U}_1\boldsymbol{y}$. Por outro lado, como posto$(\boldsymbol{\Sigma}\boldsymbol{V}_1^H) = r$, existe $\boldsymbol{x} \in \mathbb{C}^n$ tal que $\boldsymbol{y} = \boldsymbol{\Sigma}\boldsymbol{V}_1^H\boldsymbol{x}$.

A conclusão é que, dado $\boldsymbol{z} \in \mathcal{R}(\boldsymbol{U}_1)$, existe $\boldsymbol{x} \in \mathbb{C}^n$ tal que

$$\boldsymbol{z} = \boldsymbol{U}_1\boldsymbol{\Sigma}\boldsymbol{V}_1^H\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x},$$

ou seja, $\boldsymbol{z} \in \mathcal{R}(\boldsymbol{A})$, completando o argumento.

4. Semelhante à anterior.

5. Para verificar este resultado, basta usar a SVD ao calcular $\boldsymbol{A}\boldsymbol{A}^H$ e $\boldsymbol{A}^H\boldsymbol{A}$:

$$\boldsymbol{A}\boldsymbol{A}^H = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^H\left(\boldsymbol{V}\boldsymbol{S}^H\boldsymbol{U}^H\right) = \boldsymbol{U}\begin{bmatrix}\boldsymbol{\Sigma}^2 & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}\boldsymbol{U}^H,$$

onde

$$\boldsymbol{S}\boldsymbol{S}^H = \begin{bmatrix}\boldsymbol{\Sigma}^2 & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix} \in \mathbb{R}^{m\times m}$$

é uma matriz diagonal com elementos $\left\{\sigma_1^2,\sigma_2^2,\ldots,\sigma_r^2,0,\ldots,0\right\}$. O cálculo para $\boldsymbol{A}^H\boldsymbol{A}$ é semelhante. ■

8.2 Aproximação de Matrizes

A SVD também pode ser usada para encontrar a melhor aproximação $\boldsymbol{A}_k$ com posto k para uma matriz $\boldsymbol{A} \in \mathbb{C}^{m\times n}$. Esta aproximação é útil em problemas de identificação de sistemas, ou identificação de sinais em ruído. Note inicialmente que pode-se reescrever a SVD compacta da seguinte maneira:

$$\boldsymbol{A} = \boldsymbol{U}_1 \Sigma \boldsymbol{V}_1^H = \begin{bmatrix}\boldsymbol{u}_1 \dots \boldsymbol{u}_r\end{bmatrix} \begin{bmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_r\end{bmatrix} \begin{bmatrix}\boldsymbol{v}_1^H \\ \vdots \\ \boldsymbol{v}_r^H\end{bmatrix} =$$

$$= \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^H + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^H + \cdots + \sigma_r \boldsymbol{u}_r \boldsymbol{v}_r^H = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^H. \tag{8.1}$$

A partir desta decomposição, pode-se construir aproximações com posto k para $\boldsymbol{A}$ da seguinte forma:

$$\boldsymbol{A}_k = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^H. \tag{8.2}$$

O teorema a seguir prova que $\boldsymbol{A}_k$ é, na verdade, a matriz com posto k que mais se aproxima de $\boldsymbol{A}$ (segundo a norma $\|\cdot\|_2$).

Theorem 1. *Seja $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^H$ a SVD de $\boldsymbol{A} \in \mathbb{C}^{m\times n}$. Se $k < r =$ posto$(\boldsymbol{A})$, e com $\boldsymbol{A}_k$ como definido em* (8.2) *acima, então*

$$\min_{\text{posto}(\boldsymbol{B})=k} \|\boldsymbol{A} - \boldsymbol{B}\|_2 = \|\boldsymbol{A} - \boldsymbol{A}_k\|_2 = \sigma_{k+1}. \tag{8.3}$$

Demonstração. Primeiramente, é necessário verificar que o posto $\boldsymbol{A}_k$ é realmente k, e que $\|\boldsymbol{A} - \boldsymbol{A}_k\|_2 = \sigma_{k+1}$. Estes fatos são verificados facilmente, pois se

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^H, \quad \boldsymbol{S} = \begin{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

é a SVD de $\boldsymbol{A}$, as SVDs de $\boldsymbol{A}_k$ e de $\boldsymbol{A} - \boldsymbol{A}_k$ são

$$\boldsymbol{A}_k = \boldsymbol{U} \begin{bmatrix} \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_k & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \boldsymbol{V}^H, \qquad \text{e}$$

$$\boldsymbol{A} - \boldsymbol{A}_k = \boldsymbol{U} \begin{bmatrix} \begin{bmatrix} 0 & & & & & \\ & \ddots & & & & \\ & & 0 & & & \\ & & & \sigma_{k+1} & & \\ & & & & \ddots & \\ & & & & & \sigma_r \end{bmatrix} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \boldsymbol{V}^H.$$

Para provar que $\boldsymbol{A}_k$ é realmente a solução de (8.3), vamos provar que todo $\boldsymbol{B} \in \mathbb{C}^{m\times n}$ com posto k é tal que

$$\|\boldsymbol{A}-\boldsymbol{B}\|_2 \geq \sigma_{k+1}.$$

Para verificar isto, basta mostrar que existe um vetor unitário $\boldsymbol{z}$ tal que $\|(\boldsymbol{A}-\boldsymbol{B})\boldsymbol{z}\|_2 \geq \sigma_{k+1}$.

Seja então $\boldsymbol{B} \in \mathbb{C}^{m\times n}$ tal que posto$(\boldsymbol{B}) = k$. Por definição de posto, existem vetores ortonormais $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{n-k} \in \mathbb{C}^n$ tais que $\mathcal{N}(\boldsymbol{B}) = \mathrm{Sp}\big(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{n-k}\big)$.

Como $\big\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{n-k},\boldsymbol{v}_1,\ldots,\boldsymbol{v}_{k+1}\big\}$ tem $n-k+k+1 = n+1$ elementos, e como todos os vetores $\boldsymbol{x}_i,\boldsymbol{v}_j$ pertencem ao $\mathbb{C}^n$, a intersecção abaixo

$$\mathrm{Sp}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{n-k}) \cap \mathrm{Sp}(\boldsymbol{v}_1,\ldots,\boldsymbol{v}_{k+1}) \neq \big\{\boldsymbol{0}\big\}$$

necessariamente deverá conter pelo menos um vetor não-nulo, e portanto podemos escolher um vetor unitário $\boldsymbol{z}$ nesta intersecção. Para este vetor, vale

$$\begin{aligned}
&\boldsymbol{B}\boldsymbol{z} = \boldsymbol{0}, && \text{pois} \quad \boldsymbol{z} \in \mathcal{N}(\boldsymbol{B}),\\
&\boldsymbol{A}\boldsymbol{z} = \sum_{i=1}^{r}\big(\sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^H\big)\boldsymbol{z} = \sum_{i=1}^{k+1}\sigma_i\big(\boldsymbol{v}_i^H\boldsymbol{z}\big)\boldsymbol{u}_i, && \text{pois} \quad \boldsymbol{z} \in \mathrm{Sp}(\boldsymbol{v}_1,\ldots,\boldsymbol{v}_{k+1}).
\end{aligned}$$

Disto segue que

$$\begin{aligned}
\|\boldsymbol{A}-\boldsymbol{B}\|_2 &= \max_{\|\boldsymbol{y}\|_2=1}\|(\boldsymbol{A}-\boldsymbol{B})\boldsymbol{y}\|_2 \geq \|(\boldsymbol{A}-\boldsymbol{B})\boldsymbol{z}\|_2 = \|\boldsymbol{A}\boldsymbol{z}\|_2 =\\
&= \left\|\sum_{i=1}^{k+1}\sigma_i(\boldsymbol{v}_i^H\boldsymbol{z})\boldsymbol{u}_i\right\|_2 \geq \sigma_{k+1}\left\|\sum_{i=1}^{k+1}(\boldsymbol{v}_i^H\boldsymbol{z})\boldsymbol{u}_i\right\|_2 = \sigma_{k+1}.
\end{aligned}$$

Para verificar a última igualdade, lembre-se que $\boldsymbol{z} \in \mathrm{Sp}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{k+1})$ e que $\|\boldsymbol{z}\|_2 = 1$. Portanto, $\boldsymbol{v}_j^H \boldsymbol{z} = 0$ se $j \geq k+2$, e

$$\sum_{i=1}^{k+1} (\boldsymbol{v}_i^H \boldsymbol{z}) \boldsymbol{u}_i = \begin{bmatrix} \boldsymbol{u}_1 & \ldots & \boldsymbol{u}_{k+1} & \boldsymbol{u}_{k+2} & \ldots & \boldsymbol{u}_m \end{bmatrix} \begin{bmatrix} \begin{bmatrix} \boldsymbol{v}_1^H \\ \ldots \\ \boldsymbol{v}_{k+1}^H \end{bmatrix} \boldsymbol{z} \\ \boldsymbol{0} \end{bmatrix} = \boldsymbol{U}\boldsymbol{V}^H \boldsymbol{z}.$$

Assim,

$$\left\| \sum_{i=1}^{k+1} (\boldsymbol{v}_i^H \boldsymbol{z}) \boldsymbol{u}_i \right\|_2 = \|\boldsymbol{U}\boldsymbol{V}^H \boldsymbol{z}\|_2 = \|\boldsymbol{z}\|_2 = 1.$$

■

Exemplo 8.1 (Determinação de senóides em ruído). Como um exemplo de utilização do resultado acima, considere o seguinte problema de detecção de sinais senoidais em ruído: Dispõe-se de uma seqüência $x(n)$, que consiste na soma de diversos sinais senoidais de freqüências diferentes, e ruído:

$$x(n) = \sum_{i=1}^{N} cos(\omega_i t + \phi_i) + v(n),$$

onde $\{v(n)\}$ é um ruído branco.

Apenas observando o sinal $x(n)$, pode não ser simples determinar quantos sinais senoidais estão presentes, nem a sua freqüência. Para se resolver este problema, pode-se calcular a transformada discreta de Fourier (TDF) da seqüência $\{x(n)\}_{n=0}^{M}$. No entanto, pode-se mostrar que a estimativa obtida é não-consistente (isto é, a variância da estimativa não tende para zero quando o tamanho da seqüência M tende para infinito).

Dois métodos alternativos que não sofrem deste problema são conhecidos como MUSIC e ESPRIT, descritos em [11]. Estes métodos têm como base a decomposição em valores singulares da matriz de autocovariância[1] do sinal descrita abaixo.

Inicialmente, determina-se uma aproximação para a matriz de autocovariância de $x(n)$, ou seja, calcula-se

$$\boldsymbol{R}_{L,T} = \frac{1}{T-1} \sum_{-T/2}^{T/2} \begin{bmatrix} x(n)^2 & \dots & x(n)x(n-L) \\ x(n-1)x(n) & \dots & x(n-1)x(n-L) \\ \vdots & \ddots & \vdots \\ x(n-L)x(n) & \dots & x(n-L)^2 \end{bmatrix},$$

onde $T = M - L$.

Observe que cada linha de $R_{L,T}$ é composta de valores da função de autocorrelação de $x(n)$ para diferentes atrazos. Lembrando que a função de autocorrelação de um sinal senoidal com freqüência angular ω tem transformada de Fourier com apenas duas harmônicas não-nulas (em ω e em $-\omega$), pode-se mostrar que, na ausência de ruído, e no limite para $T \to \infty$, a matriz $R_{L,T}$ tende a ter posto igual a duas vezes o número de senóides em $x(n)$, ou seja,

$$\text{posto}\left(\lim_{T\to\infty} \boldsymbol{R}_{L,T}\right) = 2N, \quad \text{na ausência de ruído.}$$

[1]Como a matriz de autocovariância é positiva semi-definida, a SVD é na verdade igual à diagonalização, e os valores singulares são iguais aos autovalores.

Desta forma, na ausência de ruído, a SVD de $\boldsymbol{R}_{L,\infty}$ será

$$\boldsymbol{R}_{L,\infty} = \boldsymbol{U} \begin{bmatrix} \sigma_1 & & & & & \\ & \sigma_2 & & & & \\ & & \ddots & & & \\ & & & \sigma_{2N} & & \\ & & & & 0 & \\ & & & & & \ddots \end{bmatrix} \boldsymbol{U}^T.$$

Note que, como $\boldsymbol{R}_{L,\infty}$ é simétrica, a SVD fica $\boldsymbol{U}\boldsymbol{S}\boldsymbol{U}^T$, em vez de $\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$.

Quando o sinal está imerso em ruído, a autocovariância do ruído, $\boldsymbol{R}_v = \sigma_v^2 I$, é adicionada a $\boldsymbol{R}_{L,\infty}$, e o posto da matriz resultante será máximo (igual a $L+1$). No entanto, haverá uma distinção clara entre os valores singulares devidos ao ruído e os valores singulares devidos às senóides (a não ser que o nível de ruído seja muito elevado).

As freqüências das senóides também podem ser calculadas a partir de uma SVD. Uma vez determinado o número N de senóides distintas presentes no sinal, calcula-se a SVD $\boldsymbol{R}_{2N+1,M-2N-1} = \boldsymbol{U}_{2N+1}\boldsymbol{S}_{2N+1}\boldsymbol{U}_{2N+1}^T$. A partir do vetor singular $\boldsymbol{u}_{2N+1}$, relativo ao menor valor singular de $\boldsymbol{R}_{2N+1,M-2N-1}$, pode-se estimar as freqüências das senóides componentes do sinal $x(n)$, calculando-se os zeros do polinômio

$$p(z) = z^{2N} + \frac{u_{2N+1}^{(2)}}{u_{2N+1}^{(1)}} z^{2N-1} + \cdots + \frac{u_{2N+1}^{(2N+1)}}{u_{2N+1}^{(1)}},$$

onde $u_{2N+1}^{(i)}$ é o i-ésimo elemento de $\boldsymbol{u}_{2N+1}$ (ver [11]).

Na figura 8.1 os resultados acima estão exemplificados com uma

seqüência $\{s(n)\}$ gerada como a seguir,

$$s(n) = \text{sen}(2\pi \times 0{,}1n) + \cos(2\pi \times 0{,}2718n) + 2 * v(n),$$

onde $v(n)$ é um ruído gaussiano branco com variância $\sigma_v^2 = 4$. A figura 8.1.a mostra um trecho do gráfico de $s(n) \times n$, enquanto que a figura 8.1.b mostra os valores singulares de $R_{40,10^5}$.

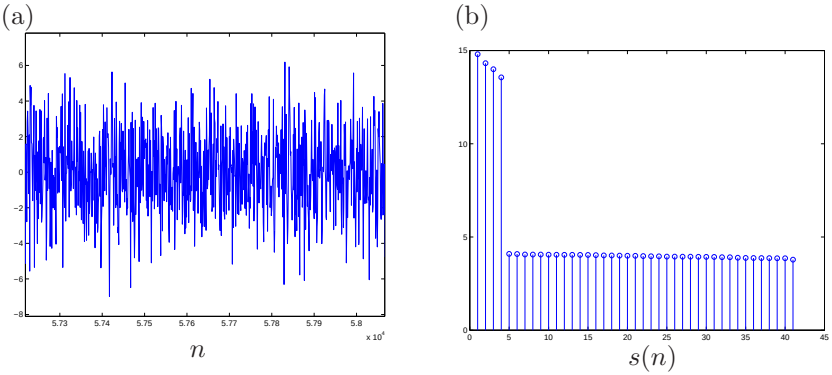

Figura 8.1: *(a) Trecho do sinal s(n) definido acima. (b) Valores singulares da matriz de autocovariância* $R_{40,10^5}$.

Outro exemplo da utilização prática de valores singulares pode ser encontrado em [1]. Trata-se de um método para codificação de sinais de voz, que utiliza a decomposição em valores singulares. A idéia é utilizar (8.1) e alocar um número maior de bits para componentes do sinal correspondentes a valores singulares maiores. Maiores detalhes podem ser encontrados na referência mencionada.

8.3 Propriedades Numéricas da SVD

Uma característica importante da SVD é que o problema "achar a decomposição em valores singulares de $\boldsymbol{A}$" é sempre bem-condicionado, e existem algoritmos muito estáveis para resolver este problema. O resultado a seguir, provado em [3, p. 260], resume estas propriedades:

Pode-se mostrar que o algoritmo descrito em [3] para cálculo da SVD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$ de uma matriz $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ calcula matrizes $\hat{\boldsymbol{U}}$, $\hat{\boldsymbol{V}}$ e $\hat{\boldsymbol{S}}$ satisfazendo

$$\begin{aligned}
\hat{\boldsymbol{U}} &= \boldsymbol{W} + \boldsymbol{\Delta U}, & \boldsymbol{W}^T\boldsymbol{W} &= \boldsymbol{I}, & \|\boldsymbol{\Delta U}\|_2 &\le \varepsilon,\\
\hat{\boldsymbol{V}} &= \boldsymbol{Z} + \boldsymbol{\Delta V}, & \boldsymbol{Z}^T\boldsymbol{Z} &= \boldsymbol{I}, & \|\boldsymbol{\Delta V}\|_2 &\le \varepsilon,\\
\hat{\boldsymbol{S}} &= \boldsymbol{W}^T(\boldsymbol{A} + \boldsymbol{\Delta A})\boldsymbol{Z}, & \|\boldsymbol{\Delta A}\|_2 &\le \varepsilon\,\|\boldsymbol{A}\|_2 = \varepsilon\sigma_1,
\end{aligned}$$

onde ε é um múltiplo pequeno de eps, a precisão da máquina.

Além disso, se $\sigma_1, \ldots, \sigma_p$ forem os valores singulares de $\boldsymbol{A}$ (com $p = \min\{m,n\}$), e $\hat{\sigma}_1, \ldots, \hat{\sigma}_p$ forem os valores singulares calculados numericamente, então

$$|\sigma_k - \hat{\sigma}_k| < \varepsilon\sigma_1 = \varepsilon\,\|\boldsymbol{A}\|_2\,, \quad \text{para } k = 1\ldots p.$$

Nota 3. O resultado acima mostra que as matrizes $\hat{\boldsymbol{U}}$ e $\hat{\boldsymbol{V}}$ calculadas não são exatamente unitárias, mas são pequenas perturbações de matrizes unitárias.

Nota 4. Conclui-se também que os valores singulares calculados $\hat{\sigma}_1, \ldots, \hat{\sigma}_p$ estão próximos dos valores singulares reais. No entanto, a precisão relativa no cálculo dos valores singulares só é realmente boa para valores singulares não muito menores que σ_1. Isto tem uma conseqüência prática: se for encontrado um valor singular com ordem de grandeza próxima de

eps $\hat{\sigma}_1 \approx$ eps σ_1, não será possível dizer com certeza se este valor singular é diferente de zero ou não (pois o erro $\varepsilon\sigma_1$ será da mesma ordem que o valor singular $\hat{\sigma}_k$). Isto gera uma incerteza sobre o posto da matriz $\boldsymbol{A}$.

Por exemplo, observe a SVD da matriz abaixo, calculada pelo Matlab:

$$\boldsymbol{A} = \begin{bmatrix} -1{,}4 & -0{,}8 & 1{,}1 & -0{,}4 \\ 0{,}3 & 0{,}6 & -0{,}45 & 0{,}3 \\ 0{,}8 & 1{,}1 & -0{,}95 & 0{,}55 \\ -0{,}4 & 0{,}2 & 0{,}1 & 0{,}1 \\ 0{,}5 & 0{,}5 & -0{,}5 & 0{,}25 \end{bmatrix} =$$

$$= \boldsymbol{U} \begin{bmatrix} 2{,}8835 & 0 & 0 & 0 \\ 0 & 0{,}69684 & 0 & 0 \\ 0 & 0 & 2{,}0386 \times 10^{-16} & 0 \\ 0 & 0 & 0 & 4{,}0255 \times 10^{-17} \\ 0 & 0 & 0 & 0 \end{bmatrix} \boldsymbol{V}^T.$$

Os cálculos feitos pelo Matlab têm precisão eps $= 2{,}2 \times 10^{-16}$, portanto, *para efeitos práticos*, não podemos distingüir os dois menores valores singulares de $\boldsymbol{A}$ de zero: *numericamente, a matriz* $\boldsymbol{A}$ *tem posto igual a 2*, mesmo sendo os valores singulares *calculados* todos não-nulos.

Em geral, os valores singulares de uma matriz qualquer são a melhor forma de se determinar, numericamente, o seu posto. Mas obviamente esta determinação é sujeita a erros. Quanto pior for a precisão dos cálculos (quanto maior eps) e quanto maior a imprecisão na obtenção da matriz (ela pode ser obtida através de medições, por exemplo), mais incerta será a determinação do posto.

Capítulo 9

O Produto de Kronecker

9.1 Definição e propriedades básicas

O produto de Kronecker, ou produto tensorial, é útil para resolver equações lineares em que a incógnita é uma matriz. Por exemplo, uma equação de Liapunov, encontrada em estudos de estabilidade de sistemas dinâmicos, é do tipo

$$\boldsymbol{A}P + P\boldsymbol{A}^T = Q, \tag{9.1}$$

em que $\boldsymbol{A}$ e Q são matrizes dadas de dimensão $n \times n$, Q é (usualmente) negativa-definida, e procura-se uma solução P positiva-definida.

A equação acima é linear em P, como pode ser facilmente verificado. No entanto, sua resolução não é imediata usando procedimentos usuais. Uma forma de se encontrar uma solução para (9.1) é usar o *Produto de Kronecker*, como veremos a seguir.

Definição 9.1 (Produto de Kronecker)**.** Dadas duas matrizes $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ e $B \in \mathbb{C}^{p\times q}$, o produto de Kronecker de $\boldsymbol{A}$ por B, denotado $\boldsymbol{A} \otimes B$, é dado por

$$\boldsymbol{A} \otimes B = \begin{bmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{n1}B & \dots & a_{nn}B \end{bmatrix}.$$

Veja que as dimensões de $\boldsymbol{A} \otimes B$ são $mp \times nq$, e que o produto de Kronecker é definido para matrizes $\boldsymbol{A}$, B de quaisquer dimensões.

Repare que, em geral, $\boldsymbol{A} \otimes B \neq B \otimes \boldsymbol{A}$. Por exemplo, se

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

então

$$\boldsymbol{A} \otimes B = \left[\begin{array}{cc|cc|cc} 1 & 1 & 2 & 2 & 3 & 3 \\ 1 & -1 & 2 & -2 & 3 & -3 \\ \hline 4 & 4 & 5 & 5 & 6 & 6 \\ 4 & -4 & 5 & -5 & 6 & -6 \end{array}\right],$$

$$B \otimes \boldsymbol{A} = \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 2 & 3 \\ 4 & 5 & 6 & 4 & 5 & 6 \\ \hline 1 & 2 & 3 & -1 & -2 & -3 \\ 4 & 5 & 6 & -4 & -5 & -6 \end{array}\right].$$

Também será útil definir como $\mathrm{vec}(\boldsymbol{A})$ o vetor de dimensão mn obtido colocando as colunas de $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ umas sobre as outras:

Definição 9.2 (vec)**.** Dada $\boldsymbol{A} \in \mathbb{C}^{m\times n}$, denota-se por $\mathrm{vec}(\boldsymbol{A})$ o vetor

$$\mathrm{vec}(\boldsymbol{A}) = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{n1} & a_{12} & \dots & a_{n2} & \dots & a_{nn} \end{bmatrix}^T.$$

As propriedades básicas do produto de Kronecker, facilmente verificáveis a partir da definição, são (nas linhas abaixo, $\boldsymbol{A}$,$\boldsymbol{A}_1 \in \mathbb{C}^{m\times n}$, B,$B_1 \in \mathbb{C}^{p\times q}$, $C \in \mathbb{C}^{n\times r}$, $D \in \mathbb{C}^{q\times s}$, $E \in \mathbb{C}^{r\times s}$, $\alpha \in \mathbb{C}$):

1. $(A + A_1) \otimes B = A \otimes B + A_1 \otimes B$,
2. $A \otimes (B + B_1) = A \otimes B + A \otimes B_1$,
3. $(\alpha A) \otimes B = A \otimes (\alpha B) = \alpha(A \otimes B)$,
4. $(A \otimes B)^H = A^H \otimes B^H$,
5. $(A \otimes B) \otimes E = A \otimes (B \otimes E)$,
6. $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$.

A propriedade 6 pode ser verificada como segue:

$$\begin{aligned}
(A \otimes B)(C \otimes D) &= \begin{bmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{bmatrix} \begin{bmatrix} c_{11}D & \dots & c_{1r}D \\ \vdots & \ddots & \vdots \\ c_{n1}D & \dots & c_{nr}D \end{bmatrix} = \\
&= \begin{bmatrix} \left(\sum_{i=1}^n a_{1i}c_{i1}\right)BD & \dots & \left(\sum_{i=1}^n a_{1i}c_{ir}\right)BD \\ \vdots & \ddots & \vdots \\ \left(\sum_{i=1}^n a_{mi}c_{i1}\right)BD & \dots & \left(\sum_{i=1}^n a_{mi}c_{ir}\right)BD \end{bmatrix} \\
&= (AC) \otimes (BD).
\end{aligned}$$

A partir das propriedades acima, fica simples demonstrar que (as matrizes a seguir têm as dimensões indicadas acima):

7. Se A e B forem hermitianas, então $A \otimes B$ é também hermitiana.

8. Se A e B forem invertíveis, então $A \otimes B$ é também invertível, e $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.

9. Se U e V forem unitárias, então $U \otimes V$ também é unitária.

10. Se as decomposições de $A \in \mathbb{C}^{m\times m}$ e $B \in \mathbb{C}^{n\times n}$ na forma de Schur forem, respectivamente, $A = UR_AU^H$ e $B = VR_BV^H$, então a decomposição de $A \otimes B$ na forma de Schur é

$$\begin{aligned} A \otimes B &= \big(UR_AU^H\big) \otimes \big(VR_BV^H\big) = \big[(UR_A) \otimes (VR_B)\big]\big(U^H \otimes V^H\big) = \\ &= \big(U \otimes V\big)\big(R_A \otimes R_B\big)\big(U \otimes V\big)^H, \end{aligned}$$

Repare que $R_A \otimes R_B$ é uma matriz triangular superior. Decorre dessa observação que os autovalores de $A \otimes B$ são da forma $\lambda_i\mu_j$, $1 \le i \le m$, $1 \le j \le n$, onde $\sigma(A) = \big\{\lambda_1, \ldots, \lambda_m\big\}$ e $\sigma(B) = \big\{\mu_1, \ldots, \mu_n\big\}$ são os autovalores de A e B, respectivamente.

11. Segue do item anterior que $\operatorname{posto}(A \otimes B) = \operatorname{posto}(A)\operatorname{posto}(B)$.

12. Se $A \in \mathbb{C}^{m\times n}$ e $B \in \mathbb{C}^{p\times q}$ tiverem decomposições em valores singulares $A = U_AS_AV_A^H$ e $B = U_BS_BV_B^H$, então a decomposição em valores singulares de $A \otimes B$ é

$$A \otimes B = \big(U_A \otimes U_B\big)\big(S_A \otimes S_B\big)\big(V_A \otimes V_B\big)^H.$$

Portanto, se os valores singulares de A forem $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_{r_a}$ e os autovalores de B forem $\eta_1 \ge \eta_2 \ge \cdots \ge \eta_{r_b}$ ($r_a = \operatorname{posto}(A)$ e $r_b = \operatorname{posto}(B)$), então o maior valor singular de $A \otimes B$ é $\sigma_1\eta_1$, e o menor, $\sigma_{r_a}\eta_{r_b}$.

13. Decorre do item anterior que $|||A \otimes B|||_2 = |||A|||_2 |||B|||_2$.

14. Se $\boldsymbol{x}$ é um autovetor de A relativo ao autovalor λ, e $\boldsymbol{y}$ é um autovetor de B relativo ao autovalor μ, então $\boldsymbol{x} \otimes \boldsymbol{y}$ é um autovetor de $A \otimes B$ relativo ao autovalor $\lambda\mu$.

9.2 Resolução de equações de Sylvester

Equações de Sylvester são equações do tipo

$$A_1 X B_1 + A_2 X B_2 + \cdots + A_k X B_k = C. \tag{9.2}$$

As equações de Liapunov mencionadas na seção anterior são portanto casos particulares de equações de Sylvester. A resolução desse tipo de equação usando produtos de Kronecker baseia-se na seguinte propriedade:

Teorema 9.1. *Se $A \in \mathbb{C}^{m\times n}$ e $B \in \mathbb{C}^{r\times s}$, $X \in \mathbb{C}n \times r$, $C \in \mathbb{C}^{m\times s}$ e $AXB = C$, então*

$$\left(B^T \otimes A\right) \operatorname{vec}(X) = \operatorname{vec}(C).$$

Prova: Para demonstrar essa propriedade, como em vários dos casos anteriores, basta escrever um termo geral de cada lado da equação e verificar a igualdade. Neste caso,

$$c_{ij} = \sum_{k=1}^{n} \sum_{l=1}^{r} a_{ik} x_{kl} b_{lj}.$$

Note que o termo correspondente a c_{ij} em $\mathrm{vec}(C)$ é o $i+j(s-1)$, ou seja, $c_{ij} = \big(\mathrm{vec}(C)\big)_{i+j(s-1)}$. A linha $i+j(s-1)$ de $B^T \otimes A$ (denotada por $\boldsymbol{l}^T_{i+j(s-1)}$ é:

$$\boldsymbol{l}^T_{i+j(s-1)} = \begin{bmatrix} b_{1j}\boldsymbol{a}_i^T & b_{2j}\boldsymbol{a}_i^T & \dots & b_{rj}\boldsymbol{a}_i^T \end{bmatrix},$$

em que $\boldsymbol{a}_i^T$ é a i-ésima linha de A. Denotando por $\boldsymbol{x}_l$ a l-ésima coluna de X, segue que

$$\begin{aligned}\Big[\big(B^T \otimes A\big)\,\mathrm{vec}(X)\Big]_{i+j(s-1)} &= \boldsymbol{l}^T_{i+j(s-1)}\,\mathrm{vec}(X) = \sum_{l=1}^{r} b_{lj}\boldsymbol{a}_i^T\boldsymbol{x}_l \\ &= \sum_{l=1}^{r}\sum_{k=1}^{n} b_{lj}a_{ik}x_{kl} = c_{ij}.\end{aligned}$$

■

Como corolário desse teorema, temos

Corolário 9.1 (Equações de Sylvester)**.** *A solução de uma equação de Sylvester como* (9.2) *é dada por*

$$\mathrm{vec}(X) = \Big[B_1^T \otimes A_1 + B_2^T \otimes A_2 + \cdots + B_k^T \otimes A_k\Big]^{-1}\mathrm{vec}(C), \qquad (9.3)$$

desde que a matriz entre colchetes seja não-singular.

Referências Bibliográficas

[1] B. S. Atal e B. E. Caspers. Beyond multipulse and CELP towards high quality speech at 4kb/s. In B. S. Atal, V. Cuperman, e A. Gersho, editors, *Advances in Speech Coding*, pp. 191–202. Kluwer Academic Publishers, 1991.

[2] F. R. Gantmacher. *The Theory of Matrices*. Chelsea Publishing Company, 2nd edition, 1990.

[3] G. H. Golub e C. F. V. Loan. *Matrix Computations*. Johns Hopkins, 3rd edition, 1996.

[4] W. H. Greub. *Linear Algebra*. Graduate Texts in Mathematics. Springer, 1981.

[5] N. J. Higham. *Accuracy and Stability of Numerical Algorithms*. SIAM, 1996.

[6] R. A. Horn e C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1987.

[7] R. A. Horn e C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, Cambridge, MA, 1991.

[8] A. J. Laub. *Matrix Analysis for Scientists and Engineers*. Society for Industrial and Applied Mathematics, PA, dez. 2005.

[9] C. D. Meyer. *Matrix Analysis and Applied Linear Algebra*. SIAM, Philadelphia, USA, 2000.

[10] W. H. Press, S. A. Teukolsky, W. T. Vetterling, e B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 2nd edition, 1994.

[11] J. G. Proakis, C. M. Rader, F. Ling, e C. L. Nikias. *Advanced Digital Signal Processing*. Maxwell Macmillan International Editions, 1992.

[12] G. W. Stewart. *Introduction to Matrix Computations*. Academic Press, 1973.

[13] G. Strang. *Linear Algebra and Its Applications*. Brooks Cole, 3rd edition, 1988.

[14] J. H. Wilkinson. *Rounding Errors in Algebraic Processes*. Her Majesty's Stationery Office, 1963.

[15] J. H. Wilkinson. *The Algebraic Eigenvalue Problem*. Oxford University Press, London, UK, 1965.